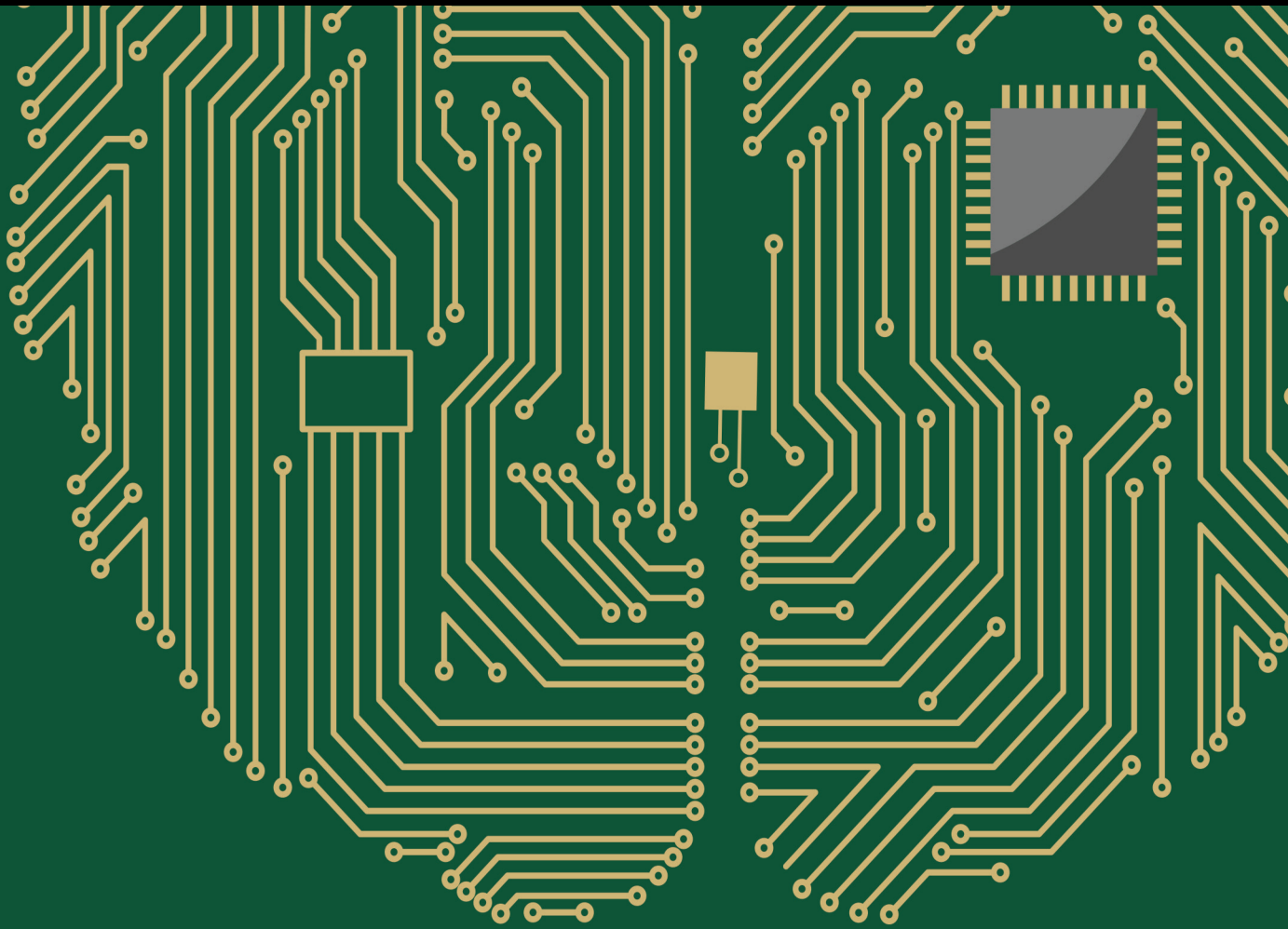


Design and Optimization of Human Cognitive Systems using Computational Intelligence Methods

Lead Guest Editor: Shengrong Gong

Guest Editors: Tongguang Ni and Wenbing Zhao





**Design and Optimization of Human Cognitive
Systems using Computational Intelligence
Methods**

Computational Intelligence and Neuroscience

**Design and Optimization of Human
Cognitive Systems using Computational
Intelligence Methods**

Lead Guest Editor: Shengrong Gong

Guest Editors: Tongguang Ni and Wenbing Zhao



Copyright © 2022 Hindawi Limited. All rights reserved.

This is a special issue published in "Computational Intelligence and Neuroscience." 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

Andrzej Cichocki, Poland

Associate Editors

Arnaud Delorme, France
Cheng-Jian Lin , Taiwan
Saeid Sanei, United Kingdom

Academic Editors


Mohamed Abd Elaziz , Egypt
Tariq Ahanger , Saudi Arabia
Muhammad Ahmad, Pakistan
Ricardo Aler , Spain
Nouman Ali, Pakistan
Pietro Aricò , Italy
Lerina Aversano , Italy
Ümit Ağbulut , Turkey
Najib Ben Aoun , Saudi Arabia
Surbhi Bhatia , Saudi Arabia
Daniele Bibbo , Italy
Vince D. Calhoun , USA
Francesco Camastra, Italy
Zhicheng Cao, China
Hubert Cecotti , USA
Jyotir Moy Chatterjee , Nepal
Rupesh Chikara, USA
Marta Cimitile, Italy
Silvia Conforto , Italy
Paolo Crippa , Italy
Christian W. Dawson, United Kingdom
Carmen De Maio , Italy
Thomas DeMarse , USA
Maria Jose Del Jesus, Spain
Arnaud Delorme , France
Anastasios D. Doulamis, Greece
António Dourado , Portugal
Sheng Du , China
Said El Kafhali , Morocco
Mohammad Reza Feizi Derakhshi , Iran
Quanxi Feng, China
Zhong-kai Feng, China
Steven L. Fernandes, USA
Agostino Forestiero , Italy
Piotr Franaszczuk , USA
Thippa Reddy Gadekallu , India
Paolo Gastaldo , Italy
Samanwoy Ghosh-Dastidar, USA

Manuel Graña , Spain
Alberto Guillén , Spain
Gaurav Gupta, India
Rodolfo E. Haber , Spain
Usman Habib , Pakistan
Anandakumar Haldorai , India
José Alfredo Hernández-Pérez , Mexico
Luis Javier Herrera , Spain
Alexander Hošovský , Slovakia
Etienne Hugues, USA
Nadeem Iqbal , Pakistan
Sajad Jafari, Iran
Abdul Rehman Javed , Pakistan
Jing Jin , China
Li Jin, United Kingdom
Kanak Kalita, India
Ryotaro Kamimura , Japan
Pasi A. Karjalainen , Finland
Anitha Karthikeyan, Saint Vincent and the
Grenadines
Elpida Keravnou , Cyprus
Asif Irshad Khan , Saudi Arabia
Muhammad Adnan Khan , Republic of
Korea
Abbas Khosravi, Australia
Tai-hoon Kim, Republic of Korea
Li-Wei Ko , Taiwan
Raşit Köker , Turkey
Deepika Koundal , India
Sunil Kumar , India
Fabio La Foresta, Italy
Kuruva Lakshmana , India
Maciej Lawrynczuk , Poland
Jianli Liu , China
Giosuè Lo Bosco , Italy
Andrea Loddo , Italy
Kezhi Mao, Singapore
Paolo Massobrio , Italy
Gerard McKee, Nigeria
Mohit Mittal , France
Paulo Moura Oliveira , Portugal
Debajyoti Mukhopadhyay , India
Xin Ning , China
Nasimul Noman , Australia
Fivos Panetsos , Spain

Evgeniya Pankratova , Russia
Rocío Pérez de Prado , Spain
Francesco Pistolesi , Italy
Alessandro Sebastian Podda , Italy
David M Powers, Australia
Radu-Emil Precup, Romania
Lorenzo Putzu, Italy
S P Raja, India
Dr.Anand Singh Rajawat , India
Simone Ranaldi , Italy
Upaka Rathnayake, Sri Lanka
Navid Razmjoo, Iran
Carlo Ricciardi, Italy
Jatinderkumar R. Saini , India
Sandhya Samarasinghe , New Zealand
Friedhelm Schwenker, Germany
Mijanur Rahaman Seikh, India
Tapan Senapati , China
Mohammed Shuaib , Malaysia
Kamran Siddique , USA
Gaurav Singal, India
Akansha Singh , India
Chiranjibi Sitaula , Australia
Neelakandan Subramani, India
Le Sun, China
Rawia Tahrir , Iraq
Binhua Tang , China
Carlos M. Travieso-González , Spain
Vinh Truong Hoang , Vietnam
Fath U Min Ullah , Republic of Korea
Pablo Varona , Spain
Roberto A. Vazquez , Mexico
Mario Versaci, Italy
Gennaro Vessio , Italy
Ivan Volosyak , Germany
Leyi Wei , China
Jianghui Wen, China
Lingwei Xu , China
Cornelio Yáñez-Márquez, Mexico
Zaher Mundher Yaseen, Iraq
Yugen Yi , China
Qiangqiang Yuan , China
Miaolei Zhou , China
Michal Zochowski, USA
Rodolfo Zunino, Italy


Contents

An Improved New Caledonian Crow Learning Algorithm for Global Function Optimization

Yanjiao Wang , Jiayu Song , and Ziming Teng 


Research Article (34 pages), Article ID 9248771, Volume 2022 (2022)

Dynamic Optimization Modeling of Smart Tourism Information System Using VRGIS in Big Data Environment

Anfeng Xu  and Wenjun Zeng



Research Article (8 pages), Article ID 7914674, Volume 2022 (2022)

A Novel Affective Analysis System Modeling Method Integrating Affective Cognitive Model and Bi-LSTM Neural Network

Zhuqing Yang , Liya Zhou, and Zhengjun Jing



Research Article (11 pages), Article ID 1856496, Volume 2022 (2022)

Comparative Analysis of the Performance of Complex Texture Clustering Driven by Computational Intelligence Methods Using Multiple Clustering Models

Jincheng Zhou, Dan Wang , Lei Ling, Mingjiang Li, and Khin-Wee Lai 

Research Article (13 pages), Article ID 8449491, Volume 2022 (2022)

A Mental Health Management and Cognitive Behavior Analysis Model of College Students Using Multi-View Clustering Analysis Algorithm

Danhui Dong  and Xiaoying Shen 

Research Article (10 pages), Article ID 2813473, Volume 2022 (2022)

A Comparative Optimization Model of Japanese Literature Characteristics for Cognitive Retrieval of Cross-Language Information

Tao Feng 

Research Article (10 pages), Article ID 8195075, Volume 2022 (2022)

Distributed Scheduling Strategy of Virtual Power Plant Using the Particle Swarm Optimization Neural Network under Blockchain Background

Changchang Lu  and Sijie Chen



Research Article (10 pages), Article ID 3222249, Volume 2022 (2022)

Comparison and Suggestions of Logistics Performance Index of Main Countries of Belt and Road Strategy Based on Bootstrap DEA Model

Wen-Tsao Pan , Bingqian Jiang , Yuting Wang , Yueyuan Cai , and Xiaoxia Ji 



Research Article (9 pages), Article ID 2159578, Volume 2022 (2022)

Font Design Optimization Model for New Media Short Video Based on Virtual Reality Digital Processing Technology


Hongwei Hu  and Wen Yao Zhu 

Research Article (11 pages), Article ID 1026918, Volume 2022 (2022)

Comparison and Analysis of Several Clustering Algorithms for Pavement Crack Segmentation Guided by Computational Intelligence

Dan Wang , Zaijun Zhang , Jincheng Zhou, Benfei Zhang, and Mingjiang Li
Research Article (13 pages), Article ID 8965842, Volume 2022 (2022)

Analysis of Balance of Income and Expenditure and Optimal Retirement Age of Pension Insurance Co-Ordination Account Based on Improved Machine Learning Algorithm

Shi Yan, Yaodong Zhou, and Youlu Zhang 
Research Article (11 pages), Article ID 5870893, Volume 2022 (2022)


A Novel Adaptive Affective Cognition Analysis Model for College Students Using a Deep Convolution Neural Network and Deep Features

Huali Feng 
Research Article (9 pages), Article ID 2114114, Volume 2022 (2022)

Application of Cognitive System Model and Gestalt Psychology in Residential Healthy Environment Design

Jicheng Yang  and Chao Yuan 
Research Article (8 pages), Article ID 5661221, Volume 2022 (2022)


Optimization of Railway Mixed Goods Loading Layout considering Stability

Juan Wang, Yinghua Yao, Yinggui Zhang , and Ximing Wang
Research Article (11 pages), Article ID 1517280, Volume 2022 (2022)

Concrete Surface Crack Recognition Based on Coordinate Attention Neural Networks

Yuhao Zhang and Zhongwei Wang 
Research Article (8 pages), Article ID 7454746, Volume 2022 (2022)




Integrated Prediction Framework for Clinical Scores of Cognitive Functions in ESRD Patients

Yutao Zhang, Quan Sheng, Xidong Fu, Haifeng Shi, and Zhuqing Jiao 
Research Article (12 pages), Article ID 8124053, Volume 2022 (2022)

Multimodal Sentiment Analysis Based on Cross-Modal Attention and Gated Cyclic Hierarchical Fusion Networks

Zhibang Quan , Tao Sun , Mengli Su, and Jishu Wei
Research Article (12 pages), Article ID 4767437, Volume 2022 (2022)

A Survival Status Classification Model for Osteosarcoma Patients Based on E-CNN-SVM and Multisource Data Fusion

Qiang Zhang , Peng Peng , and Yi Gu 
Research Article (11 pages), Article ID 9464182, Volume 2022 (2022)

Contents

A Fast CS-Based Reconstruction Model with Total Variation Constraint for MRI Enhancement in K-Space Domain

Hongxuan Duan  and Xiaochang Lv

Research Article (11 pages), Article ID 9222958, Volume 2022 (2022)

Exploration of Emotion Perception in Serious Interactive Digital Narrative

Jie Zhang  and Yaqian Liu



Research Article (10 pages), Article ID 8160695, Volume 2022 (2022)

An Intelligent System for Detecting Abnormal Behavior in Students Based on the Human Skeleton and Deep Learning

Yourong Ding , Ke Bao , and Jianzhong Zhang


Research Article (11 pages), Article ID 3819409, Volume 2022 (2022)

Adaptive Weighted Strategy Based Integrated Surrogate Models for Multiobjective Evolutionary Algorithm

Ke Bao , Wei Fang, and Yourong Ding 



Research Article (16 pages), Article ID 5227975, Volume 2022 (2022)

A Novel AI-Based Visual Stimuli Generation Approach for Environment Concept Design

Yingjing Duan and Jie Zhang 


Research Article (12 pages), Article ID 8015492, Volume 2022 (2022)

Quality Risk Management Algorithm for Cold Storage Construction Based on Bayesian Networks

Yaping Song  and Zhanguo Wei 




Research Article (12 pages), Article ID 6830090, Volume 2022 (2022)

Research on Embedded Multifunctional Data Mining Technology Based on Granular Computing

Juan Li  and Xianghong Tian



Research Article (7 pages), Article ID 4825079, Volume 2022 (2022)

Knowledge Graph-Enabled Text-Based Automatic Personality Prediction

Majid Ramezani , Mohammad-Reza Feizi-Derakhshi , and Mohammad-Ali Balafar 



Research Article (18 pages), Article ID 3732351, Volume 2022 (2022)

Optimal Control and Stability Analysis of an SEIR Model with Infectious Force in Latent Period

Li Jiayi , Li Sixian, Shi Weixuan, Hu Manfeng, and Zhang Jingxiang 


Research Article (9 pages), Article ID 7596421, Volume 2022 (2022)

Residual Life Prediction of Lithium Batteries Based on Data Mining

Dandan Ma  and Xiangge Qin 


Research Article (8 pages), Article ID 4520160, Volume 2022 (2022)

A Novel Stress State Assessment Method for College Students Based on EEG


Li Liu , Yunfeng Ji, Yun Gao, Tao Li, and Wei Xu

Research Article (11 pages), Article ID 4565968, Volume 2022 (2022)

Assessing the Impact of Continuous Vaccination and Voluntary Isolation on the Dynamics of COVID-19: A Mathematical Optimal Control of SEIR Epidemic Model

Yue Yu, Manman Shi, Manfeng Hu, and Jingxiang Zhang 
Research Article (13 pages), Article ID 3309420, Volume 2022 (2022)

Combining BERT Model with Semi-Supervised Incremental Learning for Heterogeneous Knowledge Fusion of High-Speed Railway On-Board System

Lu-jie Zhou , Zhi-peng Zhao, and Jian-wu Dang
Research Article (15 pages), Article ID 9948218, Volume 2022 (2022)


Dance Performance in New Rural Areas Based on 3D Image Reconstruction Technology

Li Xie 
Research Article (7 pages), Article ID 7122053, Volume 2022 (2022)



Hotspots and Cutting-Edge Visual Analysis of Digital Museum in China Using Data Mining Technology

Yajing Hou , Lijun Xu , and Lu Chen
Research Article (16 pages), Article ID 7702098, Volume 2022 (2022)

Design of Nuclear Radiation Monitoring System in Floor Exploration Based on Deep Learning

Bochen Zong 
Research Article (10 pages), Article ID 4351339, Volume 2022 (2022)



Machine Learning Application of Transcranial Motor-Evoked Potential to Predict Positive Functional Outcomes of Patients

Mohd Redzuan Jamaludin , Khin Wee Lai , Joon Huang Chuah , Muhammad Afiq Zaki ,
Khairunnisa Hasikin , Nasrul Anuar Abd Razak , Samiappan Dhanalakshmi , Lim Beng Saw , and
Xiang Wu 
Research Article (13 pages), Article ID 2801663, Volume 2022 (2022)


Greedy Strategies with Multiobjective Optimization for Investment Portfolio Problem Modeling

Xinchen Zhang , Linghao Zhang , Qincheng Zhou, and Xu Jin 
Research Article (12 pages), Article ID 4862772, Volume 2022 (2022)


Design and Implementation of Chinese Common Braille Translation System Integrating Braille Word Segmentation and Concatenation Rules

Ju-Xiao Zhang , Hai-Feng Chen, Bing Chen, Bei-Qin Chen , Jing-Hua Zhong, and Xiao-Qin Zeng
Research Article (14 pages), Article ID 8934241, Volume 2022 (2022)

A Novel Deep Learning Network and Its Application for Pulmonary Nodule Segmentation


Dechuan Lu, Junfeng Chu, Rongrong Zhao, Yuanpeng Zhang, and Guangyu Tian 
Research Article (6 pages), Article ID 7124902, Volume 2022 (2022)

Cost Early-Warning Model System of Large-Scale Construction Project

Jingyi Dai and Dandan Ke 
Research Article (10 pages), Article ID 3541803, Volume 2022 (2022)

Contents

Interactive Design of Personalized Website Search Interface Based on Visual Communication

Zhen Xu  and Shan Wang


Research Article (11 pages), Article ID 2125506, Volume 2022 (2022)

A Novel Bitcoin and Gold Prices Prediction Method Using an LSTM-P Neural Network Model

Xinchen Zhang , Linghao Zhang , Qincheng Zhou, and Xu Jin 



Research Article (12 pages), Article ID 1643413, Volume 2022 (2022)

A Novel Sentiment Analysis Model of Museum User Experience Evaluation Data Based on Unbalanced Data Analysis Technology

Xiang Chen , Zhiwei Chen, Lei Xiao, and Ming Zhou

Research Article (10 pages), Article ID 2096634, Volume 2022 (2022)

A Dynamic Model of Evolutionary Knowledge and Capabilities Based on Human-Machine Interaction in Smart Manufactures

Shuxian Chen, Zongqiang Ren, Xikai Yu , and Ao Huang 


Research Article (10 pages), Article ID 8584888, Volume 2022 (2022)

Research on Rice Yield Prediction Model Based on Deep Learning

Xiao Han , Fangbiao Liu, Xiaoliang He, and Fenglou Ling

Research Article (9 pages), Article ID 1922561, Volume 2022 (2022)

Research on Impulse Power Load Forecasting Based on Improved Recurrent Neural Networks

Chenyang Feng , Kang Xu, and Haoyun Ma


Research Article (9 pages), Article ID 2784563, Volume 2022 (2022)

TransEffiDet: Aircraft Detection and Classification in Aerial Images Based on EfficientDet and Transformer

Yanfeng Wang, Tao Wang, Xin Zhou , Weiwei Cai , Runmin Liu , Meigen Huang, Tian Jing, Mu Lin, Hua He, Weiping Wang, and Yifan Zhu


Research Article (10 pages), Article ID 2262549, Volume 2022 (2022)

Research on the Transformation from Financial Accounting to Management Accounting Based on Drools Rule Engine

Rui Liu, Yuqin Wang , and Jing Zou




Research Article (8 pages), Article ID 9445776, Volume 2022 (2022)

Research on Intelligent Target Tracking Algorithm Based on MDNet under Artificial Intelligence

Yu Wang 



Research Article (9 pages), Article ID 1550543, Volume 2022 (2022)

Aircraft Image Recognition Network Based on Hybrid Attention Mechanism

Yanfeng Wang , Yinan Chen , and Runmin Liu 

Research Article (9 pages), Article ID 4189500, Volume 2022 (2022)

DCCAM-MRNet: Mixed Residual Connection Network with Dilated Convolution and Coordinate Attention Mechanism for Tomato Disease Identification

Yujian Liu, Yaowen Hu, Weiwei Cai , Guoxiong Zhou , Jialei Zhan, and Liujun Li
Research Article (15 pages), Article ID 4848425, Volume 2022 (2022)


Customer Relationship Management Based on SPRINT Classification Algorithm under Data Mining Technology

Yazhou Sun and Xueqing Tan 
Research Article (11 pages), Article ID 6170335, Volume 2022 (2022)

Sentimental Analysis of Twitter Users from Turkish Content with Natural Language Processing

Cagla Balli , Mehmet Serdar Guzel , Erkan Bostanci , and Alok Mishra 
Research Article (17 pages), Article ID 2455160, Volume 2022 (2022)

Research on the Design of Intelligent Music Teaching System Based on Virtual Reality Technology

Wei Chen 
Research Article (9 pages), Article ID 7832306, Volume 2022 (2022)

Research Article

An Improved New Caledonian Crow Learning Algorithm for Global Function Optimization

Yanjiao Wang ¹, Jiaxu Song ¹, and Ziming Teng ²

¹Department of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China

²Department of Communication Engineering, Jilin University, Jilin 130012, China

Correspondence should be addressed to Jiaxu Song; songjiaxu1996@163.com

Received 29 June 2022; Revised 31 July 2022; Accepted 5 August 2022; Published 10 October 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Yanjiao 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.

The New Caledonian crow learning algorithm (NCCLA) is a novel metaheuristic algorithm inspired by the learning behavior of New Caledonian crows learning to make tools to obtain food. However, it suffers from the problems of easily falling into local optima and insufficient convergence accuracy and convergence precision. To further improve the convergence performance of NCCLA, an improved New Caledonian crow learning algorithm (INCCLA) is proposed in this paper. By determining the parent individuals based on the cosine similarity, the juveniles are guided to search toward different ranges to maintain the population diversity; a novel hybrid mechanism of complete and incomplete learning is proposed to balance the exploration and exploitation capabilities of the algorithm; the update strategy of juveniles and parent individuals is improved to enhance the convergence speed and precision of the algorithm. The test results of the CEC2013 and CEC2020 test suites show that, compared with the original NCCLA algorithm and four of the best metaheuristics to date, INCCLA has significant advantages in terms of convergence speed, convergence precision, and stability.

1. Introduction

A large number of optimization problems exist in real life, engineering design, computer technology and other fields, such as minimum cost, optimal parameters, minimum time, pipeline route design, welded beam design, etc. In order to obtain higher economic efficiency and social value, scholars strive to obtain the optimal solution to optimization problems, thus making the research of optimization methods widely concerned.

Optimization methods usually include traditional optimization methods such as the fastest descent method and metaheuristic methods. Among them, traditional optimization methods usually require the optimization problem to be derivable, and the convergence speed and convergence precision are difficult to meet the practical needs when the optimization problem has multiple extrema. And the metaheuristic algorithms proposed by simulating biological habits in nature, etc., have good exploration and exploitation performance through information exchange among

individuals in the population. Compared with traditional optimization algorithms, metaheuristic algorithms are better in convergence speed, convergence precision, robustness, and stability, with no strict requirements on the form of optimization problems. This makes the metaheuristic algorithm the most effective and widely used optimization method. Researchers have focused on two main aspects of metaheuristics to obtain the best optimization results: improving existing metaheuristics and proposing new ones.

Various metaheuristic algorithms have been proposed, and the more representative methods are as follows: the artificial bee colony algorithm (ABC) was proposed to simulate the behavior of bee colonies to find the optimal nectar source according to different internal divisions of labor [1]; the crow search algorithm (CSA) was proposed to simulate the behavior of crows to hide and search for food [2]; particle swarm optimization (PSO) algorithm was proposed to simulate the foraging behavior of a flock of birds [3]; the firefly algorithm (FA) was proposed to simulate the behavior of fireflies to attract each other [4]; the Marine

Predator Algorithm (MPA) was proposed to simulate the behavior of biological interactions between marine predators and prey [5]; the Manta Ray Foraging Optimization Algorithm (MRFO) was proposed to simulate three unique foraging modes of manta rays: chain foraging, cyclone foraging, and somersault foraging [6]; the dolphin swarm optimization algorithm (DSA) was proposed to simulate the habits of dolphins such as echolocation and information exchange [7]; and the gray wolf optimization algorithm (GWO) was proposed to simulate the hunting behavior of gray wolves [8] and so on. Some of the algorithms mentioned above have been applied to many research articles, such as applying particle swarm optimization to solve the UCP problem with deterministic and stochastic load demands [9], the ocean predator algorithm to solve the optimal reactive power dispatch (ORPD) problem [10], and the manta ray foraging optimization algorithm to solve the economic load dispatch and advance dispatches problems of microgrids [11].

Researchers have done a lot of work on the improvement of existing metaheuristic algorithms, and the more representative research results are given below. Many improvement methods for particle swarm algorithms have been proposed in recent years, and the representative results are as follows: In 2016, Samma et al. proposed the RLMPSO algorithm [12], where each particle performs five operations under the control of the RL algorithm to improve the search performance of the particle swarm algorithm. In 2018, Zhang et al. proposed the DLPSO algorithm [13] to address the shortcomings of the PSO algorithm in multimodal indivisible problems that tend to fall into local optima, which extracts good vectors from the vectors distributed in the search space to form a new vector with a greater possibility of jumping out of local optima. In 2021, Lu et al. proposed the EMCPSO algorithm [14], which takes advantage of the technology of multiple populations in order to overcome the problem of premature convergence of PSO, which divides the population into four identical subpopulations, and the optimal individual in each subpopulation is used to represent the evolutionary state of that subpopulation, and by sharing information among the four populations, the evolutionarily stagnant subpopulations search for the optimal solution again, while introducing an exclusion mechanism to prevent premature convergence of particles further. In 2022, Wang et al. proposed the RLLPSO algorithm for large-scale optimization problems [15], which constructs a level-based population structure to improve population diversity, a reinforcement learning strategy as well as a level competition strategy to improve the search efficiency of the algorithm in order to overcome the complexity of large-scale optimization problems.

Many improvement methods have been proposed for the artificial bee colony algorithm, and the representative results are as follows: In 2018, Cui et al. proposed the DPABC algorithm [16], which uses a dual population framework to divide the population into a convergence population and a diversity population, responsible for developing promising regions as well as maintaining population diversity, respectively, to improve the overall performance of the

algorithm. In 2019, Awadallah et al. improved the onlooker bee stage [17], combining four selection methods, including global optimum, tournament, linear ranking, and exponential ranking, to guide the search process of the onlooker bee in order to determine the impact of the selection scheme on the onlooker bee stage. In 2021, Zhou et al. proposed the ABC-MNT algorithm [18], which applies three different neighborhoods to different individuals, helping the algorithm to achieve a better balance between exploration and exploitation, in addition to employing a global neighborhood search strategy and opposition-based learning that preserves the search experience of the scout bee phase. In 2022, Ye et al. proposed the RNSABC algorithm [19], which uses a random neighborhood structure so that each solution has a random neighborhood, in addition to a depth-first search method to enhance the search capability of the following bee to improve the algorithm's ability to search for the optimal solution.

Representative results of the improvement of other mainstream metaheuristic algorithms are as follows: In 2017, Wang et al. proposed a firefly algorithm with neighborhood attractiveness (NaFA) [20], where each firefly selects attractive individuals from a predefined region instead of the whole population, and the proposed strategy can effectively improve the solution accuracy and reduce the time complexity. In 2018, Sun et al. proposed an improved whale optimization algorithm (MWOA) for solving large-scale optimization problems [21], which uses a nonlinear dynamic strategy based on the cosine function to update the control parameters, balances the exploration and exploitation capabilities of the algorithm, uses a Levy flight strategy to make the algorithm jump out of the local optimum, and uses quadratic interpolation for the optimal individuals of the population to enhance the local exploitation capabilities of the algorithm. In 2019, Zamani et al. proposed a conscious neighborhood-based crow search algorithm (CCSA) [22], which introduces three search strategies, neighborhood-based local search strategy, non-neighborhood global search strategy, and roaming-based search strategy, to enhance the balance between local and global search. In 2020, Gupta et al. proposed a memory-based gray wolf optimization algorithm (mGWO) [23], which modified crossover and greedy selection based on the historical optimum of individuals, enhancing the algorithm's ability to perform the global search, local exploitation, and the balance between the two. In 2022, Long et al. proposed a velocity-based butterfly optimization algorithm (VBOA) [24], which introduced velocity and memory to guide individuals in the local search phase and introduced a refraction-based learning strategy, effectively enhancing the diversity of populations and the exploration ability of the algorithm.

Many excellent metaheuristic algorithms have been proposed in recent years. In 2018, Wang proposed the moth search algorithm (MSA) [25] inspired by the phototropism of moths and Levy flight, which treats moths as individuals. Moths with smaller distances from the optimal individual perform Levy flight, while moths with more considerable distances approach the optimal individual in a straight line. The above two stages optimize the algorithm. In 2019,

Heidari et al. proposed the Harris Hawk optimization algorithm (HHO) [26] based on the inspiration of the collaborative group behavior of Harris hawks during predation, which uses Harris hawks as individuals, and the search process includes three stages: exploration, exploration to exploitation conversion, and exploitation, and the algorithm is characterized by few control parameters and excellent global search capability. In 2020, Li et al. proposed the slime mold algorithm (SMA) [27], inspired by the behavioral and morphological changes in *Physarum polycephalum* during foraging, which creates three different forms to optimize the problem by using weights to simulate the positive and negative feedback generated by slime molds during foraging. In the same year, Al-Sorori and Mohsen proposed the New Caledonian crow learning algorithm (NCCLA) [28] based on the behavior of New Caledonian crows to obtain food by learning to make tools. The advantage of this algorithm is its stochastic nature, which guarantees that the algorithm does not get trapped at the local optimum. In the same year, Mohamed et al. proposed a gaining-sharing knowledge-based algorithm (GSK) [29] inspired by the process of acquiring and sharing knowledge in the human life cycle, which treats people as individuals and improves their knowledge by using junior gaining and sharing phase and senior gaining and sharing phase, i.e., solving optimization problems on continuous space. In 2021, Tu et al. proposed the colony predation algorithm (CPA) [30], inspired by the supportive behavior of herd animals and the behavior of selective hunting, which is based on the coexistence of social animals and focuses on optimizing the problem through five stages: communicating and collaborating, dispersing food, surrounding food, supporting the closest individual, and finding food. In 2022, Hashim et al. proposed the snake optimizer (SO) [31], based on the behavior of snakes to forage or breed under different temperature and food availability conditions, in which individuals explore and exploit the conditions of temperature as well as food.

The various metaheuristic algorithms mentioned above provide new ideas for solving optimization problems and further advance the development of optimization techniques. Compared with the more classical PSO and DE, they have significantly improved in terms of convergence speed and convergence precision. However, unfortunately, for the highly nonlinear and complex optimization problems that emerge one after another in practical engineering, the convergence speed and convergence precision of the existing metaheuristic algorithms are obviously insufficient and even fall into local optimum, making it difficult to obtain highly satisfactory economic and social values. Therefore, improving the optimization performance of each new metaheuristic algorithm has been one of the main research contents in the field of evolution.

In this context, given the literature [28] and a large number of experimental studies, the NCCLA algorithm is a very excellent metaheuristic algorithm because of its simple operation and significantly better convergence capability than optimization algorithms such as GWO, CSA, and WOA, and is highly promising in fields such as engineering optimization. In this paper, we only study NCCLA. In order

to further improve the problems of insufficient convergence precision and convergence speed and easy falling into local optimum when NCCLA deals with very complex optimization problems, we propose an improved New Caledonian Crow Learning Algorithm (INCCLA) in this paper.

The main innovations and contributions of INCCLA are as follows: (1) A cosine similarity-based parent individual selection approach is proposed. The globally optimal individual and another excellent individual with a significant difference in similarity are selected as the parent, and the juvenile crow individuals are guided to search toward different ranges to maintain the population diversity while maintaining the convergence speed of the algorithm. (2) Improving the learning phase of juvenile crows. A new hybrid learning mechanism of complete learning and incomplete learning is set up, in which the individual juvenile crows in the complete learning stage can select learning objects according to their own conditions, which effectively improves the convergence speed of the algorithm while maintaining the population diversity to a certain extent; while the juvenile crows in the incomplete learning stage learn the behavioral attributes of different individuals in order to maintain the population diversity of the algorithm. (3) Improving the reinforcement phase. For the juvenile reinforcement stage, a weighting factor is introduced to enable the algorithm to have a strong exploration ability in the early evolutionary stage and a strong exploitation ability in the late evolutionary stage, and at the same time, a small range of random perturbations is added to increase the possibility of convergence of the algorithm to the global optimum; for the parents' reinforcement stage, the update methods of the two parents' individuals are improved further to respectively balance the exploration and exploitation ability of the algorithm. The results of testing on the CEC2013 and CEC2020 test suites show that the INCCLA proposed in this paper has significant advantages in terms of convergence speed, convergence precision, and stability compared with four other more representative optimization algorithms.

The rest of the paper is organized as follows: Section 2 describes the working principle and flow of the NCCLA algorithm. Section 3 analyzes the defects of the original NCCLA algorithm and further proposes an improved INCCLA algorithm. Section 4 shows the simulation results and analysis of the INCCLA algorithm with the original NCCLA algorithm and other more mainstream improved algorithms on the CEC2013 and CEC2020 test function suites. Section 5 concludes the proposed algorithm in this paper.

2. New Caledonian Crow Learning Algorithm

In nature, New Caledonian crows are divided into the juvenile and the parent crows, which enhance their tool-design skills through learning and their own experience and knowledge, respectively, to obtain food from the pandanus tree. Inspired by the above behavior, Wedad and Abdulqader proposed the New Caledonian crow learning algorithm (NCCLA). In NCCLA, individuals represent the

manufacturing behavior of New Caledonian crows and fitness values represent the behavioral advantage of each crow. The algorithm has three main phases: initialization, learning phase, and reinforcement phase. The pseudo-code of NCCLA is shown in Algorithm 1, and the key steps are briefly described as follows.

2.1. Population Initialization. Suppose the number of individuals in population X is N . Each individual X_i ($i = 1, 2, \dots, N$) represents the behavior of a crow, which can be expressed as $X_i = [X_{i,1}, X_{i,2}, X_{i,3}, \dots, X_{i,D}]$, where D is the dimension of the optimization problem, and $X_{i,j}$ denotes the j -th behavioral attribute of the i -th crow. At the beginning of the algorithm, the initial behavior of each crow is generated randomly according to Equation.

$$X_{i,j}(0) = X_L + U(0, 1) \times (X_U - X_L), \quad (1)$$

where X_U and X_L correspond to the upper and lower bounds of the j -th dimensional search space in the optimization problem, respectively, and $U(0,1)$ is a uniformly distributed random number in the range $[0,1]$.

2.2. Learning Phase. In NCCLA, only juveniles enter the learning phase, and each behavioral attribute $X_{i,j}$ of juveniles will be socially or asocially learned according to the probability SL_{prob} or $1 - SL_{\text{prob}}$, respectively. SL_{prob} is recommended to be set to 0.95, but can be set to other values.

2.2.1. Social Learning. After the j -th behavioral attribute of juvenile crow i , $X_{i,j}$, is determined to require social learning according to the probability SL_{prob} , it is then decided to perform vertical learning or horizontal learning according to the predetermined probability VSL_{prob} or $1 - VSL_{\text{prob}}$. The details are shown in Equation.

$$X_{i,j}(t) = \begin{cases} X_{k_1,j}(t-1), & \text{if } \text{rand} \leq VSL_{\text{prob}}, \\ X_{k_2,j}(t-1), & \text{else,} \end{cases} \quad (2)$$

where $X_{i,j}(t)$ is the j -th new behavioral attribute acquired by juvenile crow i after social learning in iteration t , VSL_{prob} is recommended to be set to 0.99, and can also be set to other values.

In (2), when $\text{rand} \leq VSL_{\text{prob}}$, the juvenile crow X_i performs vertical learning to its parent, i.e., it copies the corresponding behavioral attributes of its parent X_{p1} or X_{p2} with probability $P1_{\text{prob}}$, obviously $k_1 = 1$ or 2 ; Otherwise, the juvenile crow X_i performs horizontal learning, i.e., it randomly selects a sibling k_2 that is more experienced and copies its corresponding behavioral attributes, and the expression formula for k_2 is shown in equation (3). It should be noted that for the juvenile crow with the best fitness value, only vertical learning is performed, not horizontal learning.

$$k_2 = 3 + [\text{rand} \times (i - 3)], \quad (3)$$

where rand is a random number uniformly distributed in the range $[0, 1]$, $[\cdot]$ means rounding is performed.

2.2.2. Asocial Learning. When a crow X_i performs asocial learning, its behavioral attributes are randomly updated using (1) according to the probability TaE_{prob} , or retained the previous behavioral attributes according to $1 - TaE_{\text{prob}}$. This is shown in Equation .

$$X_{i,j}(t) = \begin{cases} X_L + U(0, 1) \times (X_U - X_L), & \text{rand} \leq TaE_{\text{prob}}, \\ X_{i,j}(t-1), & \text{else,} \end{cases} \quad (4)$$

where TaE_{prob} is recommended to be set to 0.99, but can also be set to other values.

2.3. Reinforcement Phase. After completion of the learning phase, certain attributes of the learned juvenile crow behavior and parent behavior are reinforced according to the reinforcement probability RP_{prob} . RP_{prob} is recommended to be set to 0.99, and can be set to other values.

2.3.1. Juvenile Reinforcement. Each behavioral attribute of juvenile crows was reinforced according to Equation.

$$X_{i,j}(t) = X_{i,j}(t) \pm RW, \quad (5)$$

where RW is shown in Equation.

$$RW = \begin{cases} \beta - \alpha, & i < N, \\ r1 \times ((r2 \times \beta) - \alpha), & \text{otherwise,} \end{cases} \quad (6)$$

where $r1$ and $r2$ are random numbers between 0 and 1, α represents the difference between their behavioral attributes before and after learning, as shown in (7), and β represents the social learning effect developed over time, as shown in Equation (8).

$$\alpha = |X_{i,j}(t) - X_{i,j}(t-1)|, \quad (7)$$

$$\beta = X_{i,j}(t-1) \times e^{-lf \times r \times t \times \text{mean}(j)}, \quad (8)$$

where t represents the number of current iterations, $X_{i,j}(t-1)$ is the j -th behavioral attribute of crow i before the current generation of learning, r is a normally distributed random number in the range $[0,1]$, $\text{mean}(j)$ is the average of the j -th behavioral attribute of all individuals in the population, and lf is a learning factor, as shown in Equation .

$$lf = lf_{\min} + \left(\frac{lf_{\max} - lf_{\min}}{\max_t} \right) \times t, \quad (9)$$

where \max_t represents the maximum number of iterations, lf_{\max} and lf_{\min} represent the maximum and minimum values of the learning factor, respectively, and are recommended to be set to 0.02 and 0.0005, respectively, or can be set by oneself.

2.3.2. Parents Reinforcement. The parents X_{p1} and X_{p2} update certain attributes according to the reinforcement probability RP_{prob} . When $\text{rand} \leq RP_{\text{prob}}$, the j -th dimensional behavioral attribute of crow i is reinforced, as shown in (10); otherwise, it retains the original behavioral attribute.

```

Input:  $N, D, RP_{\text{prob}}, SL_{\text{prob}}, VSL_{\text{prob}}, P1_{\text{prob}}, TaE_{\text{prob}}, \text{MaxIter}$ 
Output: best solution and its fitness
(1) Initialization of variables ( $N, D, RP_{\text{prob}}, SL_{\text{prob}}, VSL_{\text{prob}}, P1_{\text{prob}}, TaE_{\text{prob}}, \text{MaxIter}$ )
(2) Generate the initial population  $X$  according to Section 2.1
(3) Calculate the fitness value  $F_i$  of each individual  $X_i$ 
(4) While  $t \leq \text{MaxIter}$  do
(5) Rank all individuals and select the two best individuals,  $X_1$  and  $X_2$ , as parents, noted as  $X_{p1}$  and  $X_{p2}$ , and the rest as juvenile crows
(6) For Each juvenile crow  $X_i$  in population  $X$  do
(7) For Each behavior attribute  $j$  in  $X_i$  do
//Learning phase
(8) If  $\text{rand} \leq SL_{\text{prob}}$  then
(9) Decide to perform vertical or horizontal learning according to probability  $VSL_{\text{prob}}$  in Section 2.2.1//Social learning
(10) else
(11) Random update of behavior attributes or retention of previous behavior attributes using (1) as decided in Section 2.2.2 based on the probability  $TaE_{\text{prob}}$ //Asocial learning
(12) end If
(13) Based on the probability  $RP_{\text{prob}}$ , reinforce the attributes of  $X_i$  after learning using (4)//Juvenile reinforcement
(14) end For
(15) end For
(16) Based on the probability  $RP_{\text{prob}}$ , reinforce the attributes of the parents  $X_{p1}$  and  $X_{p2}$  using (9)//Parents reinforcement
(17)  $t = t + 1$ 
(18) end While
(19) Output the global optimal solution

```

ALGORITHM 1: NCCLA.

$$X_{i,j}(t) = \begin{cases} X_{i,j}(t-1) - \left(X_{1,j}(t-1) + e^{r1 \times (\text{mean}(j) - X_{i,j}(t-1))} \right), & i = p1, \\ X_{i,j}(t-1) - \left(r2 \times \left(X_{1,j}(t-1) - e^{r1 \times (\text{mean}(j) - X_{i,j}(t-1))} \right) \right), & i = p2, \end{cases} \quad (10)$$

where $r1$ is a normally distributed random number, $r2$ is a uniformly distributed random number in the range $[0, 1]$, and $\text{mean}(j)$ is the mean of the j -th behavioral attribute of all individuals in the current population.

3. Proposed Algorithm

To further improve the convergence performance of NCCLA, this section proposes an improved New Caledonian Crow Learning Algorithm (INCCLA), whose pseudo-code is shown in Algorithm 2.

3.1. Determination of Parent Individuals Based on Cosine Similarity. In NCCLA, VSL_{prob} and $P1_{\text{prob}}$ are set to 0.99 and 0.95, respectively, meaning that each juvenile crow will perform vertical learning with a probability of 0.99×0.95 toward the parent individual during the learning phase. The parent individuals are always the two individuals with the best fitness in the population. As evolution proceeds, the two-parent individuals will rapidly approach each other, showing a high degree of similarity, which will lead most of

the juvenile crows to approach them through vertical learning rapidly, and can only search around the parent individuals, lacking exploration of other ranges. Although rapid convergence can be achieved in the early stage of evolution, the loss of population diversity is apparent, and the algorithm is straightforward to fall into the local optimum.

In order to solve the above problem, the following cosine similarity-based parent individual selection method is proposed in this section. First, the best individual in the population is determined as the parent individual X_{p1} ; then, the cosine similarity of the remaining individuals in the population to X_{p1} is calculated according to (11), and they are arranged in order from smallest to largest and evenly divided into two groups; finally, the individual with the best fitness value is selected as the parent individual X_{p2} from the other group different from the group in which X_{p1} is located. It is important to note that the parent individuals are selected for every P generation above. Generally, $P=50$ is sufficient to achieve good results, but it can also be set according to the optimization problem.

```

Input:  $N, D, RP_{\text{prob}}, SL_{\text{prob}}, \text{MaxIter}$ 
Output: best solution and its fitness
(1) Initialization of variables ( $N, D, RP_{\text{prob}}, SL_{\text{prob}}, \text{MaxIter}$ )
(2) Generate the initial population  $X$  according to Section 2.1
(3) Calculate the fitness value  $F_i$  of each individual  $X_i$ 
(4) While  $t \leq \text{MaxIter}$  do
(5) Rank all individuals and select the parents according to Section 3.1, noted as  $X_{p1}$  and  $X_{p2}$ , and the rest as juvenile crows
(6) Randomly select  $R$  individual juvenile crows to form a subpop
(7) For Each juvenile crow  $X_i$  in population  $X$  do
(8)   if  $X_i \in \text{subpop}$  do
(9)     Perform complete learning of the juvenile crow  $X_i$  according to Section 3.2.1
(10)   else
(11)     Perform incomplete learning of the juvenile crow  $X_i$  according to Section 3.2.2
(12)   end if
(13)   operations on juvenile crows  $X_i$  according to Section 3.3.1
(14) end for
(15) Perform parent reinforcement operations on parents  $X_{p1}$  and  $X_{p2}$  according to Section 3.3.2
(16)  $t = t + 1$ 
(17) end While
(18) Output the global optimal solution

```

ALGORITHM 2: INCCLA.

$$\text{sim}_{i,p1}(t-1) = \frac{\sum_{j=1}^D X_{p1,j}(t-1) \times X_{i,j}(t-1)}{\sqrt{\sum_{j=1}^D (X_{p1,j}(t-1))^2} \times \sqrt{\sum_{j=1}^D (X_{i,j}(t-1))^2}}, \quad i = 1, 2, \dots, N, \quad (11)$$

where $X_{p1,j}(t-1)$ denotes the j -th dimension of the globally optimal individual $X_{p1}(t-1)$ determined by relying on the previous iteration of the population, and $\text{sim}_{i,p1}(t-1)$ denotes the cosine similarity of the i -th individual $X_i(t-1)$ without any evolutionary operation to $X_{p1}(t-1)$ in the current generation.

To further illustrate the above method of determining parent individuals based on cosine similarity, taking the optimized 2-dimensional Rotated High Conditioned Elliptic function as an example, the specific determination process is given in Figure 1. It can be seen that, according to the original parent individual determination method, X_1 and X_2 , which are close to each other, will be selected as parent individuals, while according to the proposed method in this section, X_1 and X_4 , which are farther apart, will be selected as parent individuals. The fitness value of X_4 is not much worse than X_2 .

In summary, compared with the original approach of selecting two more similar optimal individuals as parents, this section retains the optimal individual of the population as a parent individual, which does not affect the convergence speed too much. In contrast, the other parent individual is selected to be less similar to the optimal individual of the population but with better fitness value. This way can guide juveniles in their search across a range and ensure that they learn from more experienced individuals, maintaining good population diversity while not reducing the convergence speed too much.

3.2. Improving Juvenile Crow Learning Phase. As seen in Algorithm 1, the learning phase generates its own learning objects for juvenile crows intending to provide excellent evolutionary directions for the reinforcement phase. In-depth analysis can be found that the various behavioral attributes of the learning objects generated in the learning phase are not almost wholly derived from the same individual. Although the population diversity can be better maintained, it is difficult to ensure the superiority of the learning objects composed of them due to the complete separation of each behavioral attribute. Therefore, it is difficult to guarantee the convergence speed of the reinforcement phase. The excellent individuals themselves are already better integrated with each behavioral attribute, which has a vital role in the rapid convergence of the algorithm but is not conducive to the maintenance of population diversity. Given this, this section proposes a novel juvenile learning approach as shown in Figure 2. R juveniles are randomly selected to perform complete learning, i.e., each behavioral attribute of the corresponding learning object originates from the same individual completely, which promotes fast convergence of the algorithm; while the remaining juveniles perform incomplete learning, i.e., each behavioral attribute of the learning object originates from different individuals, which ensures population diversity.

Individual	Fitness value	Cosine similarity
X1	7.024E+07	1
X2	8.349e+07	0.9994
X3	9.407e+07	0.9995
X4	7.106e+08	-0.4944
X5	9.451e+08	0.6486
X6	1.080e+09	-0.9981
X7	2.050e+09	-0.8219
X8	3.919e+09	-0.8888
X9	4.212e+09	0.4412
X10	4.394e+09	0.6834
X11	4.750e+09	-0.9542
X12	5.597e+09	-0.6412
X13	5.730e+09	-0.9144
X14	6.249e+09	0.6197
X15	1.051e+10	0.1554
X16	2.179e+10	0.0352

Sort the cosine similarity and divide it evenly into two groups

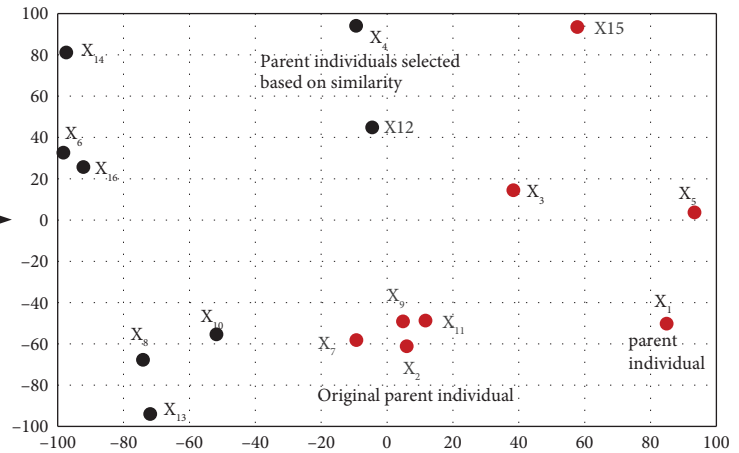


FIGURE 1: Schematic of the process of determining parent individuals based on cosine similarity.

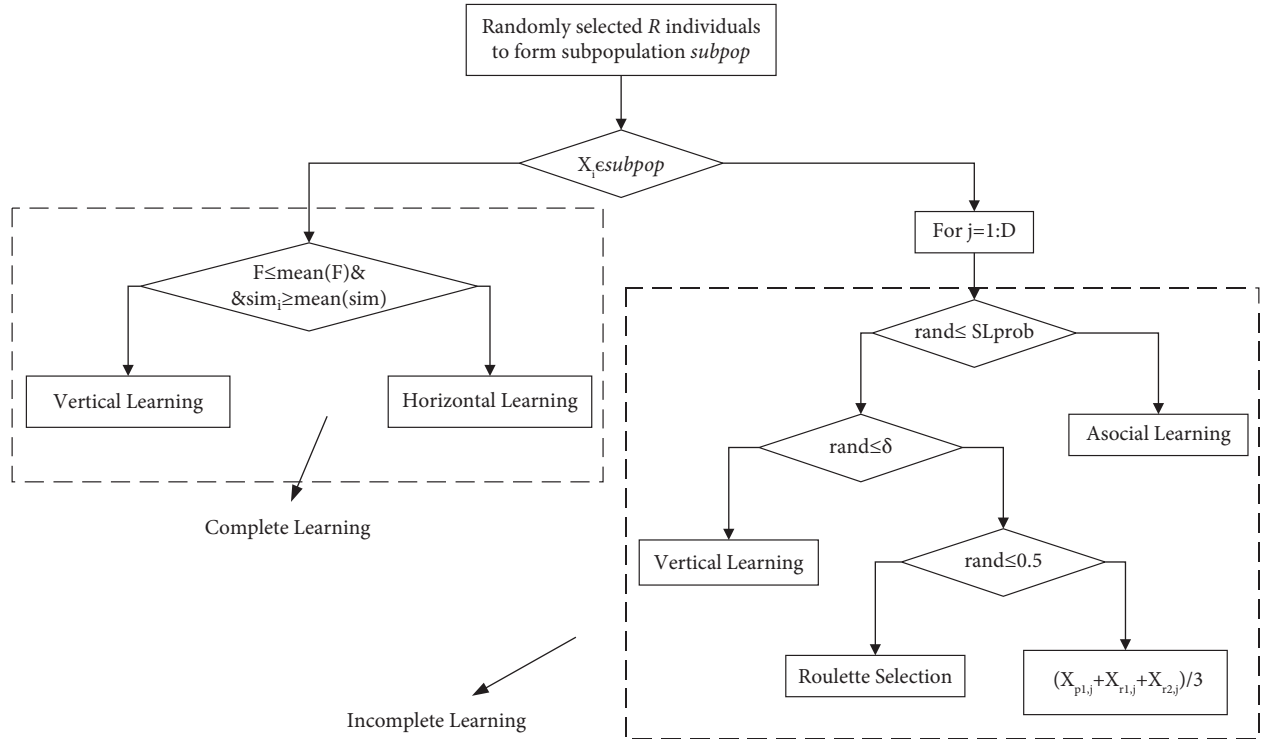


FIGURE 2: Schematic of juvenile crow learning phase.

3.2.1. Complete Learning. As mentioned above, the complete learning phase aims to enhance the convergence speed of the algorithm by copying all the behavioral attributes of a particular outstanding individual. In contrast to social learning, asocial learning focuses on maintaining population diversity; therefore, in the complete learning phase, asocial learning is eliminated, and only social learning is used.

Social learning in NCCLA includes vertical and horizontal learning, where vertical learning is learning from the two best parent individuals in the population, while horizontal learning is learning from other individuals who are

better than oneself. Obviously, compared with vertical learning, horizontal learning is more capable of maintaining population diversity. In NCCLA, vertical or horizontal learning is chosen according to the fixed probability VSL_{prob} . If VSL_{prob} is high, the population will quickly approach the parent individuals, accelerating the algorithm's convergence. However, the rapid loss of population diversity can easily cause the algorithm to fall into local optimum. If VSL_{prob} is low, each individual may have different learning objects, and introducing multiple learning objects makes the evolutionary direction of each individual more diffuse, which is

not conducive to the rapid convergence of the population. The NCCLA recommends that VSL_{prob} be set to 0.95, allowing all juveniles to primarily use vertical learning as their social learning method. In vertical learning, individuals need to choose to learn from the best parent X_{p1} or the second-best parent X_{p2} according to the fixed probability SL_{prob} . If SL_{prob} is large or small, all juvenile crows will search around a parent almost exclusively, which is not conducive to the maintenance of population diversity and increases the possibility of the algorithm falling into local optimum; If SL_{prob} is set to about 0.5, it will learn from the parent

individuals with equal chances, which is similar to random selection and has certain blindness, slowing down the convergence of the algorithm to some extent. In short, the selection methods of vertical learning or social learning by fixed probability and the selection of a certain parent individual for learning by fixed probability in vertical learning are not very reasonable.

Given this, to effectively improve the convergence speed of the algorithm and not destroy the population diversity too much, we propose the complete social learning method as shown in Equation.

$$X_i(t) = \begin{cases} X_{k_1}(t-1), & \text{if } F_i(t-1) \leq \bar{F}_i(t-1) \& \text{sim}_{i,p1}(t-1) \geq \overline{\text{sim}_{i,p1}}(t-1) \\ X_{k_2}(t-1), & \text{else} \end{cases}, \quad (12)$$

where $F_i(t-1)$ denotes the fitness value of juvenile X_i , $\bar{F}_i(t-1)$ denotes the mean of the fitness values of all individuals in the population, $\text{sim}_{i,p1}(t-1)$ denotes the cosine similarity of juvenile X_i to the best individual X_{p1} in the population. $\overline{\text{sim}_{i,p1}}(t-1)$ denotes the mean of the cosine similarity of all individuals in the population to X_{p1} . When both $F_i(t-1) \leq \bar{F}_i(t-1)$ and $\text{sim}_{i,p1}(t-1) \geq \overline{\text{sim}_{i,p1}}(t-1)$ are satisfied, the juvenile X_i performs vertical learning from the parent X_{p1} or X_{p2} according to (13); otherwise, the juvenile X_i will perform horizontal learning in the original way of NCCLA, i.e., it randomly selects a juvenile with a better adaptation value than itself as the learning object.

$$X_i(t-1) = \begin{cases} X_{p1}(t-1), & \text{if } \frac{e^{-\text{sim}_{i,p1}(t-1)}}{1 + e^{-5 \times NF_{p1}}} \geq \frac{e^{-\text{sim}_{i,p2}(t-1)}}{1 + e^{-5 \times NF_{p2}}}, \\ X_{p2}(t-1), & \text{else,} \end{cases} \quad (13)$$

where $\text{sim}_{i,p1}$ and $\text{sim}_{i,p2}$ denote the cosine similarity of juvenile X_i to the parent individuals X_{p1} and X_{p2} , respectively, and NF_i denotes the normalization ability of juvenile X_i as shown in equation (14). The better the individual fitness value, the stronger its normalization ability.

$$NF_i = \begin{cases} \left| \frac{\max(F) - F_i}{\sum_{i=1}^N F_i} \right|, & \text{if } \sum_{i=1}^N F_i \neq 0, \\ \frac{1}{N}, & \text{else.} \end{cases} \quad (14)$$

In summary, this section proposes the complete social learning approach with the following advantages. First, compared with the original fixed-probability selection of vertical or horizontal learning, the new selection approach proposed in this section, as shown in (12), can rely on individuals' conditions to adaptively select the learning mode. Only juvenile individuals with convergence potential, i.e., those who are more similar to the optimal individuals and have better fitness values, will perform vertical learning. In

contrast, the rest of the individuals will perform horizontal learning. Obviously, this approach makes a small number of dominant individuals, who are not far from the optimal parent and are more excellent, focus on mining in the region with more convergence prospects and then quickly determine the more excellent evolutionary direction, driving the rapid convergence of the rest individuals. In turn, most of the remaining individuals learn from other better juvenile crows, which can develop other search areas, facilitating the maintenance of population diversity and reducing the risk of the algorithm falling into a local optimum. Second, the new vertical learning approach as shown in (13) proposed in this section, which relies on a comprehensive judgment of individual fitness values and the degree of similarity with individuals of the two parents, selects individuals learning from the parent with less similarity and excellent performance, and enables the juvenile crows to explore different promising areas as much as possible, which can better maintain the population diversity while ensuring the convergence speed. In short, the complete learning approach proposed in this section selects the learning mode and learning objects in a targeted way according to the juvenile crows' conditions, which effectively improves the convergence speed of the algorithm and maintains the population diversity to a certain extent. In addition, the above process no longer uses the parameters VSL_{prob} and SL_{prob} , which avoids the trouble of parameter debugging.

3.2.2. Incomplete Learning. As mentioned above, the incomplete learning phase will generate new learning objects by copying and absorbing the behavioral attributes of different individuals. To further ensure that the algorithm maintains population diversity and avoids falling into local optimum while improving the convergence speed, this section improves the asocial learning behavior and social learning behavior in the original NCCLA separately and proposes a new incomplete learning approach. The details are as follows.

(1) Improvement of social learning behavior

This section improves the conditions of vertical and horizontal learning in the social learning approach in

NCCLA, as well as the horizontal learning approach, and proposes a new incomplete social learning approach as follows. First, the adaptive selection factor $\delta_i(t)$ is calculated for the i -th juvenile crow individual X_i as shown in (15); then, a random number rand between $[0,1]$, and if $\text{rand} \leq \delta_i(t)$ is generated, vertical learning is performed, i.e., the behavioral attributes corresponding to the j -th dimension of the parent individual X_{p1} or X_{p2} as shown in equation (13) are copied, otherwise, horizontal learning is performed, i.e., the behavioral attributes of the j -th dimension of the individual shown in equation (16) are copied.

$$\delta_i(t) = 1 - e^{-|F_i(t-1) - F_{gbest}(t-1)|}, \quad (15)$$

where $F_{gbest}(t-1)$ and $F_i(t-1)$ represent the fitness values of the globally optimal individual and the juvenile X_i before the current generation's juvenile crow learning operation, respectively.

$$X_{i,j}(t) = \begin{cases} X_{s,j}(t-1), & \text{if } \text{rand} \leq 0.5, \\ \frac{X_{p1,j}(t-1) + X_{r1,j}(t-1) + X_{r2,j}(t-1)}{3}, & \text{else,} \end{cases} \quad (16)$$

where X_{p1} is the optimal parent individual, X_{r1} is a randomly selected individual with a better fitness value than the juvenile crow X_i , X_{r2} is a randomly selected individual in the population, and X_s is the individual selected among all juvenile crow individuals using roulette selection according to the probability corresponding to equation (17).

$$P_k(t) = \frac{e^{-\text{sim}_{i,k}(t-1)}}{1 + e^{-5 \times \text{NF}_k}}, \quad (17)$$

where $\text{sim}_{i,k}$ denotes the cosine similarity between the juvenile crow X_i and the juvenile crow X_k before learning was performed, and NF_k denotes the normalization ability of the juvenile crow X_k .

(2) Improvement of asocial learning behavior

In NCCLA, the probability of asocial learning behavior occurring is only $1 - \text{VSL}_{\text{prob}} \times \text{PI}_{\text{prob}}$, which is only $1 - 0.99 \times 0.95$ by its proposed parameter setting. In essence, such a small probability of asocial learning behavior occurring is intended to provide new evolutionary genes and reduce the possibility of the algorithm falling into a local optimum. Obviously, there is no need to keep certain properties unchanged with a certain probability. For this reason, the asocial learning in the incomplete learning phase only randomly updates each behavioral attribute according to equation (1).

The analysis shows that the incomplete learning approach proposed in this section has the following advantages. First, each behavioral attribute of new juvenile individuals after incomplete learning may

originate from different individuals, including parent individuals X_{p1} and X_{p2} , any other individual in the population, and new genes generated by asocial learning, forming various combinations that can ensure excellent population diversity. Second, the incomplete learning phase mainly uses incomplete social learning and rarely uses asocial learning. Compared with the original social learning approach, the new social incomplete learning proposed in this section better adapts to the needs of the algorithm performance at different stages of the algorithm, as follows: at the early stage of evolution, the individuals in the population are more distributed, with better population diversity, and the difference between individuals and optimal individuals is large. According to the adaptive selection factor, most of the dimensions of the new individuals come from the parent individuals, and a few dimensions come from other juvenile individuals, which further promotes the rapid convergence of the population; while in the late evolutionary stage, all individuals gradually approach the globally optimal individuals, the population shows a certain aggregation, and the population diversity decreases, at this time, most of the dimensions of the new individuals are derived from other more experienced juvenile individuals, which maintains the population diversity without slowing down the convergence of the algorithm. Third, when individuals perform horizontal learning in a certain dimension, they can no longer learn and communicate only with other juveniles who are better than themselves but have the opportunity to exchange information with any individual in the population, which further enhances population diversity, and learning from individuals who are less similar to and better than themselves based on roulette selection does not reduce the convergence speed while maintaining population diversity. In summary, the incomplete learning approach proposed in this section can indeed achieve the goal of maintaining population diversity and maximizing convergence speed.

3.3. Improvement in Reinforcement Phase

3.3.1. Improvements of Juvenile Crows Reinforcement Phase.

In fact, the juvenile crow reinforcement phase in NCCLA is an offset search of RW near the learning objectives identified in the juvenile crow learning phase. The offset range RW is a combination of α and β , where α represents self-perception and β represents social perception, which aims to enhance the exchange of evolutionary information with other individuals and thus explore more search space. The experimental study shows that the unreasonable settings of RW and β make the quality of the reinforced individuals still have much room for improvement, for which they are improved separately.

(1) Improvement of β calculation method

An in-depth analysis of (8) reveals that β is actually a reference to the information of other individuals and scales the individual itself by $\exp(-\text{lf} \times r \times t \times \text{mean}(j))$ times. As the number of iterations t increases, if

mean $(j) < 0$, $\exp(-lf \times r \times t \times \text{mean}(j))$ will be huge, even tending to infinity. The corresponding β and RW are also extremely large, making it extremely easy for the reinforced individuals to exceed the search range, resulting in ineffective reinforcement and making it difficult to provide new individuals with more excellence, leading to slow convergence of the algorithm or even failure to converge to the global optimum. Given this, a new calculation of the social perception factor β is proposed, as shown in equation (18).

$$\beta = (X_{i,j}(t-1) - X_{k,j}(t-1)) \times w(t), \quad (18)$$

where k is a randomly selected individual different from X_i in the unlearned juvenile crow population, i.e., $i \neq k$, and $w(t)$ is a weighting factor, as shown in equation.

$$w(t) = w_{\max} - \frac{w_{\max} - w_{\min}}{1 + e^{-0.1 \times r \times t \times f}}, \quad (19)$$

where w_{\max} and w_{\min} are denoted as the maximum and minimum values of the weighting factors, respectively, generally, when w_{\max} and w_{\min} are 2 and 0, better results can be obtained.

The change process of the weight factor with iteration is shown in Figure 3. At the early stage of evolution, the weight factor maintains a large value, which disguisedly increases the social learning phase in the reinforcement phase. The communication between individuals is more extensive, which makes individuals search extensively in different regions, further enhancing the global search ability of individuals and reducing the risk of the algorithm falling into the local optimum; as the evolution proceeds, the weight factor gradually decreases, especially at the late stage of evolution, the weight factor maintains a small value for a long time, which weakens the communication between individuals and other individuals, and enhances the local exploration ability of individuals in their neighborhood, which is more conducive to finding the global optimum.

(2) Improvement of RW calculation method

An in-depth analysis of RW calculation, as shown in (6), reveals that when α and β are zero simultaneously, the reinforcement phase is ineffective and fails to provide new individuals. In practical optimization, the possibility of this situation is not low. For example, in the late stage of algorithm evolution, almost all individuals in the population will gather around the optimal individual, and the vast majority of them have similar behavioral attributes, or even some behavioral attributes of some individuals are entirely identical when α and β are likely to be zero at the same time. Obviously, in order to enable individuals to perform a refined search around the obtained optimal value and thus converge to the globally optimal position, when both α and β are zero at the same time, a new stochastic reinforcement

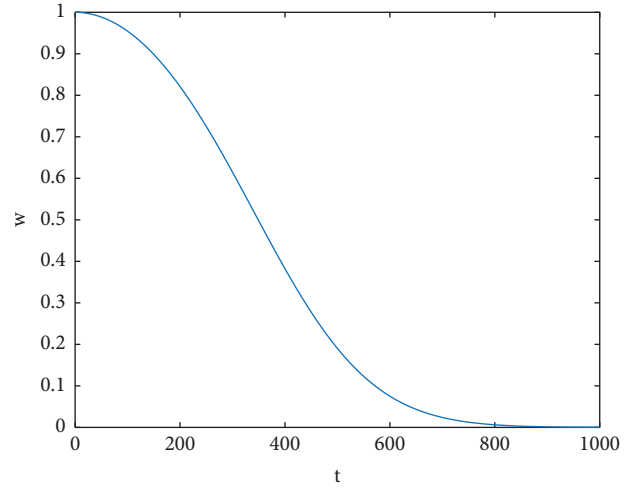


FIGURE 3: Schematic diagram of the variation of weighting factors with iterations.

strategy is proposed to calculate RW as shown in equation (20). Here, it should be noted that when α and β are not zero simultaneously, RW is still calculated according to the original way shown in equation (6).

$$RW = r \times (X_{s1,j}(t-1) - X_{s2,j}(t-1)), \quad (20)$$

where $s1$ and $s2$ are two different individuals randomly selected in the unlearned juvenile crow population, i.e., $s1 \neq s2 \neq i$, and r is a random number uniformly distributed in the range $[0, 1]$.

As seen in (20), when the individuals within the population are more similar, the above random reinforcement strategy makes the individuals perform a small random perturbation near themselves, increasing the possibility of convergence of the algorithm to the global optimum.

3.3.2. Improvements of Parent Crows Reinforcement Phase.

In NCCLA, the parent individuals in this iteration are the two relatively better individuals identified in the previous iteration of the population, which directly guide the evolution of the current generation of juveniles play an important role in the exploration and exploitation of the algorithm. As seen from the parent reinforcement phase in Section 2.3, the two-parent individuals each self-update in their independent reinforcement according to the probabilistic RP_{prob} . Further simplification reveals that the parent individuals X_{p1} and X_{p2} perform the reinforcement operation according to Equations 21 and 22.

$$X_{p1,j}(t) = e^{r1 \times (\text{mean}(j) - X_{p1,j}(t-1))}, \quad (21)$$

$$X_{p2,j}(t) = X_{2,j}(t-1) - r2 \times X_{1,j}(t-1) + e^{r1 \times (\text{mean}(j) - X_{p2,j}(t-1))}. \quad (22)$$

Both of these reinforcement methods include $e^{r1 \times (\text{mean}(j) - X_{p1,j}(t-1))}$. From the properties of the e exponential function, we can find that the effect of $e^{r1 \times (\text{mean}(j) - X_{p1,j}(t-1))}$ is to amplify the gap between $\text{mean}(j)$ and the two-parent individuals, especially when $\text{mean}(j) - X_{i,j}(t-1) > 0$, the gap amplification is more pronounced. In general, early in evolution, individual distribution is extremely dispersed, and the gap between individuals is not small, which is easily exceeded by the search space of the optimization problem after exponential amplification, resulting in ineffective reinforcement and waste of computational resources. Moreover, in the late stage of evolution, although the gap between individuals is not large, the further amplification of the gap by $e^{r1 \times (\text{mean}(j) - X_{i,j}(t-1))}$ will cause the parent individuals to produce a not small offset in this dimension, which is also very easy to deviate from the excellent evolutionary direction and will likewise cause ineffective reinforcement. Obviously, the reinforcement mentioned above of the two-parent individuals cannot effectively meet the needs of algorithm evolution. Given the different roles of the two-parent individuals in the algorithm,

the reinforcement methods of these two-parent individuals are improved separately further to balance the exploration and exploitation capabilities of the algorithm.

(1) Novel reinforcement of parent individual X_{p1}

The parent individual X_{p1} is the optimal individual determined by relying on the previous iteration of the population, which plays a vital role in guiding the algorithm's convergence. However, if its evolutionary direction points directly to the optimum local peak, it will increase the possibility of the algorithm falling into the local optimum. In view of the fact that the juvenile individuals of this generation have already achieved self-improvement based on the two more excellent parent individuals, carrying more excellent evolutionary information, which can provide more excellent reference information for the parent individuals to determine the evolutionary direction, a novel reinforcement was designed for the first parent individual X_{p1} as shown in Equation .

$$X_{p1,j}(t) = \begin{cases} X_{p1,j}(t-1) + G_1 \times (\text{mean}(j) - X_{p1,j}(t-1)), & \text{if } X_{p1,j}(t-1) \neq X_{k,j}(t) \\ X_{p1,j}(t-1) + r1 \times e^{\frac{-16t^2}{\max - t^2}} \times (X_{p1,j}(t-1) - X_{k,j}(t)), & \text{else} \end{cases}, \quad (23)$$

where $r1$ is a random number uniformly distributed in the range $[-1.5, 1.5]$, $X_{k,j}(t)$ is the j -th behavioral attribute of a randomly selected individual different from $X_{p1}(t-1)$ in the current population, and G_1 is a number conforming to the Gaussian distribution $N \sim (X_{p1,j}(t-1), 1)$, as shown in Equation

$$G_1 = \frac{1}{\sqrt{2\pi}} \times e^{\frac{-(\text{mean}(j) - X_{p1,j}(t-1))^2}{2}}. \quad (24)$$

From the above novel reinforcement approach for the parent individual X_{p1} , it can be seen that if for the j -th dimension of X_{p1} relying on the probability RP_{prob} determines that reinforcement is needed, an individual $X_k(t)$ needs to be randomly selected from the current population. By comparing whether $X_{k,j}(t)$ is equal to $X_{p1,j}(t-1)$, it is determined that learning toward the mean (j) or $X_{k,j}(t)$ is performed. At the early stage of evolution, $X_{k,j}(t)$ is almost not equal to $X_{p1,j}(t-1)$. The j -th dimension of X_{p1} is reinforced by learning from $X_{k,j}(t)$. Since the individuals in the current population already carry more excellent evolutionary information, the optimal evolutionary direction can be determined as soon as possible with reference to

their evolutionary direction. As evolution proceeds, the range $r1 \times e^{-16t^2/\max - t^2}$ of X_{p1} tries to explore a more optimal region gradually decreases, and gradually locks in a smaller range near the optimal individual to exploit a more optimal individual. This is in accordance with the process and law of evolutionary algorithms that gradually approach the region where the global optimal solution is located, avoiding the phenomenon of missing the optimal solution due to too large a search range. In addition, because the individuals of X_{p1} reinforcement learning are not the same on each behavioral attribute, it further enriches the range of dominant regions that can be explored by X_{p1} , increasing the possibility of the algorithm locking the actual optimal region and reducing the risk of the algorithm falling into a local optimum due to a single evolutionary direction. Moreover, as evolution proceeds, individuals are more similar by the late stage of evolution, and the j -th dimension of X_{p1} will be reinforced with a high probability by learning from mean (j). Compared with $X_{k,j}(t)$, the j -th dimension of X_{p1} is more likely to be different from mean (j), and the difference between mean (j) and the j -th dimension of X_{p1} is smaller, making the optimal individual X_{p1} able to perform a more

refined search within the current range, making it easier for the individual to find the optimal value of the current dimension, thus speeding up the convergence of the algorithm.

(2) Novel reinforcement of parent individual X_{p2}

Unlike the parent individual X_{p1} , which focuses on determining the evolutionary direction of the algorithm and improving the overall convergence speed, the parent individual X_{p2} focuses on providing excellent evolutionary information to guide the evolution of juveniles while maintaining population diversity. Given this, for the parent individual X_{p2} , let it learn from the optimal individual X_{p1} and other individuals in the population together and propose a novel reinforcement approach as shown in Equation .

$$\begin{aligned} X_{p2,j}(t) &= X_{p2,j}(t-1) \\ &+ r1 \times (X_{p1,j}(t-1) - X_{p2,j}(t-1)) \quad (25) \\ &+ G_2 \times (X_{p2,j}(t-1) - X_{q,j}(t)), \end{aligned}$$

where individual X_q is a juvenile randomly selected from the population of juvenile crows after reinforcement learning, and G_2 is a number that fits the Gaussian distribution $N \sim (X_{p2,j}(t-1), 0.5)$, as shown in Equation .

$$G_2 = \frac{1}{\sqrt{2\pi} \times 0.5} \times e^{-\frac{(\text{mean}(j) - X_{p1,j}(t-1))^2}{2 \times 0.5}}. \quad (26)$$

Compared with the parent individual X_{p1} reinforcement method, the parent individual X_{p2} learns from the optimal individual X_{p1} when performing reinforcement to ensure that it does not deviate from the optimal evolutionary direction, but when referring to the evolutionary information of other individuals, it can explore in a more extensive search range, potentially providing information on other excellent locations that are not in the same region as the parent individual, making it possible for the juvenile individuals learning from it to explore in other regions that are not the same as X_{p1} , further balancing the exploration and exploitation capabilities of the algorithm. In addition, the parent individual X_{p2} controls the scope of learning from the optimal individual X_{p1} and individual X_q with $r1$ and G_2 , respectively. In the early evolutionary stage, G_2 is greater than $r1$ with a greater probability, i.e., compared with learning from the optimal individual, the parent individual X_{p2} learns more from the remaining individuals, which can better maintain population diversity, ensure the global search of the algorithm, and increase the possibility of convergence of the algorithm to the global optimum. In the late evolutionary stage, G_2 is smaller than $r1$ with a greater

probability, i.e., the parent individual X_{p2} learns mainly from the optimal individual, and learns supplementally from the rest of the individuals, further ensuring that individuals perform a refined search near the optimal individual, thus locking the global optimal position.

3.4. Algorithm Complexity Analysis. Assuming that the population size is N , the maximum number of iterations is T , the problem dimension is D , and the numbers of parent and juvenile jays are N_p and N_j , respectively. The INCCLA algorithm mainly consists of the juvenile crow learning phase (T_{JL}), the juvenile crow reinforcement phase (T_{JR}), and the parent reinforcement phase (T_{PR}). The worst time complexity of each stage of the INCCLA algorithm in a single run is analyzed as follows: in the juvenile crow learning stage (T_{JL}), at most $(N_j - R) \times N$ times (1) needs to be computed, then the worst time complexity of this stage is $O((N_j - R) \times N)$; in the juvenile crow reinforcement stage (T_{JR}), at most $N_j \times D$ times (5) needs to be computed, then the worst time complexity of this stage is $O(N_j \times D)$; in the parent reinforcement stage (T_{PR}), at most $N_p \times D$ times (23) or (25) needs to be computed, then the worst time complexity of this stage is $O(N_p \times D)$.

Therefore, the worst time complexity required for a single run of INCCLA is $O((N_j - R) \times N) + O(N_j \times D) + O(N_p \times D) \approx O(T \times N \times (N_j - R + D))$.

4. Experimental Results and Discussion

In this section, to verify the performance of the INCCLA algorithm, the following four parts of experiments will be conducted in this paper: (1) The parameter sensitivity analysis; (2) Verification of the effectiveness of the proposed three improvement strategies; (3) Compare the performance of the improved INCCLA algorithm with the original NCCLA algorithm and four other representative and excellent performance evolutionary algorithms; (4) Comparison of the effectiveness of each algorithm in engineering applications.

This section uses the CEC2013 and CEC2020 test suites for experimental simulation. The CEC2013 test suite contains a total of 28 test functions, among which $F1 \sim F5$ are unimodal functions, which have only one optimal value and are used to verify the convergence performance of the algorithm; $F6 \sim F20$ are multimodal functions, which have multiple locally optimal solutions and are used to verify the ability of the algorithm to escape from the local optimum; $F21 \sim F28$ are composition functions. The CEC2020 test suite contains a total of 10 test functions, of which $F1$ is a unimodal function, $F2 \sim F4$ are multimodal shifted and rotated functions, $F5 \sim F7$ are hybrid functions, and $F8 \sim F10$ are composition functions. The relevant functions for the CEC2013 and CEC2020 test suites can be found in the literature [32, 33], respectively.

In this section, to ensure the fairness of the algorithm comparison, all algorithms were run on a computer with Windows 11 operating system, i5-11400H CPU, and programmed with MATLAB R2021a.

4.1. Sensitivity Analysis of Parameters. The INCCLA algorithm proposed in this paper involves five parameters: RP_{prob} , SL_{prob} , R , lf_{min} , and lf_{max} . Compared with the NCCLA algorithm, the INCCLA algorithm retains the originally proposed settings of RP_{prob} and SL_{prob} , changes the settings of lf_{min} and lf_{max} , and adds the parameter R . Given this, to analyze further the effect of parameters on the performance of the INCCLA algorithm, this section only analyzes the effect of parameters R , lf_{min} and lf_{max} on the performance of the INCCLA algorithm. To ensure fairness of comparison, the population size in each algorithm is $N = 50$, the dimension of the test function is $D = 30$, and the maximum number of function evaluations is $\text{MaxFEs} = 150,000$, $RP_{\text{prob}} = 0.9$, $SL_{\text{prob}} = 0.99$.

Table 1 gives the mean and average values of the optimal results obtained from 30 independent runs of the INCCLA algorithm on the CEC2013 test set when R is set to different parameters. In this experiment, $lf_{\text{min}} = 0.0001$ and $lf_{\text{max}} = 0.09$. Table 2 give the mean and average values of the optimal results of the INCCLA algorithm for 30 independent runs on the CEC2013 test set when lf_{min} and lf_{max} are set to different combinations of parameters. For this experiment, $R = 15$. Tables 1 and 2 blacken the parameters that achieved the best results on each function and count the number of functions that performed best on each parameter set in the last row.

According to the data in Table 1, it can be seen that INCCLA achieves the best convergence on 19 test functions when $R = 15$. When $R = 5, 25$, and 35 , the best convergence is achieved on 9, 11, and 14 test functions, respectively. Thus, the performance of INCCLA is sensitive to the setting of the parameter R , and the algorithm performs best when $R = 15$. Similarly, according to the data in Table 2, the performance of INCCLA is also sensitive to the settings of lf_{min} and lf_{max} , and INCCLA achieves the best convergence among the 19 tested functions when $lf_{\text{min}} = 0.0001$ and $lf_{\text{max}} = 0.09$. In summary, if there is no special requirement, INCCLA with R , lf_{min} , and lf_{max} set to 15, 0.0001, and 0.09, respectively, the algorithm can obtain better optimization results.

4.2. Experiment on Effectiveness of Each Improvement Strategy. According to Section 3, it is known that the INCCLA algorithm improves the NCCLA algorithm in three aspects. In this paper, to verify the effectiveness of the three improvement strategies, the NCCLA algorithm is combined with the three improvement strategies individually to form three new algorithms, namely, the INCCLA algorithm based on cosine similarity, the INCCLA algorithm based on improved juvenile learning phase, and the INCCLA algorithm based on improved reinforcement phase, named as INCCLA1, INCCLA2, and INCCLA3, respectively, and

compared with the original NCCLA algorithm on the CEC2013 test suit.

In this section, to ensure fairness of comparison, the population size in each algorithm is $N = 50$, the dimension of the test function is $D = 30$, and the maximum number of function evaluations is $\text{MaxFEs} = 5000 \times D = 150,000$. The other parameters of each algorithm are set as shown in Table 3. To avoid the contingency of a single operation of the algorithms, each algorithm was run 30 times independently on each test function.

Table 4 presents the running results of each algorithm on 28 test functions in 30 dimensions, where the “ \pm ” before and after represents the mean and standard deviation of the optimal values in 30 experiments, respectively, and the data that outperform the original NCCLA algorithm on the same function are marked in bold. To compare the significance of the performance of each improved strategy with the performance of NCCLA and to verify that the obtained results are not coincidental, Tables 5 and 6 present the results of the Wilcoxon rank sum test and Friedman test [34] between each improved strategy and NCCLA algorithm on 28 test functions, respectively. In Table 5, when the p value is greater than 0.05, it indicates that there is no significant difference between the improvement strategy and NCCLA, which is indicated by the symbol “ $=$.” When the p value is less than 0.05, and the mean value of the optimal solution of the result obtained in 30 experiments of the improvement strategy is better than NCCLA, it indicates that the improvement strategy is significantly better than NCCLA, which is indicated by the symbol “ $+$ ”; otherwise, it indicates that the performance of the improvement strategy is significantly worse than NCCLA, which is indicated by the symbol “ $-$.” In Table 6, the smaller the rank mean value corresponding to the algorithm, the better the algorithm’s overall performance.

Table 4 shows that INCCLA1 obtains better mean values on all the remaining 24 test functions except $F1, F8, F20$, and $F21$ compared to NCCLA. INCCLA2 obtained better mean values on all 24 tested functions except $F4, F8, F15$, and $F23$; INCCLA3 obtained better mean values on all 26 tested functions except $F17$ and $F27$. As can be seen from Table 5, the Wilcoxon rank sum test results for INCCLA1 and NCCLA on the four tested functions of $F4, F5, F10$, and $F13$ are “ $+$,” indicating that INCCLA1 outperforms NCCLA on these four functions. The Wilcoxon rank sum test result of “ $-$ ” on the $F21$ test function indicates the inferior performance of INCCLA1 over NCCLA on $F21$, while the Wilcoxon rank sum test result of “ $=$ ” on the remaining 23 test functions indicates that they perform similarly; INCCLA2 outperforms NCCLA on 14 test functions, has similar performance to NCCLA on 13 test functions, and inferior performance to NCCLA on $F23$ test functions; while INCCLA3 outperforms NCCLA on 15 test functions and has similar performance to NCCLA on 13 test functions. As can be seen from Table 6, the rank means of INCCLA2 is the smallest, indicating that the overall performance of the algorithm is superior, and the rank means of both INCCLA1 and INCCLA2 are smaller than those of NCCLA, indicating

TABLE 1: Effect of parameter R on the performance of the algorithm.

Function	$R = 5$	$R = 15$	$R = 25$	$R = 35$
F1	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F2	7.62E+05 ± 3.02E+05	8.10E+05 ± 4.96E+05	7.76E+05 ± 3.66E+05	2.45E+06 ± 3.89E+06
F3	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F4	1.13E+04 ± 3.87E+03	1.10E+04 ± 3.46E+03	1.43E+04 ± 3.64E+03	1.82E+04 ± 4.62E+03
F5	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F6	2.87E+01 ± 2.49E+01	2.40E+01 ± 2.04E+01	2.91E+01 ± 2.50E+01	3.39E+01 ± 2.58E+01
F7	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F8	2.09E+01 ± 4.40E-02	2.10E+01 ± 5.23E-02	2.09E+01 ± 8.22E-02	2.10E+01 ± 5.06E-02
F9	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F10	2.68E-01 ± 9.36E-02	2.10E-01 ± 5.91E-02	1.91E-01 ± 7.21E-02	1.24E-01 ± 5.80E-02
F11	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F12	4.00E+00 ± 2.15E+01	0.00E+00 ± 0.00E+00	4.08E+00 ± 2.23E+01	0.00E+00 ± 0.00E+00
F13	4.86E+00 ± 2.62E+01	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F14	6.82E+00 ± 4.57E+00	6.29E+00 ± 4.92E+00	3.88E+01 ± 7.61E+01	5.26E+01 ± 7.95E+01
F15	4.70E+03 ± 5.73E+02	3.94E+03 ± 6.65E+02	3.69E+03 ± 4.42E+02	3.69E+03 ± 7.34E+02
F16	1.03E+00 ± 3.89E-01	9.33E-01 ± 2.63E-01	9.36E-01 ± 3.55E-01	9.57E-01 ± 2.58E-01
F17	4.28E+00 ± 7.40E+00	5.43E+00 ± 7.92E+00	7.20E+00 ± 9.07E+00	8.64E+00 ± 4.20E+00
F18	1.18E+02 ± 3.47E+01	9.65E+01 ± 2.06E+01	9.90E+01 ± 2.63E+01	8.60E+01 ± 2.39E+01
F19	6.73E+00 ± 1.85E+00	5.32E+00 ± 2.09E+00	5.42E+00 ± 1.54E+00	5.53E+00 ± 1.84E+00
F20	1.94E-01 ± 1.05E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F21	4.15E+02 ± 8.03E+01	4.00E+02 ± 0.00E+00	4.00E+02 ± 0.00E+00	4.00E+02 ± 0.00E+00
F22	9.01E+01 ± 6.02E+01	7.70E+01 ± 5.90E+01	1.10E+02 ± 6.44E+01	1.11E+02 ± 7.74E+01
F23	3.96E+03 ± 5.75E+02	4.06E+03 ± 5.45E+02	3.55E+03 ± 7.00E+02	3.67E+03 ± 6.44E+02
F24	2.00E+02 ± 1.81E-02	2.00E+02 ± 2.44E-02	2.00E+02 ± 2.13E-02	2.00E+02 ± 1.93E-02
F25	2.59E+02 ± 3.90E+01	2.40E+02 ± 3.15E+01	2.42E+02 ± 3.21E+01	2.23E+02 ± 2.13E+01
F26	3.06E+02 ± 1.60E+01	2.87E+02 ± 3.45E+01	2.97E+02 ± 1.82E+01	3.00E+02 ± 5.68E-07
F27	3.27E+02 ± 7.05E+01	3.13E+02 ± 8.90E-01	3.13E+02 ± 9.46E-01	3.13E+02 ± 9.22E-01
F28	9.64E+02 ± 3.39E+01	9.14E+02 ± 2.15E+01	8.83E+02 ± 1.81E+01	8.60E+02 ± 1.80E+01
	9	19	11	14

Bold indicates the best results obtained on each function.

TABLE 2: Effect of parameters lf_{\min} and lf_{\max} on the performance of the algorithm.

Function	$lf_{\min} = 0.0001, lf_{\max} = 0.005$	$lf_{\min} = 0.0001, lf_{\max} = 0.09$	$lf_{\min} = 0.0005, lf_{\max} = 0.02$	$lf_{\min} = 0.005, lf_{\max} = 0.2$
F1	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F2	1.03E+06 ± 6.20E+05	8.10E+05 ± 4.96E+05	8.62E+05 ± 5.43E+05	9.26E+05 ± 5.10E+05
F3	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F4	1.42E+04 ± 4.01E+03	1.10E+04 ± 3.46E+03	1.25E+04 ± 3.36E+03	1.28E+04 ± 3.47E+03
F5	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F6	2.99E+01 ± 2.55E+01	2.40E+01 ± 2.04E+01	2.61E+01 ± 2.29E+01	2.60E+01 ± 2.23E+01
F7	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F8	2.10E+01 ± 4.85E-02	2.10E+01 ± 5.23E-02	2.10E+01 ± 4.17E-02	2.10E+01 ± 5.84E-02
F9	8.85E-01 ± 4.85E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F10	2.12E-01 ± 9.61E-02	2.10E-01 ± 5.91E-02	2.16E-01 ± 6.67E-02	2.15E-01 ± 7.12E-02
F11	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F12	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F13	4.75E+00 ± 2.60E+01	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F14	6.81E+02 ± 8.45E+02	6.29E+00 ± 4.92E+00	1.38E+02 ± 3.17E+02	9.30E+00 ± 2.29E+01
F15	4.21E+03 ± 5.84E+02	3.94E+03 ± 6.65E+02	3.89E+03 ± 7.58E+02	3.78E+03 ± 6.28E+02
F16	1.02E+00 ± 3.13E-01	9.33E-01 ± 2.63E-01	9.73E-01 ± 2.67E-01	9.77E-01 ± 3.63E-01
F17	3.17E+01 ± 4.06E+01	5.43E+00 ± 7.92E+00	1.05E+01 ± 1.10E+01	4.87E+00 ± 8.60E+00
F18	1.29E+02 ± 2.59E+01	9.65E+01 ± 2.06E+01	1.19E+02 ± 3.04E+01	9.69E+01 ± 2.65E+01
F19	7.11E+00 ± 2.59E+00	5.32E+00 ± 2.09E+00	6.10E+00 ± 2.32E+00	5.27E+00 ± 1.94E+00
F20	2.72E-01 ± 1.49E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00	0.00E+00 ± 0.00E+00
F21	4.00E+02 ± 0.00E+00	4.00E+02 ± 0.00E+00	3.93E+02 ± 3.65E+01	4.00E+02 ± 0.00E+00
F22	5.18E+02 ± 8.92E+02	7.70E+01 ± 5.90E+01	1.25E+02 ± 7.26E+01	1.16E+02 ± 5.89E+01
F23	4.06E+03 ± 7.22E+02	4.06E+03 ± 5.45E+02	4.29E+03 ± 6.34E+02	4.02E+03 ± 6.07E+02
F24	2.00E+02 ± 2.30E-02	2.00E+02 ± 2.44E-02	2.00E+02 ± 2.24E-02	2.00E+02 ± 2.46E-02
F25	2.34E+02 ± 3.04E+01	2.40E+02 ± 3.15E+01	2.51E+02 ± 3.59E+01	2.56E+02 ± 3.80E+01
F26	3.01E+02 ± 7.13E+00	2.87E+02 ± 3.45E+01	2.97E+02 ± 2.91E+01	2.98E+02 ± 2.03E+01
F27	3.14E+02 ± 1.49E+00	3.13E+02 ± 8.90E-01	3.13E+02 ± 1.03E+00	3.13E+02 ± 1.14E+00
F28	9.10E+02 ± 2.50E+01	9.14E+02 ± 2.15E+01	9.07E+02 ± 1.73E+01	9.19E+02 ± 2.33E+01
	7	19	13	13

Bold indicates the best results obtained on each function.

TABLE 3: Algorithm-related parameters.

Algorithm	Parameter
NCCLA	$RP_{\text{prob}} = 0.9; SL_{\text{prob}} = 0.99; VSL_{\text{prob}} = 0.99; P1_{\text{prob}} = 0.95; TaE_{\text{prob}} = 0.3; lf_{\text{min}} = 0.0005; lf_{\text{max}} = 0.02$
INCCLA1	$RP_{\text{prob}} = 0.9; SL_{\text{prob}} = 0.99; VSL_{\text{prob}} = 0.99; P1_{\text{prob}} = 0.95; TaE_{\text{prob}} = 0.3; lf_{\text{min}} = 0.0005; lf_{\text{max}} = 0.02$
INCCLA2	$RP_{\text{prob}} = 0.9; SL_{\text{prob}} = 0.99; R = 15; lf_{\text{min}} = 0.0005; lf_{\text{max}} = 0.02$
INCCLA3	$RP_{\text{prob}} = 0.9; SL_{\text{prob}} = 0.99; VSL_{\text{prob}} = 0.99; P1_{\text{prob}} = 0.95; TaE_{\text{prob}} = 0.3; lf_{\text{min}} = 0.0001; lf_{\text{max}} = 0.09$

TABLE 4: Results of each improvement strategy in the 30-dimensional CEC2013 test suite.

Function	NCCLA	INCCLA1	INCCLA2	INCCLA3
F1	$1.99E-05 \pm 5.60E-05$	$4.22E-05 \pm 1.59E-04$	$4.19E-06 \pm 1.30E-05$	$2.10E-30 \pm 9.42E-30$
F2	$7.59E+06 \pm 4.32E+06$	$6.38E+06 \pm 3.75E+06$	$6.37E+06 \pm 3.13E+06$	$7.13E+05 \pm 2.69E+05$
F3	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$
F4	$7.59E+03 \pm 2.43E+03$	$6.33E+03 \pm 2.47E+03$	$8.13E+03 \pm 2.51E+03$	$3.28E+03 \pm 1.98E+03$
F5	$8.91E-04 \pm 3.77E-03$	$3.42E-05 \pm 8.93E-05$	$6.86E-05 \pm 2.37E-04$	$0.00E+00 \pm 0.00E+00$
F6	$6.64E+01 \pm 3.21E+01$	$5.82E+01 \pm 2.90E+01$	$5.15E+01 \pm 2.35E+01$	$3.12E+01 \pm 2.78E+01$
F7	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$
F8	$2.10E+01 \pm 5.25E-02$	$2.10E+01 \pm 5.42E-02$	$2.10E+01 \pm 4.17E-02$	$2.10E+01 \pm 3.49E-02$
F9	$5.26E+00 \pm 9.82E+00$	$2.23E+00 \pm 6.87E+00$	$0.00E+00 \pm 0.00E+00$	$2.52E+00 \pm 7.71E+00$
F10	$9.26E+00 \pm 7.28E+00$	$6.80E+00 \pm 5.65E+00$	$6.67E+00 \pm 3.25E+00$	$2.16E-01 \pm 1.98E-01$
F11	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$	$0.00E+00 \pm 0.00E+00$
F12	$1.16E+02 \pm 3.57E+01$	$8.96E+01 \pm 6.22E+01$	$0.00E+00 \pm 0.00E+00$	$1.02E+02 \pm 7.05E+01$
F13	$1.31E+02 \pm 6.87E+01$	$1.18E+02 \pm 6.15E+01$	$0.00E+00 \pm 0.00E+00$	$1.07E+02 \pm 8.38E+01$
F14	$1.78E+01 \pm 3.12E+01$	$8.96E+00 \pm 4.34E+00$	$4.24E+00 \pm 2.81E+00$	$1.15E+01 \pm 2.49E+01$
F15	$4.79E+03 \pm 1.50E+03$	$4.33E+03 \pm 1.14E+03$	$7.17E+03 \pm 2.72E+02$	$3.65E+03 \pm 5.90E+02$
F16	$2.68E+00 \pm 3.22E-01$	$2.50E+00 \pm 3.43E-01$	$2.62E+00 \pm 2.98E-01$	$1.06E+00 \pm 3.87E-01$
F17	$9.59E+00 \pm 1.21E+01$	$7.98E+00 \pm 1.19E+01$	$5.44E+00 \pm 6.87E+00$	$1.10E+01 \pm 1.37E+01$
F18	$1.77E+02 \pm 4.90E+01$	$1.66E+02 \pm 4.61E+01$	$1.43E+02 \pm 4.88E+01$	$1.54E+02 \pm 3.64E+01$
F19	$1.57E+01 \pm 7.70E+00$	$1.47E+01 \pm 6.78E+00$	$7.74E+00 \pm 2.36E+00$	$9.11E+00 \pm 2.73E+00$
F20	$1.31E+01 \pm 3.16E+00$	$1.34E+01 \pm 2.31E+00$	$0.00E+00 \pm 0.00E+00$	$1.20E+01 \pm 3.58E+00$
F21	$4.00E+02 \pm 2.73E-01$	$4.00E+02 \pm 3.12E-01$	$4.00E+02 \pm 1.97E-01$	$4.00E+02 \pm 7.62E-12$
F22	$1.17E+02 \pm 6.11E+01$	$7.99E+01 \pm 7.61E+01$	$9.09E+01 \pm 6.46E+01$	$1.04E+02 \pm 8.25E+01$
F23	$4.21E+03 \pm 6.46E+02$	$4.07E+03 \pm 1.11E+03$	$6.90E+03 \pm 4.60E+02$	$3.87E+03 \pm 9.03E+02$
F24	$2.04E+02 \pm 1.37E+01$	$2.01E+02 \pm 7.70E-01$	$2.00E+02 \pm 4.13E-02$	$2.02E+02 \pm 1.10E+01$
F25	$2.88E+02 \pm 3.40E+01$	$2.79E+02 \pm 3.98E+01$	$2.48E+02 \pm 4.21E+01$	$2.85E+02 \pm 3.25E+01$
F26	$3.08E+02 \pm 4.25E+01$	$2.99E+02 \pm 4.17E+01$	$2.97E+02 \pm 1.82E+01$	$2.97E+02 \pm 3.89E+01$
F27	$4.46E+02 \pm 2.75E+02$	$3.94E+02 \pm 1.81E+02$	$3.22E+02 \pm 2.68E+00$	$6.16E+02 \pm 3.16E+02$
F28	$1.46E+03 \pm 8.95E+02$	$1.21E+03 \pm 5.33E+02$	$1.08E+03 \pm 5.76E+01$	$1.05E+03 \pm 3.53E+02$

Bold indicates the best results obtained on each function.

that all three improvement strategies proposed in this paper achieve better results.

In summary, all three improvement strategies proposed in this paper have certain improvement effects on NCCLA. The improvement strategies in the juvenile learning phase and the improvement strategies in the reinforcement phase have the most obvious improvement effects.

4.3. Performance Comparison of INCCLA with Other Algorithms. In this section, to verify the superior performance of the INCCLA algorithm in terms of convergence precision and convergence speed, this section compares the INCCLA with the NCCLA algorithm and the four better evolutionary algorithms on the CEC2013 and CEC2020 test suites, including the artificial bee colony algorithm based on new neighborhood selection mechanism (NSABC) [35], the sine cosine algorithm based on transition parameters and mutation operators (MSCA) [36], artificial tree algorithm

based on two populations (IATTP) [37] and the improved crow search algorithm (ICSA) [38]. To ensure the fairness of the comparison, the population size in each algorithm is $N = 50$, and the maximum number of function evaluations is $\text{MaxFEs} = 150,000$. The other parameters of each algorithm are set as shown in Table 7, where the parameter values of each comparison algorithm are taken as in the original paper.

4.3.1. Comparison of INCCLA with Other Algorithms on Convergence Precision

(1) *Testing at CEC 2013.* In order to fully compare the performance of INCCLA with other algorithms in terms of convergence precision, tests were conducted on the CEC2013 test suite with three different dimensions, $D = 10$, $D = 30$, and $D = 100$, respectively. Tables 8–10 give the mean and standard deviation of 30 independent experiments for each algorithm

TABLE 5: Wilcoxon rank sum test results between NCCLA and each improvement strategy.

Function	<i>p</i> value (vs.NCCLA)		
	INCCLA1	INCCLA2	INCCLA3
F1	0.876 (=)	0.053 (=)	0.000 (+)
F2	0.304 (=)	0.379 (=)	0.000 (+)
F3	1.000 (=)	1.000 (=)	1.000 (=)
F4	0.029 (+)	0.363 (=)	0.000 (+)
F5	0.003 (+)	0.004 (+)	0.000 (+)
F6	0.559 (=)	0.082 (=)	0.000 (+)
F7	1.000 (=)	1.000 (=)	1.000 (=)
F8	0.067 (=)	0.340 (=)	0.077 (=)
F9	0.179 (=)	0.005 (+)	0.218 (=)
F10	0.035 (+)	0.270 (=)	0.000 (+)
F11	1.000 (=)	1.000 (=)	1.000 (=)
F12	0.302 (=)	0.000 (+)	0.385 (=)
F13	0.025 (+)	0.000 (+)	0.616 (=)
F14	0.491 (=)	0.000 (+)	0.012 (+)
F15	0.099 (=)	0.000 (+)	0.000 (+)
F16	0.057 (=)	0.395 (=)	0.000 (+)
F17	0.090 (=)	0.118 (=)	0.137 (=)
F18	0.420 (=)	0.006 (+)	0.045 (+)
F19	0.706 (=)	0.000 (+)	0.000 (+)
F20	0.264 (=)	0.000 (+)	0.273 (=)
F21	0.019 (-)	0.994 (=)	0.000 (+)
F22	0.067 (=)	0.050 (=)	0.911 (=)
F23	0.112 (=)	0.000 (-)	0.108 (=)
F24	0.482 (=)	0.000 (+)	0.000 (+)
F25	0.807 (=)	0.000 (+)	0.473 (=)
F26	0.529 (=)	0.000 (+)	0.000 (+)
F27	0.599 (=)	0.000 (+)	0.510 (=)
F28	0.258 (=)	0.002 (+)	0.000 (+)
+/-/-	4/23/1	14/13/1	15/13/0

TABLE 6: Friedman test results for four algorithms.

	NCCLA	INCCLA1	INCCLA2	INCCLA3
Avg.rank	3.48	2.48	1.96	2.07
Sort	4	3	1	2

TABLE 7: Related parameter settings of each algorithm.

Algorithm	Parameter
NCCLA	$RP_{\text{prob}} = 0.9$; $SL_{\text{prob}} = 0.99$; $VSL_{\text{prob}} = 0.99$; $P1_{\text{prob}} = 0.95$; $TaE_{\text{prob}} = 0.3$; $lf_{\text{min}} = 0.0005$; $lf_{\text{max}} = 0.02$
NSABC	$C = 1.5$; $\text{limit} = 100$; $k = 10$
MSCA	$a = 2$; $b = 0.5$
IATTP	$h1 = h2 = h3 = 0.5$, $h4 = 0.8$, $m = 50$, $q = 0.8$
ICSA	$AP = 0.1$; $FL = 1.5$
INCCLA	$RP_{\text{prob}} = 0.9$; $SL_{\text{prob}} = 0.99$; $R = 15$; $lf_{\text{min}} = 0.0001$; $lf_{\text{max}} = 0.09$

on the 10-dimension, 30-dimension, and 100-dimension CEC2013 datasets. The functions that achieved the best optimization results on the same functions are bold. To further verify the differences between INCCLA and each algorithm, the Wilcoxon rank sum test with a 5% significance level was performed between INCCLA and each algorithm, and the results are shown in Table 11. To comprehensively evaluate the overall performance of all algorithms, the results of the Friedman test are given in Table 12.

For the 10-dimensional optimization problem, as shown in Table 8, INCCLA achieves better mean values in all 28 test functions compared with NCCLA, and INCCLA achieves global optimum on 9 test functions, including F1, F3, F5, F7, F9, F11, F12, F13, and F20. Both the INCCLA algorithm and NSABC achieved the global optimum on nine test functions, including F1, F3, F5, F7, F9, F11, F12, F13, and F20, and the two algorithms achieved comparable mean values on functions F25 and F26. On five functions, including F19,

TABLE 8: Data result of INCCLA with other algorithms on the 10-dimensional CEC2013 test suite.

Fun.	NSABC		MSCA		IATTP		ICSA		NCCLA		INCCLA	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
F1	0.00E+00	0.00E+00	2.17E+02	2.79E+02	8.91E-03	1.05E-02	4.53E-07	2.34E-06	1.39E-08	4.82E-08	0.00E+00	0.00E+00
F2	7.16E+05	4.42E+05	2.62E+06	2.16E+06	5.11E+04	4.09E+04	3.07E+05	2.59E+05	7.28E+05	1.16E+06	4.01E+04	3.74E+04
F3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F4	7.16E+03	2.30E+03	5.31E+03	1.98E+03	7.64E+00	3.64E+00	1.04E+03	6.49E+02	1.55E+03	2.11E+03	5.52E+02	4.35E+02
F5	0.00E+00	0.00E+00	4.85E+01	1.34E+01	4.64E-02	3.16E-02	7.30E-01	3.80E+00	5.24E-07	1.43E-06	0.00E+00	0.00E+00
F6	5.71E+00	4.78E+00	2.03E+01	2.04E+01	7.76E+00	3.97E+00	2.09E+01	2.60E+01	2.82E+01	3.25E+01	4.03E+00	4.80E+00
F7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F8	6.78E-01	3.71E+00	4.44E-16	3.01E-31	4.44E-16	3.01E-31	6.77E-01	3.71E+00	1.22E+01	1.01E+01	4.44E-16	3.01E-31
F9	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F10	7.48E-01	2.40E-01	2.86E+01	2.44E+01	7.28E-01	1.61E-01	5.27E+00	4.75E+00	5.90E+00	4.26E+00	1.31E-01	8.01E-02
F11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F12	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F13	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F14	1.92E+00	5.25E+00	1.68E+03	2.04E+02	1.38E+03	1.70E+02	1.20E+02	1.10E+02	2.60E+00	3.27E+00	7.04E-01	1.56E+00
F15	6.26E+02	1.55E+02	1.54E+03	1.68E+02	1.03E+03	1.78E+02	5.33E+02	2.85E+02	7.31E+02	2.81E+02	4.43E+02	2.14E+02
F16	5.16E-01	1.79E-01	1.20E+00	2.36E-01	1.11E+00	1.50E-01	1.13E+00	2.30E-01	6.28E-01	2.82E-01	4.10E-01	1.85E-01
F17	4.98E+00	5.10E+00	7.16E+01	1.88E+01	3.12E+01	5.59E+00	6.15E+00	4.54E+00	4.89E+00	5.12E+00	4.43E+00	4.92E+00
F18	2.06E+01	3.94E+00	6.70E+01	1.84E+01	3.41E+01	6.04E+00	2.52E+01	6.02E+00	2.51E+01	1.10E+01	1.79E+01	5.21E+00
F19	7.05E-01	1.89E-01	5.42E+00	1.34E+00	2.23E+00	3.18E-01	8.89E-01	2.04E-01	1.65E+00	1.01E+00	8.08E-01	2.49E-01
F20	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.50E-01	8.23E-01	0.00E+00	0.00E+00
F21	3.19E+02	9.82E+01	4.09E+02	2.50E+01	4.00E+02	2.56E-04	3.84E+02	5.14E+01	3.98E+02	7.01E+00	3.90E+02	5.59E+01
F22	7.17E+01	5.88E+01	1.32E+03	2.90E+02	1.34E+03	2.20E+02	1.20E+02	8.23E+01	8.72E+01	6.55E+01	6.54E+01	6.12E+01
F23	6.37E+02	1.59E+02	1.65E+03	2.05E+02	1.43E+03	1.52E+02	3.02E+02	1.39E+02	8.88E+02	3.61E+02	3.96E+02	2.06E+02
F24	1.00E+02	8.14E-03	1.73E+02	4.50E+01	1.00E+02	1.15E-02	1.17E+02	3.79E+01	1.69E+02	4.73E+01	1.50E+02	5.09E+01
F25	1.00E+02	1.28E-04	1.17E+02	3.76E+01	1.01E+02	1.41E-01	1.03E+02	1.98E+00	1.00E+02	1.08E-04	1.00E+02	0.00E+00
F26	1.00E+02	4.96E-13	1.17E+02	3.79E+01	1.00E+02	4.28E-13	1.00E+02	4.65E-12	1.23E+02	4.30E+01	1.20E+02	4.07E+01
F27	3.09E+02	3.35E-01	3.16E+02	4.02E+01	3.08E+02	1.45E-01	3.10E+02	5.89E-01	3.21E+02	4.33E+01	3.09E+02	3.19E-01
F28	3.31E+02	2.48E+02	6.65E+02	1.29E+02	3.37E+02	2.77E+02	4.43E+02	3.01E+02	5.18E+02	2.82E+02	5.06E+02	2.53E+02

Bold indicates the best results obtained on each function.

TABLE 9: Data result of INCCLA with other algorithms on the 30-dimensional CEC2013 test suite.

Fun.	NSABC			MSCA			IATTP			ICSA			NCCLA			INCCLA		
	Mean	Std		Mean	Std		Mean	Std		Mean	Std		Mean	Std		Mean	Std	
F1	0.00E+00	0.00E+00	9.60E+03	4.32E+03	3.00E+00	3.36E+01	3.41E+01	3.41E+01	1.99E-05	5.60E-05	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F2	1.01E+07	3.38E+06	7.18E+07	2.22E+07	3.71E+06	2.11E+07	1.04E+07	1.04E+07	7.59E+06	4.32E+06	8.10E+05	4.96E+05	8.10E+05	4.96E+05	8.10E+05	4.96E+05	4.96E+05	4.96E+05
F3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F4	8.13E+04	1.31E+04	3.74E+04	7.65E+03	4.01E+02	1.33E+04	2.91E+03	2.91E+03	7.59E+03	2.43E+03	1.10E+04	3.46E+03	1.10E+04	3.46E+03	1.10E+04	3.46E+03	1.10E+04	3.46E+03
F5	0.00E+00	0.00E+00	2.72E+03	1.54E+03	1.58E+01	1.58E+01	9.80E+00	3.09E+02	8.91E-04	3.77E-03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F6	2.27E+01	1.20E+01	6.45E+02	3.67E+02	8.24E+01	1.35E+02	3.38E+01	3.38E+01	6.64E+01	3.21E+01	2.40E+01	2.04E+01	2.40E+01	2.04E+01	2.40E+01	2.04E+01	2.40E+01	2.04E+01
F7	0.00E+00	0.00E+00	8.51E-02	3.51E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F8	2.10E+01	5.54E-02	2.10E+01	5.62E-02	2.10E+01	2.10E+01	6.35E-02	6.35E-02	2.10E-01	5.25E-02	2.10E+01	5.23E-02	2.10E+01	5.23E-02	2.10E+01	5.23E-02	2.10E+01	5.23E-02
F9	0.00E+00	0.00E+00	2.37E+00	6.82E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	5.26E+00	9.82E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F10	2.20E+00	4.88E-01	1.06E+03	2.58E+02	2.14E+01	1.56E+02	7.69E+01	7.69E+01	9.26E+00	7.28E+00	2.10E-01	5.91E-02	2.10E-01	5.91E-02	2.10E-01	5.91E-02	2.10E-01	5.91E-02
F11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F12	2.03E+01	4.62E+01	3.05E+01	6.95E+01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.16E+02	3.57E+01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F13	8.17E+00	3.11E+01	4.51E+01	8.33E+01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.31E+02	6.87E+01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F14	6.38E+00	2.25E+01	7.02E+03	4.08E+02	6.87E+03	2.73E+03	6.75E+02	6.75E+02	1.78E+01	3.12E+01	6.29E+00	4.92E+00	6.29E+00	4.92E+00	6.29E+00	4.92E+00	6.29E+00	4.92E+00
F15	3.99E+03	3.54E+02	7.97E+03	2.26E+02	5.19E+02	6.11E+03	1.01E+03	1.01E+03	4.79E+03	1.50E+03	3.94E+03	6.65E+02	3.94E+03	6.65E+02	3.94E+03	6.65E+02	3.94E+03	6.65E+02
F16	1.08E+00	2.70E-01	2.79E+00	2.88E-01	2.57E+00	2.60E+00	2.71E-01	2.71E-01	2.68E+00	3.22E-01	9.33E-01	2.63E-01	9.33E-01	2.63E-01	9.33E-01	2.63E-01	9.33E-01	2.63E-01
F17	6.77E+00	1.19E+01	6.47E+02	1.18E+02	2.05E+02	1.91E+01	3.07E+01	3.07E+01	9.59E+00	1.21E+01	5.43E+00	7.92E+00	5.43E+00	7.92E+00	5.43E+00	7.92E+00	5.43E+00	7.92E+00
F18	1.66E+02	2.57E+01	6.17E+02	1.94E+02	2.21E+02	2.22E+02	3.41E+01	3.41E+01	1.77E+02	4.90E+01	9.65E+01	2.06E+01	9.65E+01	2.06E+01	9.65E+01	2.06E+01	9.65E+01	2.06E+01
F19	8.23E+00	1.70E+00	3.10E+03	4.81E+03	1.73E+01	1.58E+01	8.31E+00	8.31E+00	1.57E+01	7.70E+00	5.32E+00	2.09E+00	5.32E+00	2.09E+00	5.32E+00	2.09E+00	5.32E+00	2.09E+00
F20	8.44E+00	6.04E+00	3.10E+00	5.26E+00	5.56E-01	2.13E+00	4.69E+00	4.69E+00	6.58E+00	3.16E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F21	3.97E+02	1.69E+01	1.01E+03	2.52E+02	4.02E+02	5.22E-01	2.11E+01	2.11E+01	4.00E+02	2.73E-01	4.00E+02	0.00E+00	4.00E+02	0.00E+00	4.00E+02	0.00E+00	4.00E+02	0.00E+00
F22	8.12E+01	6.21E+01	7.45E+03	5.44E+02	7.20E+03	3.25E+02	7.62E+02	7.62E+02	1.17E+02	6.11E+01	7.70E+01	5.90E+01	7.70E+01	5.90E+01	7.70E+01	5.90E+01	7.70E+01	5.90E+01
F23	4.40E+03	3.98E+02	7.87E+03	3.16E+02	7.60E+03	4.00E+02	4.00E+02	4.00E+02	5.40E+03	4.21E+03	4.06E+03	5.45E+02	4.06E+03	5.45E+02	4.06E+03	5.45E+02	4.06E+03	5.45E+02
F24	2.00E+02	4.34E-02	2.07E+02	1.04E+01	2.00E+02	9.22E-02	2.95E-01	2.95E-01	2.04E+02	1.37E+01	2.00E+02	2.44E-02	2.00E+02	2.44E-02	2.00E+02	2.44E-02	2.00E+02	2.44E-02
F25	2.27E+02	2.13E+01	2.78E+02	3.44E+01	2.42E+02	3.48E+01	2.12E+01	2.12E+01	2.88E+02	3.40E+01	2.40E+02	3.15E+01	2.40E+02	3.15E+01	2.40E+02	3.15E+01	2.40E+02	3.15E+01
F26	2.01E+02	1.77E-01	3.06E+02	5.46E+01	2.58E+02	4.81E+01	3.01E+01	3.01E+01	3.08E+02	4.25E+01	2.87E+02	3.45E+01	2.87E+02	3.45E+01	2.87E+02	3.45E+01	2.87E+02	3.45E+01
F27	3.31E+02	2.04E+01	6.23E+02	2.42E+02	3.37E+02	9.82E+01	8.59E+01	8.59E+01	4.46E+02	2.75E+02	3.13E+02	8.90E-01	3.13E+02	8.90E-01	3.13E+02	8.90E-01	3.13E+02	8.90E-01
F28	9.39E+02	1.59E+02	2.58E+03	4.31E+02	1.01E+03	2.17E+01	3.69E+02	3.69E+02	1.60E+03	1.46E+03	9.14E+02	2.15E+01	9.14E+02	2.15E+01	9.14E+02	2.15E+01	9.14E+02	2.15E+01

Bold indicates the best results obtained on each function.

TABLE 10: Data result of INCCLA with other algorithms on the 100-dimensional CEC2013 test suite.

Fun.	NSABC			MSCA			IATTP			ICSA			NCCLA			INCCLA		
	Mean	Std		Mean	Std		Mean	Std		Mean	Std		Mean	Std		Mean	Std	
F1	1.85E-20	2.55E-20	8.85E+04	9.37E+03	2.84E+02	1.36E+04	3.64E+03	1.73E+03	2.84E+02	1.36E+04	3.64E+03	1.15E+00	1.85E+00	7.80E-12	1.18E-11			
F2	6.23E+07	1.21E+07	9.81E+08	2.70E+08	4.82E+07	2.20E+08	7.15E+07	2.20E+08	4.82E+07	2.20E+08	7.15E+07	3.58E+07	8.52E+06	2.02E+07	4.71E+06			
F3	0.00E+00	0.00E+00	1.28E+15	3.11E+15	2.34E+10	1.53E+10	1.45E+12	1.53E+10	2.34E+10	7.89E+11	1.45E+12	0.00E+00	0.00E+00	0.00E+00	0.00E+00			
F4	3.11E+05	3.36E+04	2.21E+05	1.57E+04	6.28E+03	3.95E+04	1.66E+04	3.95E+04	6.28E+03	9.43E+04	1.66E+04	6.76E+04	1.36E+04	1.19E+05	1.52E+04			
F5	2.19E-12	2.45E-12	4.07E+04	1.22E+04	2.17E+02	8.53E+03	3.07E+03	1.63E+03	2.17E+02	8.53E+03	3.07E+03	1.81E+02	4.28E+00	1.74E-06	1.22E-06			
F6	2.27E+02	3.41E+01	1.37E+04	2.93E+03	1.28E+02	2.03E+03	3.77E+02	8.88E+02	1.28E+02	2.03E+03	3.77E+02	3.20E+02	5.22E+01	2.36E+02	5.69E+01			
F7	4.26E+02	2.99E+02	1.26E+04	9.57E+03	1.51E+02	5.76E+02	4.53E+02	2.49E+02	1.51E+02	5.76E+02	4.53E+02	3.93E+02	4.94E+02	0.00E+00	0.00E+00			
F8	2.13E+01	2.11E-02	2.13E+01	3.44E-02	1.82E-02	2.13E+01	2.34E-02	2.13E+01	1.82E-02	2.13E+01	2.34E-02	2.13E+01	2.89E-02	2.13E+01	2.14E-02			
F9	1.28E+02	3.51E+00	1.24E+02	8.02E+00	6.41E+00	1.18E+02	1.05E+01	1.18E+02	6.41E+00	7.86E+01	1.05E+01	1.01E+02	8.79E+00	7.02E+01	3.63E+01			
F10	2.49E+01	1.08E+01	9.81E+03	1.29E+03	2.26E+02	2.29E+03	4.02E+02	1.41E+03	2.26E+02	2.29E+03	4.02E+02	1.60E+02	4.54E+01	4.63E+00	1.45E+00			
F11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	6.96E-01	3.81E+00	8.47E+01	2.79E+01	0.00E+00	0.00E+00	0.00E+00	0.00E+00			
F12	8.02E+02	7.60E+01	1.25E+03	1.38E+02	7.86E+01	5.77E+02	7.69E+01	7.08E+02	7.86E+01	5.77E+02	7.69E+01	5.71E+02	1.02E+02	5.21E+02	7.37E+01			
F13	9.18E+02	7.95E+01	1.22E+03	1.55E+02	6.97E+02	6.97E+02	4.48E+01	6.97E+02	6.93E+01	6.27E+02	4.48E+01	8.15E+02	1.51E+02	6.24E+02	4.86E+01			
F14	1.80E+02	1.49E+02	3.14E+04	7.61E+02	1.24E+03	1.89E+04	2.67E+03	3.15E+04	1.24E+03	1.89E+04	2.67E+03	2.65E+02	2.15E+02	7.58E+03	5.27E+03			
F15	1.80E+04	9.42E+02	3.21E+04	5.71E+02	6.01E+02	3.05E+04	5.18E+02	3.10E+04	6.01E+02	3.05E+04	5.18E+02	3.01E+04	2.86E+03	2.03E+04	1.71E+03			
F16	2.56E+00	3.13E-01	4.53E+00	2.88E-01	2.58E-01	4.30E+00	2.67E-01	4.36E+00	2.58E-01	4.30E+00	2.67E-01	4.27E+00	2.93E-01	2.06E+00	3.75E-01			
F17	4.28E+00	9.78E+00	4.72E+03	3.43E+02	7.39E+01	1.28E+03	2.28E+02	1.31E+03	7.39E+01	1.28E+03	2.28E+02	4.60E+01	8.90E+00	9.60E+01	2.45E+01			
F18	1.75E+03	1.44E+02	4.90E+03	4.89E+02	8.63E+01	1.74E+03	1.58E+02	1.32E+03	8.63E+01	1.74E+03	1.58E+02	1.14E+03	2.04E+02	7.03E+02	1.24E+02			
F19	1.10E+02	1.27E+01	4.60E+05	1.90E+05	2.81E+02	6.99E+02	4.08E+03	6.99E+02	2.81E+02	7.19E+03	4.08E+03	1.80E+02	4.93E+01	7.04E+01	1.88E+01			
F20	5.00E+01	0.00E+00	5.00E+01	0.00E+00	2.63E-09	5.00E+01	1.50E-04	5.00E+01	2.63E-09	5.00E+01	1.50E-04	5.00E+01	1.83E-02	4.99E+01	1.65E-01			
F21	3.79E+02	3.43E+01	7.31E+03	2.84E+02	4.26E+02	5.05E+03	7.92E+02	1.29E+03	4.26E+02	5.05E+03	7.92E+02	4.46E+02	6.81E+00	3.49E+02	4.88E+01			
F22	2.54E+02	1.24E+02	3.16E+04	6.60E+02	1.33E+03	1.78E+04	3.63E+03	3.11E+04	1.33E+03	1.78E+04	3.63E+03	2.89E+02	1.27E+02	7.59E+03	6.74E+03			
F23	2.23E+04	1.73E+03	3.35E+04	6.13E+02	6.55E+02	3.07E+04	8.30E+02	3.31E+04	6.55E+02	3.07E+04	8.30E+02	3.11E+04	3.75E+03	2.22E+04	1.48E+03			
F24	4.72E+02	7.75E+01	5.65E+02	4.14E+01	6.67E+01	4.12E+02	3.62E+01	4.40E+02	6.67E+01	4.12E+02	3.62E+01	4.06E+02	1.25E+02	2.42E+02	8.01E+01			
F25	6.06E+02	1.35E+01	6.73E+02	2.88E+01	1.70E+01	5.51E+02	1.74E+01	5.57E+02	1.70E+01	5.51E+02	1.74E+01	5.76E+02	2.28E+01	5.09E+02	4.82E+01			
F26	2.19E+02	7.44E+01	5.45E+02	4.62E+01	1.06E+02	4.27E+02	5.66E+01	4.52E+02	1.06E+02	4.27E+02	5.66E+01	4.47E+02	9.21E+01	3.10E+02	3.85E+01			
F27	3.30E+03	2.41E+02	3.30E+03	4.34E+02	6.09E+02	1.97E+03	4.12E+02	2.49E+03	6.09E+02	1.97E+03	4.12E+02	2.16E+03	7.89E+02	5.76E+02	5.95E+02			
F28	4.58E+03	1.28E+03	1.47E+04	1.01E+03	3.02E+03	1.02E+04	1.02E+03	9.16E+03	3.02E+03	1.02E+04	1.02E+03	7.16E+03	3.07E+03	4.08E+03	1.19E+03			

Bold indicates the best results obtained on each function.

TABLE 11: Wilcoxon rank sum test results of INCCLA with other algorithms on the 10-dimensional and 30-dimensional CEC2013 test suite.

Function	$D = 10$ p value (vs.INCCLA)					$D = 30$ p value (vs.INCCLA)					$D = 100$ p value (vs.INCCLA)				
	NSABC	MSCA	IATTP	ICSA	NCCLA	NSABC	MSCA	IATTP	ICSA	NCCLA	NSABC	MSCA	IATTP	ICSA	NCCLA
F1	1.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F2	0.000 (-)	0.000 (-)	0.290 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F3	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)
F4	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (-)	0.002 (-)	0.000 (-)	0.000 (-)	0.000 (+)	0.004 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (+)	0.000 (+)
F5	1.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	1.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F6	0.030 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.190 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F7	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F8	0.333 (=)	1.000 (=)	1.000 (=)	0.333 (=)	0.000 (-)	0.865 (=)	0.935 (=)	0.258 (=)	0.654 (=)	0.695 (=)	0.970 (=)	0.137 (=)	0.728 (=)	0.180 (=)	0.911 (=)
F9	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.011 (-)	1.000 (=)	1.000 (=)	0.005 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.200 (=)	0.000 (-)
F10	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F11	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.081 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.333 (=)	0.000 (-)	1.000 (=)
F12	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.021 (-)	0.021 (-)	1.000 (=)	1.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.006 (-)	0.065 (=)
F13	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.160 (=)	0.002 (-)	1.000 (=)	1.000 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.706 (=)	0.000 (-)
F14	0.745 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.004 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (+)
F15	0.000 (-)	0.000 (-)	0.000 (-)	0.195 (=)	0.000 (-)	0.807 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.039 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F16	0.030 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.002 (-)	0.046 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F17	0.075 (=)	0.000 (-)	0.000 (-)	0.037 (-)	0.176 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.007 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (+)
F18	0.015 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.004 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F19	0.055 (=)	0.000 (-)	0.000 (-)	0.411 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F20	1.000 (=)	1.000 (=)	1.000 (=)	1.000 (=)	0.333 (=)	0.000 (-)	0.002 (-)	0.160 (=)	0.000 (-)	0.000 (-)	0.041 (-)	0.000 (-)	0.151 (=)	0.151 (=)	0.043 (=)
F21	0.001 (+)	0.000 (-)	0.000 (-)	0.038 (-)	0.000 (-)	0.333 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F22	0.711 (=)	0.000 (-)	0.000 (-)	0.001 (-)	0.013 (-)	0.166 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.028 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (+)
F23	0.000 (-)	0.000 (-)	0.000 (-)	0.102 (=)	0.000 (-)	0.011 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.297 (=)	0.652 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F24	0.003 (+)	0.000 (-)	0.061 (=)	0.395 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F25	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.888 (=)	0.000 (-)	0.455 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F26	0.011 (+)	1.000 (=)	0.011 (+)	0.011 (+)	0.479 (=)	0.000 (+)	0.000 (-)	0.911 (=)	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F27	0.000 (-)	0.000 (-)	0.000 (+)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
F28	0.002 (+)	0.000 (-)	0.317 (=)	0.771 (=)	0.055 (=)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)
+/-/-	4/14/10	0/9/19	3/11/14	1/13/14	0/10/18	2/12/14	0/4/24	1/10/17	0/7/21	1/5/22	7/5/16	0/2/26	1/3/24	1/4/23	4/4/20

TABLE 12: Friedman test results of 6 algorithms.

Dimension		NSABC	MSCA	IATTP	ICSA	NCCLA	INCCLA
$D = 10$	Avg.rank	2.68	4.96	3.54	3.50	4.07	2.25
	sort	2	6	3	4	5	1
$D = 30$	Avg.rank	2.50	5.34	3.57	4.07	3.73	1.79
	sort	2	6	3	5	4	1
$D = 100$	Avg.rank	2.91	5.59	4.04	3.93	2.86	1.68
	sort	3	6	5	4	2	1

F_{21} , F_{24} , F_{26} , and F_{28} , the INCCLA algorithm had inferior mean values to NSABC, and on the remaining 12 functions, the INCCLA algorithm obtained better mean values than NSABC. MSCA, IATTP, and ICSA all achieved theoretical optima on only seven test functions, including F_3 , F_7 , F_9 , F_{11} , F_{12} , F_{13} , and F_{20} . As seen in Table 11, NCCLA has similar performance on 10 functions compared to INCCLA, but significantly inferior performance on the 18 tested functions; NSABC performs significantly better on 4 functions but significantly inferior on 10 functions; MSCA has similar performance on 9 of the tested functions, but significantly inferior performance on the remaining 19 tested functions; IATTP has significantly better performance on F_4 , F_{26} , and F_{27} only, but significantly inferior performance on the 14 tested functions; ICSA has significantly better performance on F_{26} only, but significantly inferior performance on all 14 functions. In summary, it shows that for low-dimensional optimization problems, the INCCLA proposed in this paper has some advantages in terms of convergence precision compared with other representative methods.

For the 30-dimensional optimization problem, it can be seen from Table 9 that INCCLA achieves the global optimum on nine test functions, including F_1 , F_3 , F_5 , F_7 , F_9 , F_{11} , F_{12} , F_{13} , and F_{20} . Like INCCLA, NCCLA also achieves the global optimum on these nine test functions, and INCCLA achieves worse mean values than NCCLA only on F_4 , but better mean values on the remaining test functions; NSABC achieved theoretical optima on six functions, including F_1 , F_3 , F_5 , F_7 , F_9 , and F_{11} ; MSCA obtained the theoretical optimal results on F_3 and F_{11} only; IATTP obtained the theoretical optimal for six functions, including F_3 , F_7 , F_9 , F_{11} , F_{12} , and F_{13} ; ICSA obtained the theoretical optimal for five functions, including F_3 , F_7 , F_9 , F_{12} , and F_{13} . As seen in Table 11, compared to INCCLA, NCCLA has similar performance on 5 functions and significantly better performance on F_4 only, but significantly inferior performance on the 22 tested functions; NSABC has significantly better performance on F_6 and F_{26} , but significantly inferior performance on 14 functions; MSCA has similar performance on 4 tested functions, but significantly inferior performance on the remaining 24 tested functions; IATTP had significantly better performance on F_4 only, but significantly inferior performance on 17 test functions; ICSA has similar performance on 7 test functions only, but significantly inferior performance on the remaining 21 test functions. In summary, it shows that for the 30-dimensional CEC2013 test suite, the INCCLA proposed in this paper has a significant advantage in

convergence accuracy compared with other representative methods, and the performance gap between algorithms is significantly more significant than that of the 10-dimensional optimization problem.

For the 100-dimensional optimization problem, it can be seen from Table 10 that INCCLA achieves the global optimum on three test functions, including F_3 , F_7 , and F_{11} . NCCLA achieves the global optimum on F_3 and F_{11} , except that INCCLA and NCCLA obtain the same mean value on F_8 , NCCLA obtains a better mean value than INCCLA on F_4 , F_{14} , F_{17} , and F_{22} , but NCCLA obtains a worse mean value than INCCLA on the rest of the 21 functions; NSABC obtained the global optimum on F_3 and F_{11} ; MSCA obtained the theoretical optimum on F_{11} only; neither IATTP nor ICSA obtained the theoretical optimum on any function. As seen in Table 11, NCCLA performs significantly better on only four functions compared to INCCLA, including F_4 , F_{14} , F_{17} , and F_{22} , but performs significantly inferior on the 20 test functions; NSABC has similar performance on 5 tested functions, significantly better performance on 7 functions, but significantly inferior performance on 16 functions; MSCA has similar performance on F_8 and F_{11} , but significantly inferior performance on the remaining 26 test functions; IATTP has significantly better performance on F_4 only, but significantly inferior performance on the 24 test functions; ICSA has significantly better performance on F_4 only, but significantly inferior performance on the 23 test functions. In summary, it shows that for high-dimensional optimization problems, the INCCLA proposed in this paper has a significant advantage in convergence precision compared with other representative methods, and the performance gap between the algorithms is more significant than that for optimization problems in 10 and 30 dimensions.

Table 12 shows that the overall performance of INCCLA is the best among the six optimization algorithms for the 10-dimensional, 30-dimensional, and 100-dimensional test functions, and its advantage is more obvious as the dimensionality of the optimization problem increases. In both 10-dimensional and 30-dimensional test functions, NSABC ranks second in overall performance, and IATTP, ICSA, and NCCLA are slightly worse than NSABC in both dimensions. In the 100-dimensional test function, the overall performance of NCCLA ranked second, and NSABC, ICSA, and IATTP were slightly worse than NCCLA in this dimension. The overall performance of MSCA was the worst among the six algorithms. In summary, compared with other algorithms, the INCCLA proposed in this paper has certain advantages in terms of convergence precision.

TABLE 13: Data result of INCCLA with other algorithms on the CEC2020 test suite.

Function	NSABC		MSCA		IATTP		ICSA		NCCLA		INCCLA	
	Mean (std)	P value (vs. INCCLA)	Mean (std)	P value (vs. INCCLA)	Mean (std)	P value (vs. INCCLA)	Mean (std)	P value (vs. INCCLA)	Mean (std)	P value (vs. INCCLA)	Mean (std)	P value (vs. INCCLA)
F1	1.53E+03(1.89E+03)	0.000 (-)	1.42E+08 (9.70E+07)	0.000 (-)	2.66E+04 (1.66E+04)	0.000 (-)	4.17E+04 (1.28E+05)	0.000 (-)	3.25E+03 (2.75E+03)	0.000 (-)	2.38E-01 (5.98E-01)	
F2	7.31E+01 (6.93E+01)	0.739 (=)	1.15E+03 (2.35E+02)	0.000 (-)	1.17E+03 (1.42E+02)	0.000 (-)	1.71E+02 (1.47E+02)	0.145 (=)	3.23E+02 (2.38E+02)	0.000 (-)	9.33E+01(8.63E+01)	
F3	1.25E+01 (1.87E+00)	0.000 (+)	5.63E+01 (9.38E+00)	0.000 (-)	4.01E+01 (4.60E+00)	0.000 (-)	1.53E+01 (2.38E+00)	0.112 (=)	3.12E+01 (7.79E+00)	0.000 (-)	1.63E+01(3.44E+00)	
F4	6.49E-01(1.19E-01)	0.684 (=)	7.65E+00 (2.43E+00)	0.000 (-)	2.03E+00 (3.89E-01)	0.000 (-)	5.21E+00 (1.86E+01)	0.008 (-)	1.28E+00 (5.88E-01)	0.000 (-)	6.20E-01 (2.53E-01)	
F5	3.53E+04 (2.75E+04)	0.000 (-)	9.21E+03 (6.13E+03)	0.000 (-)	4.94E+02 (2.68E+02)	0.000 (-)	4.68E+03 (1.74E+04)	0.000 (-)	6.71E+03 (1.07E+04)	0.000 (-)	1.34E+02 (1.04E+02)	
F6	2.03E+01 (4.13E+01)	0.994 (=)	1.21E+02 (5.85E+01)	0.000 (-)	4.16E+01 (4.74E+01)	0.011 (+)	4.63E+01 (6.54E+01)	0.008 (-)	1.17E+02 (8.56E+01)	0.000 (-)	4.53E+01(6.66E+01)	
F7	4.11E+03 (3.95E+03)	0.000 (-)	2.83E+03 (1.39E+03)	0.000 (-)	1.42E+02 (6.70E+01)	0.000 (-)	1.02E+02 (8.29E+01)	0.000 (-)	4.84E+03 (5.59E+03)	0.000 (-)	2.86E+01 (3.67E+01)	
F8	1.07E+02 (1.49E+01)	0.570 (=)	1.21E+02 (6.45E+00)	0.000 (-)	1.07E+02 (1.52E+01)	0.000 (-)	1.10E+02 (2.20E-01)	0.000 (-)	1.10E+02 (1.10E-04)	0.000 (-)	1.04E+02 (2.43E+01)	
F9	2.41E+02(1.10E+02)	0.000 (+)	3.74E+02 (6.78E+00)	0.000 (-)	2.08E+02 (1.18E+02)	0.007 (+)	3.13E+02 (7.19E+01)	0.013 (+)	3.47E+02 (6.53E+00)	0.001 (-)	3.18E+02 (7.44E+01)	
F10	4.15E+02 (2.27E+01)	0.100 (=)	4.33E+02 (2.81E+01)	0.000 (-)	4.17E+02 (2.25E+01)	0.610 (=)	4.31E+02 (2.23E+01)	0.000 (-)	4.18E+02 (2.33E+01)	0.085 (=)	4.14E+02 (2.21E+01)	
+/-/-	2/5/3		0/0/10		2/1/7		1/2/7		0/1/9			
Avg.rank	2.45		5.60		3.25		3.65		4.25		1.80	
Sort	2		6		3		4		5		1	

(2) *Testing at CEC 2020.* To further examine the performance of the INCCLA algorithm, the INCCLA algorithm and four other algorithms are tested on the 10-dimensional CEC2020 test suite. The parameters of each algorithm were set as above. Table 13 counts the results of 30 independent experiments for each algorithm, including the mean and standard deviation, the results of the Wilcoxon rank sum test results for each algorithm with a significance level of 5% with the INCCLA algorithm, the number of functions for which each algorithm is significantly better than INCCLA, significantly worse than INCCLA, and not significantly different from INCCLA, and the results of the Friedman test and ranking results for each algorithm.

As shown in Table 13, for the 10-dimensional CEC2020 test function set, compared with NCCLA, INCCLA showed similar performance only on F_{10} , achieved better mean values on the remaining 9 test functions, and performed significantly better on them. Compared with NSABC, INCCLA performs significantly inferior only on F_9 and F_3 , achieves similar performance on 5 test functions, and performs significantly better on 3 test functions; compared to MSCA, INCCLA achieves better mean values on all 10 tested functions as well as showing significantly better performance. Compared to IATTP, INCCLA only showed inferior performance on F_6 and F_9 , similar performance on F_{10} , and significantly better performance on 7 test functions; compared to ICSA, INCCLA only showed inferior performance on F_9 , similar performance on F_2 and F_3 , and significantly better performance on 7 test functions. It is obvious from the rank sum calibration results that INCCLA has an advantage over several other comparative algorithms on the CEC2020 test suite.

In summary, compared with NCCLA and the other four superior optimization algorithms, INCCLA shows a significant advantage in convergence precision. The advantages become more obvious as the dimension of the optimization problem increases.

4.3.2. Comparison of INCCLA with Other Algorithms on Convergence Speed. In order to compare the differences in convergence speed among the algorithms more intuitively, Figure 4 shows the convergence curves of each algorithm on the 28 test functions of the CEC2013 test suite when the dimension is 30, where the horizontal coordinate is the number of function evaluations and the vertical coordinate is the logarithm of the fitness value.

As shown in Figure 4, for the unimodal functions $F_1 \sim F_5$, INCCLA can obtain the global optimum on F_1 , F_3 , and F_5 test functions, but the convergence speed is slightly slower than NSABC on F_1 and F_5 test functions; INCCLA shows better convergence precision and the fastest convergence speed than the other five algorithms on F_2 ; INCCLA is second only to IATTP in terms of convergence precision and convergence speed on F_4 . For multimodal functions $F_6 \sim F_{20}$, INCCLA can obtain the global optimum on F_7 , F_9 , F_{11} , F_{12} , F_{13} , and F_{20} test functions, and the convergence speed is second only to MSCA on F_{11} ; INCCLA shows better convergence precision and the fastest convergence speed

than the other five algorithms on F_6 , F_{10} , and F_{19} . INCCLA shows better convergence precision and relatively better convergence speed on F_8 , F_{14} , F_{15} , F_{16} , F_{17} , and F_{18} test functions. In the early evolutionary stage, the convergence speed of INCCLA is slightly slower than the other algorithms. When it reaches the late evolutionary stage, INCCLA can continue searching for better solutions compared to the reduced search performance of the other algorithms. For the composition functions $F_{21} \sim F_{28}$, INCCLA shows better convergence precision and fastest convergence speed than the other five algorithms on the test functions F_{25} , F_{27} , and F_{28} ; INCCLA shows higher convergence precision and relatively faster convergence speed on F_{22} and F_{23} . The convergence speed of INCCLA on F_{21} is similar to that of the three algorithms NSABC, NCCLA and IATTP, where the convergence precision of INCCLA and NSABC is slightly better than that of the other algorithms; the convergence speed of INCCLA on F_{24} is similar to that of the four algorithms NCCLA, ICSA, IATTP, and NSABC, with the convergence precision of INCCLA being higher. On F_{26} , the convergence speed and precision of INCCLA are similar to those of ICSA and IATTP, and the convergence precision of INCCLA is second only to that of NSABC. In summary, it shows that INCCLA has some advantages in terms of convergence speed compared with NCCLA and the other four superior optimization algorithms.

In summary, compared with NCCLA and the other four superior optimization algorithms, the INCCLA proposed in this paper is superior in the overall performance in terms of convergence precision and convergence speed, although it has the shortcoming of not converging fast enough in the early stage for individual complex functions. The essential reasons for this are all caused by the three improved evolutionary strategies in Sections 3.1–3.3 of this paper, which are analyzed in depth as follows. The update strategy of the INCCLA algorithm has a large step size in the early evolutionary stage, which makes individuals search a farther distance around themselves, thus ensuring the global search of the algorithm and reducing the possibility of the algorithm falling into a local optimum, for more complex optimization problems, which also inevitably slows down the rapid aggregation of the population around a dominant region. Therefore, it makes INCCLA not converge fast enough in the early stage on individual complex functions. However, as the iteration proceeds, the individual search step size gradually decreases, and it is easier to search to have the advantage of the area, easy to refine the search, which improves the convergence speed in the late evolutionary stage and facilitates the algorithm to obtain a more excellent convergence precision. In addition, the parental selection mechanism can guide the evolutionary direction of juveniles during the learning phase, allowing the algorithm to increase algorithmic diversity while maintaining convergence. The juvenile crow hybrid learning mechanism can determine the state of the algorithm in the current evolutionary stage according to the individual's attributes and select the learning mode in a targeted manner so that the algorithm can effectively maintain the balance of convergence speed and diversity in the evolutionary process. The interplay and

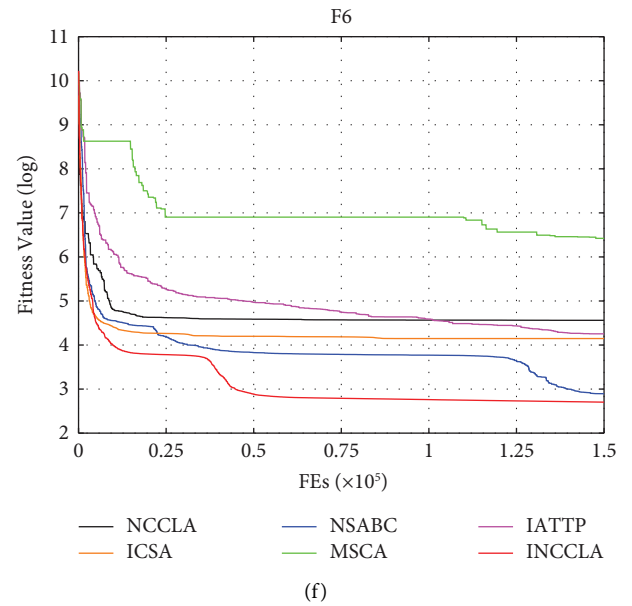
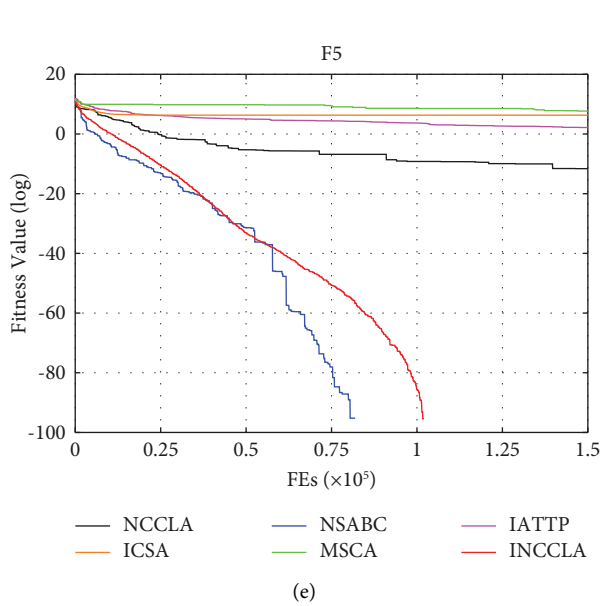
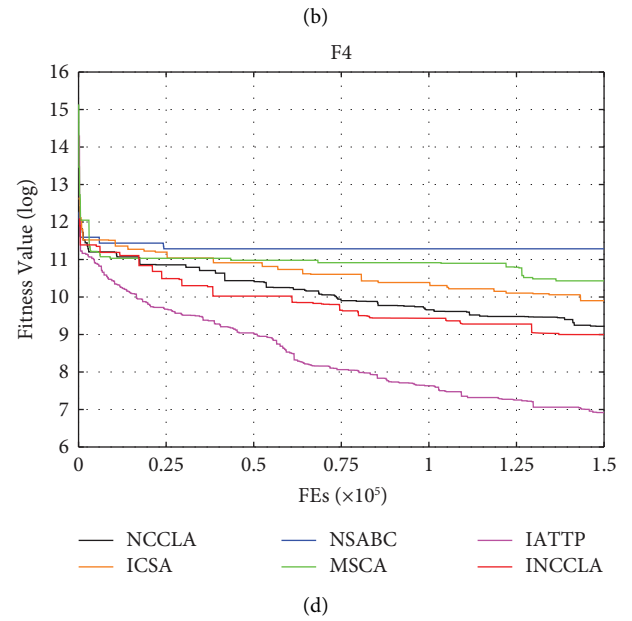
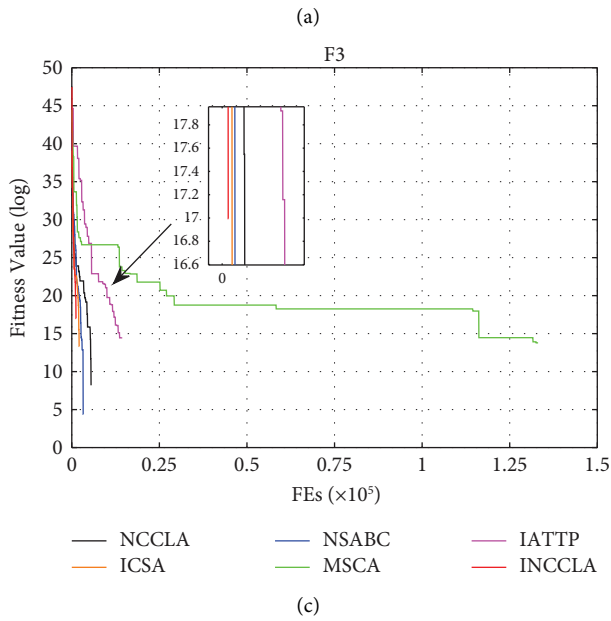
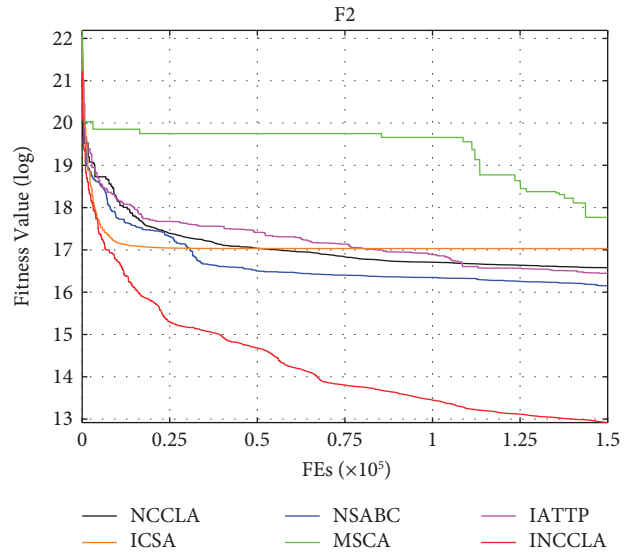
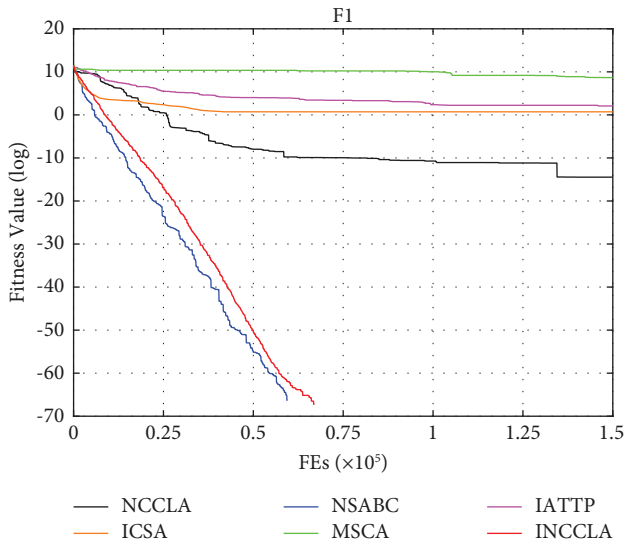


FIGURE 4: Continued.

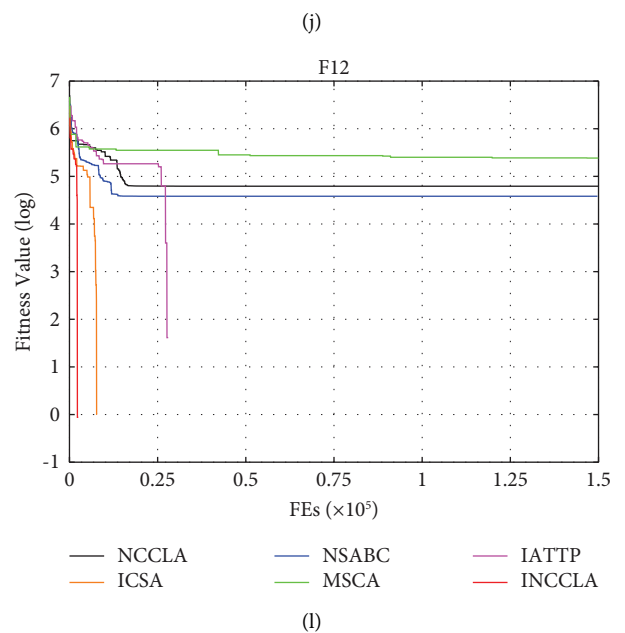
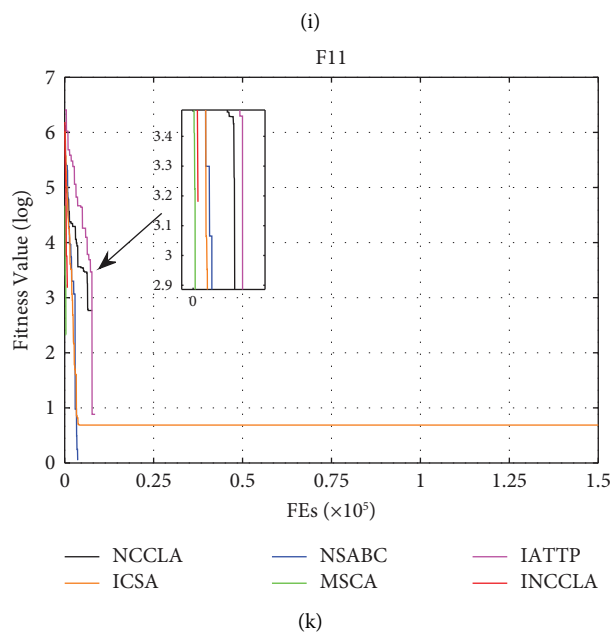
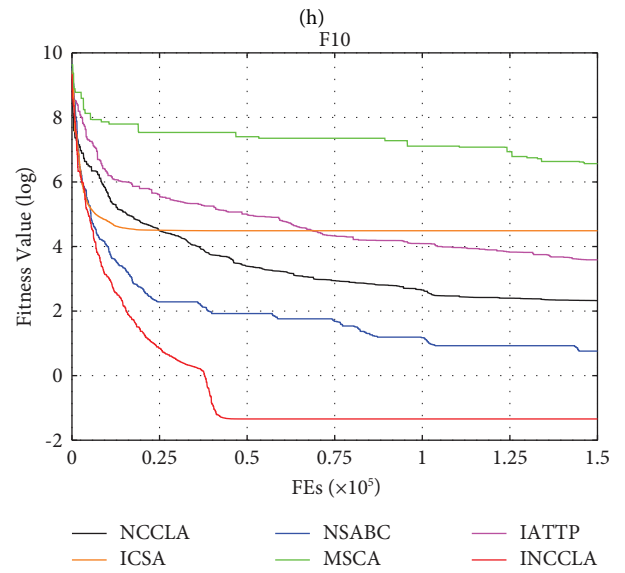
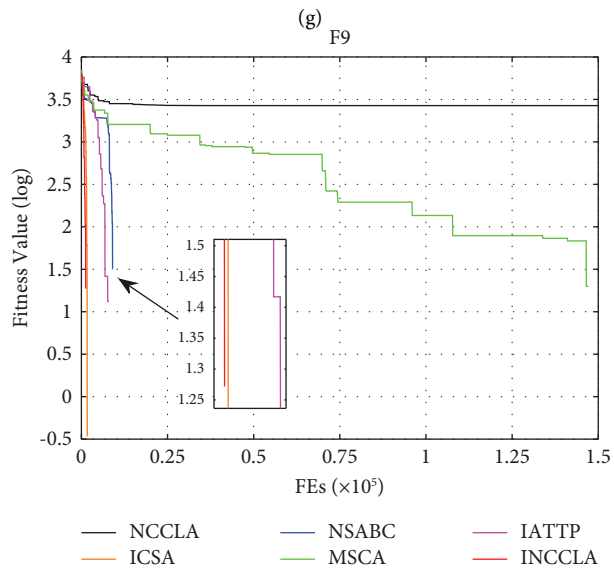
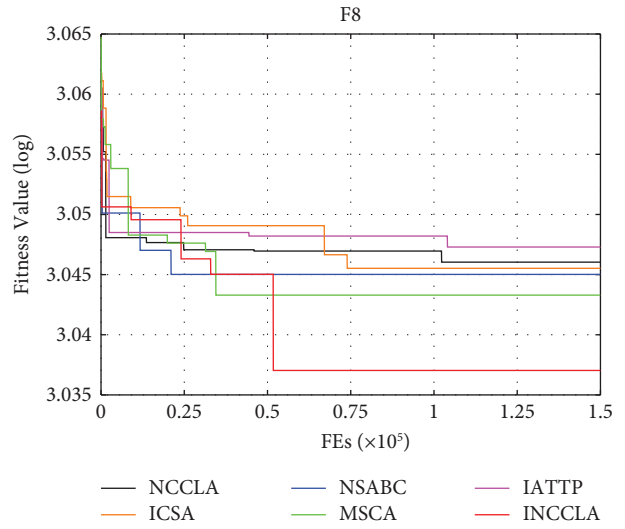
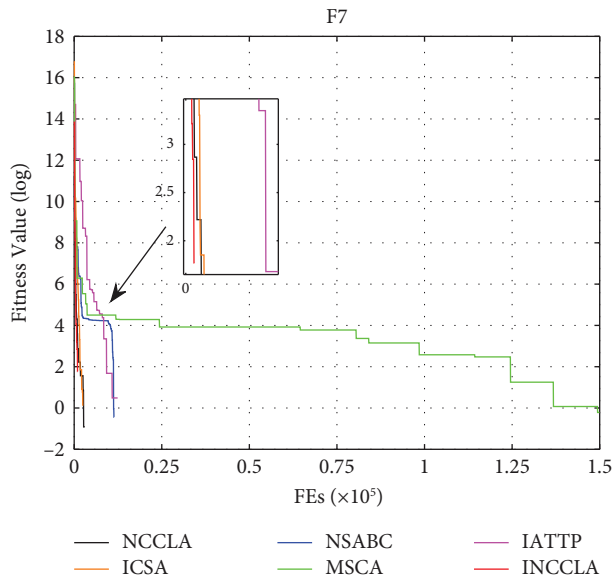
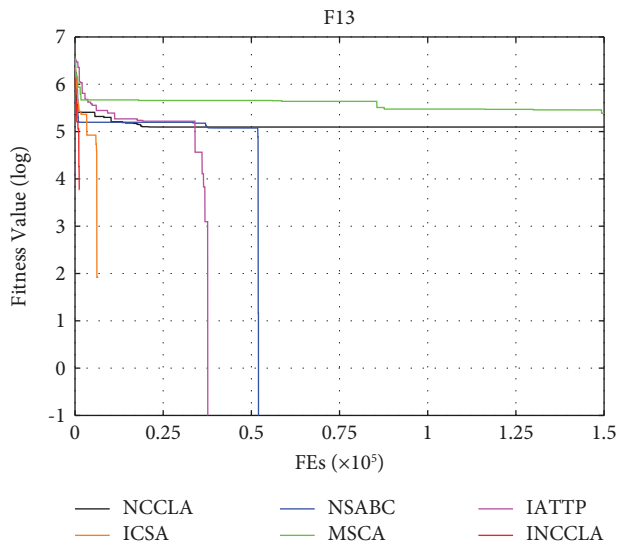
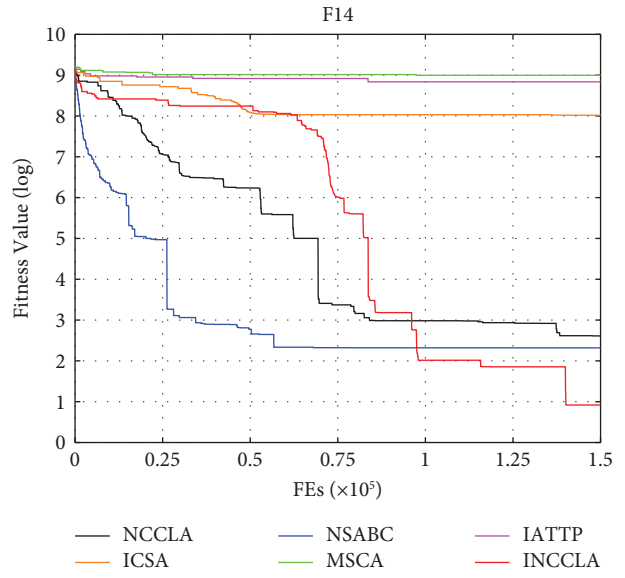


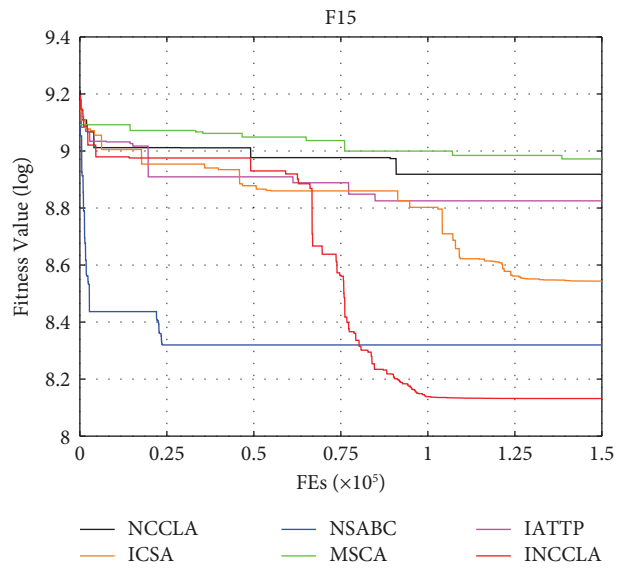
FIGURE 4: Continued.



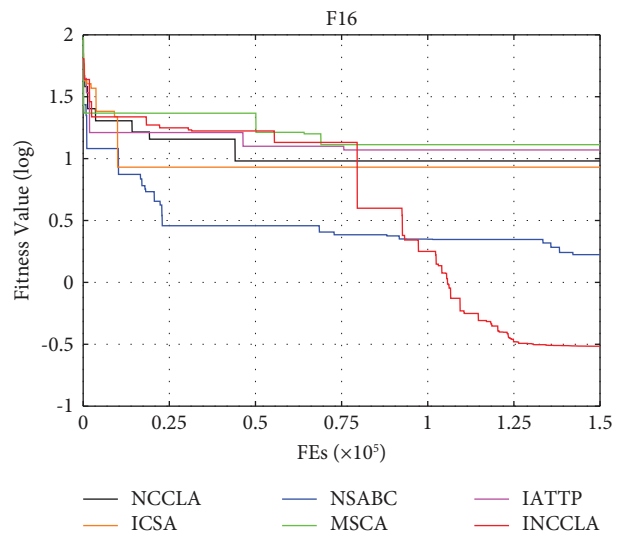
(m)



(n)



(o)



(p)

FIGURE 4: Continued.

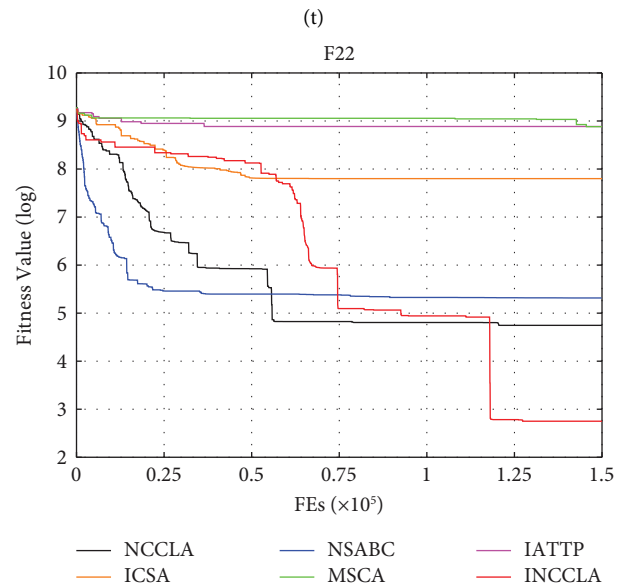
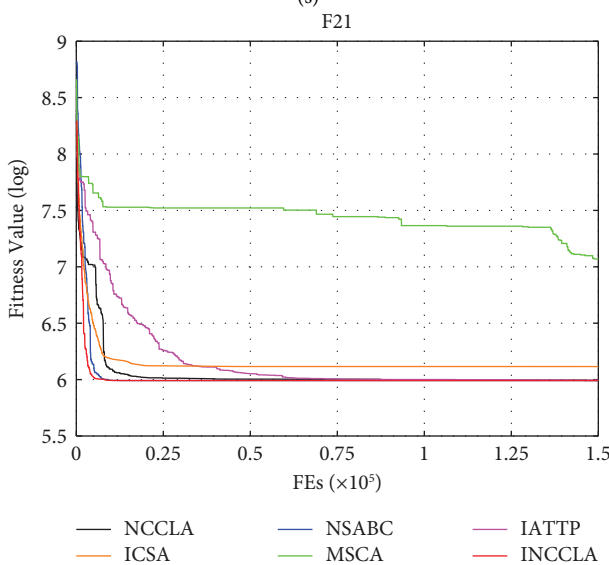
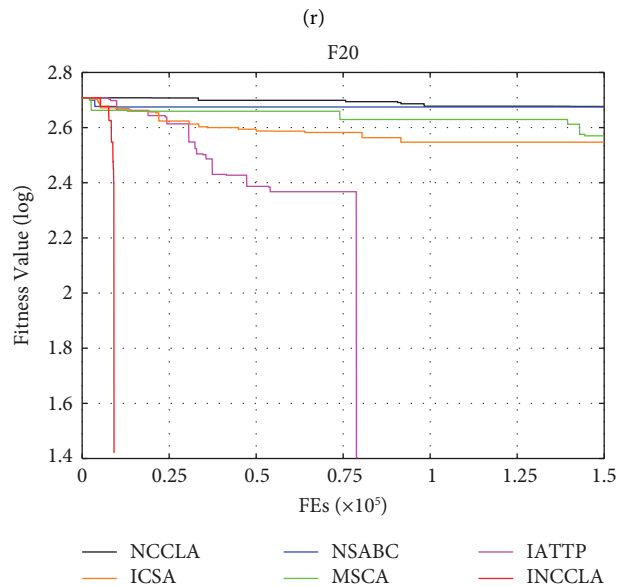
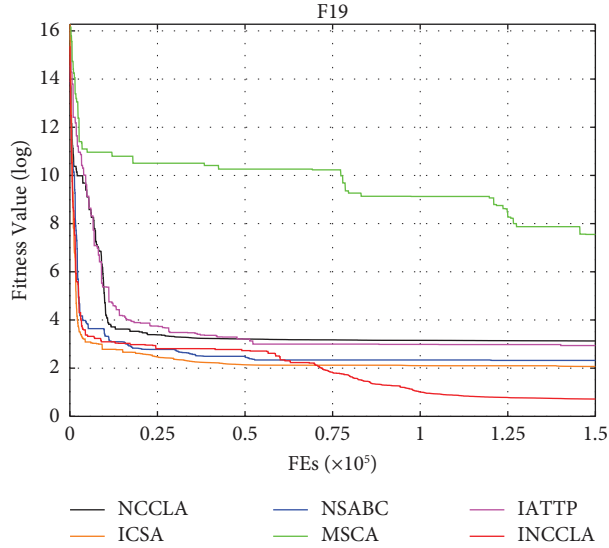
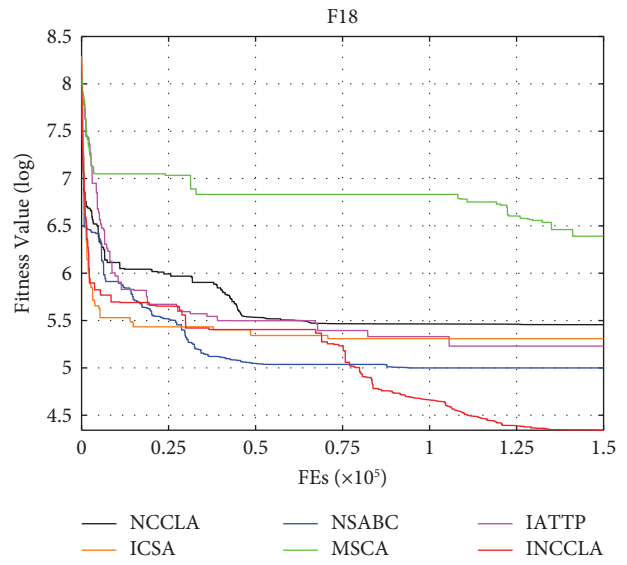
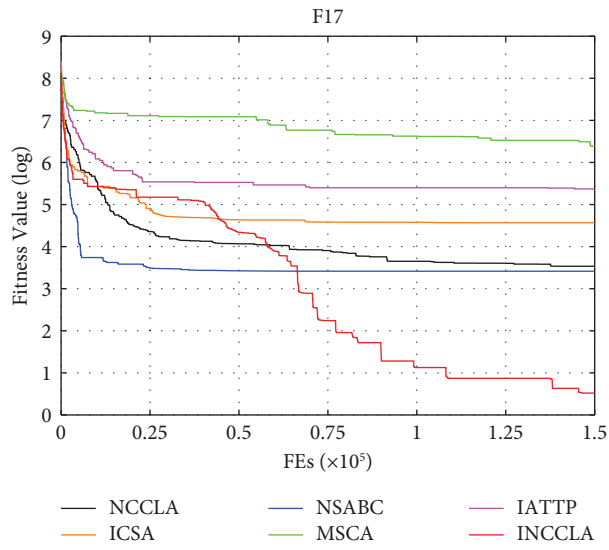


FIGURE 4: Continued.

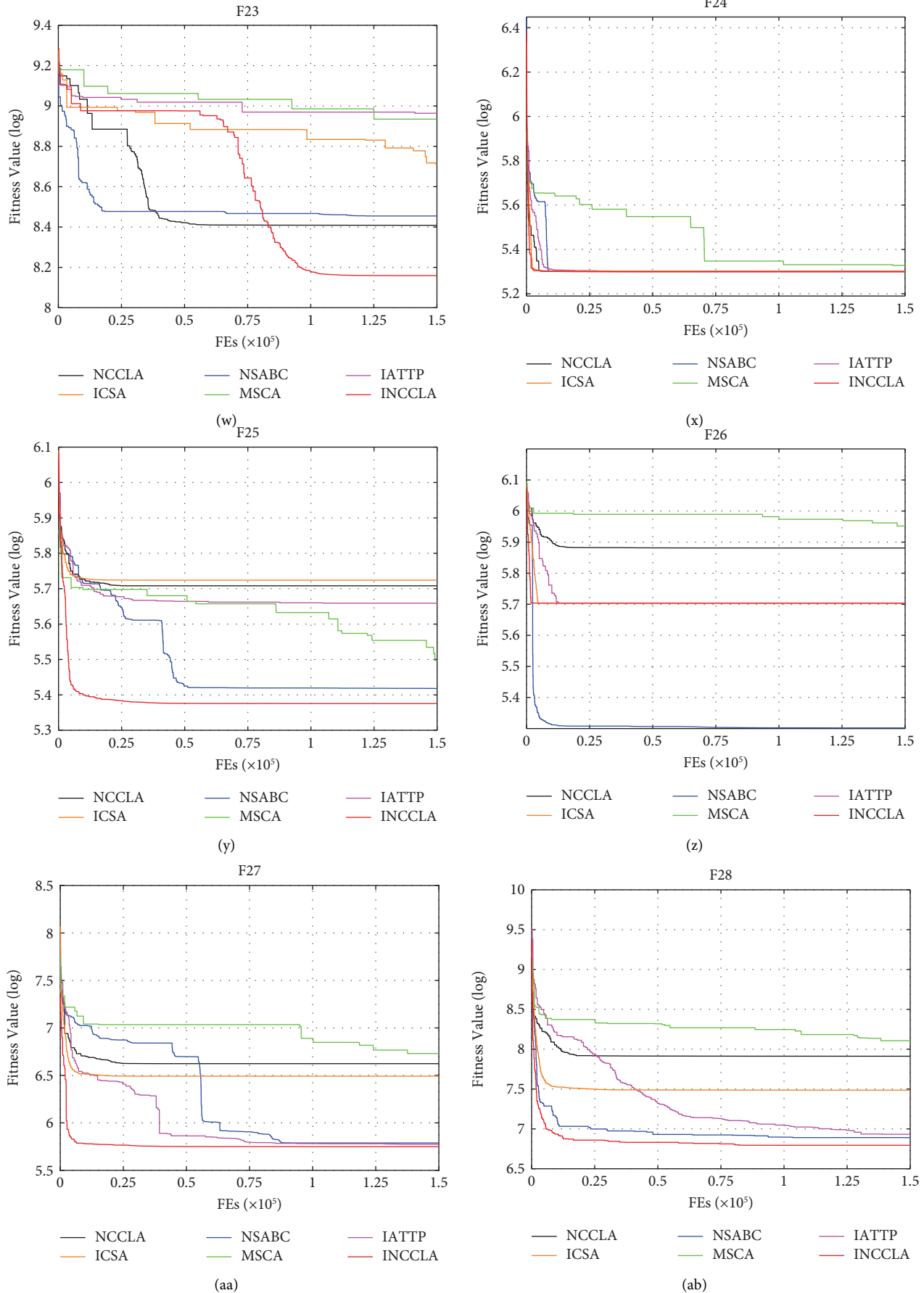


FIGURE 4: Convergence curves of each algorithm on the CEC2013 test suite. (a)-(ab) denote the 28 test functions $F1 \sim F28$ in the CEC2013 test set, respectively.

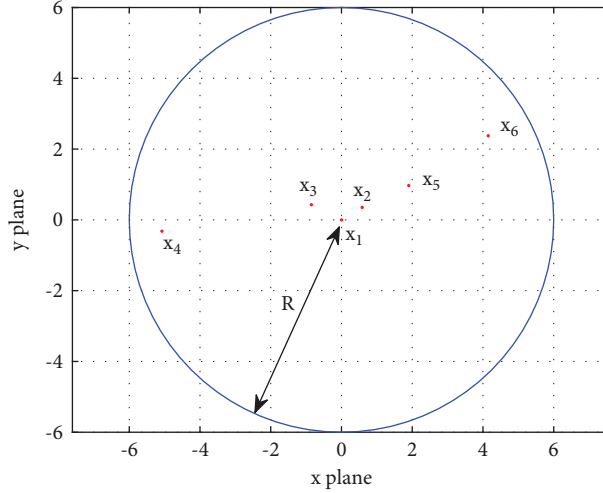


FIGURE 5: Distribution of synergy nodes.

influence of these three mechanisms ensures the overall advantage of INCCLA in terms of convergence precision and convergence speed. In the next step, we will work on further improving the convergence speed of the INCCLA algorithm in the pre-evolutionary stage while ensuring no significant change in the convergence precision.

4.4. Comparison of the Effects of Engineering Applications. To further compare the effectiveness of INCCLA with each comparison algorithm for engineering applications, this section will be validated by dealing with the collaborative beamforming optimization problem. The collaborative beamforming optimization problem is a typical problem in antenna arrays. The amplitude $\xi \in [0, 1]$ and phase $\alpha \in [-\pi, \pi]$ of the transmit signal weights of each collaborative node are used as decision variables in the algorithm optimization process, and the peak side valve level PSL minimization as shown in (27) is achieved by algorithm optimization.

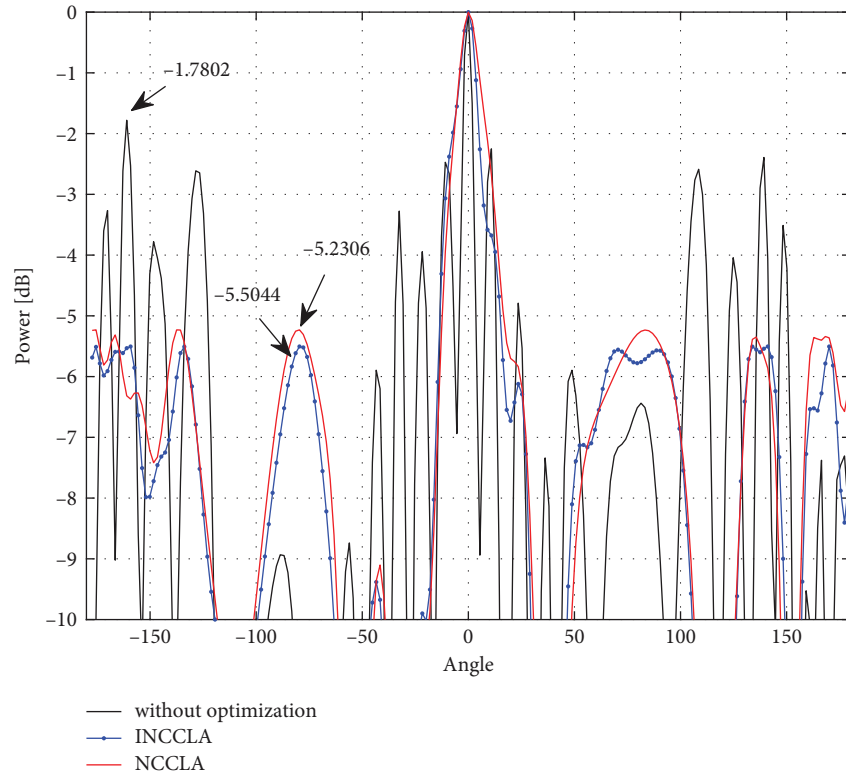
$$PSL = 20 \log_{10} \frac{\max |AF(\theta_{SL}, w)|}{AF(\phi, w)}, \quad (27)$$

where, $AF(\theta, w)$ represents the array factor, as shown in (28) and ϕ is the main beam direction. The positions of θ_{SL} can be found by finding all the peak points of the array factor (other than the main lobe's peak) for the domain $\theta \in [-\pi, \phi) \cup t(\phi, \pi]$, the denominator $AF(\phi, w)$ is the main beam power, which can be calculated as described in the literature [39], and molecule $\max |AF(\theta_{SL}, w)|$ is the maximum beam power in the side flap.

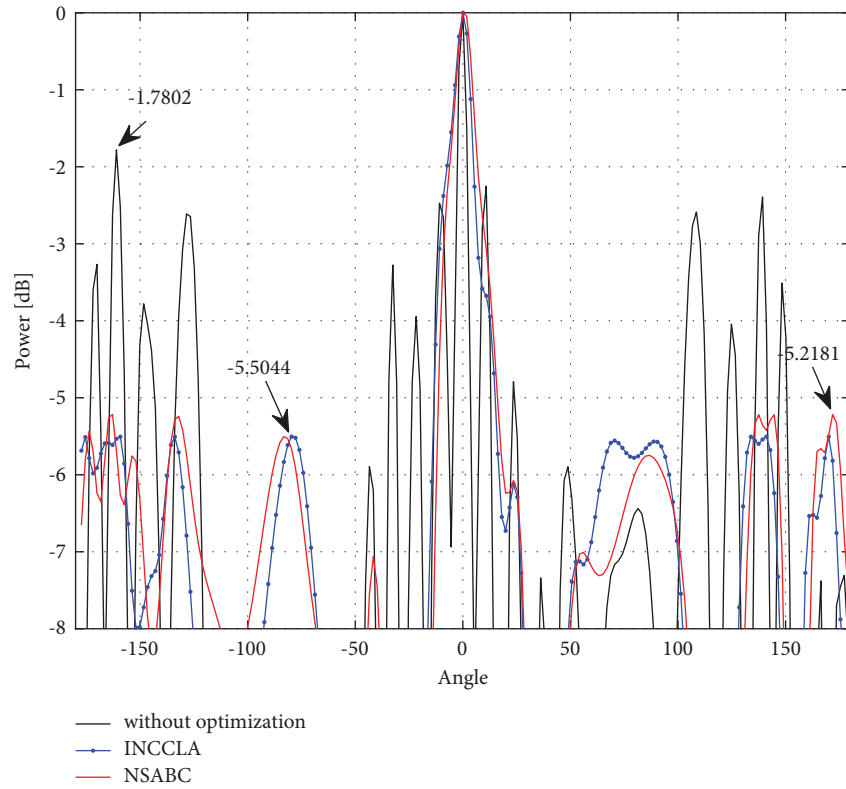
$$AF(\theta, w) = \sum_{k=1}^K w_k e^{j(2\pi/\lambda)r_k [\cos(\theta - \psi_k)]}. \quad (28)$$

This section examines the practical engineering optimization of INCCLA and the other comparative algorithms by optimizing the collaborative beamforming problem shown in Figure 5, in which the wavelength of the transmitting signal is λ . The six synergistic nodes are distributed in a circular domain of radius 6λ , where one synergistic node is located in the center of the circular domain.

The parameter settings of each algorithm are shown in Table 7. In order to avoid adverse effects of chance on algorithm evaluation, each algorithm was run 10 times

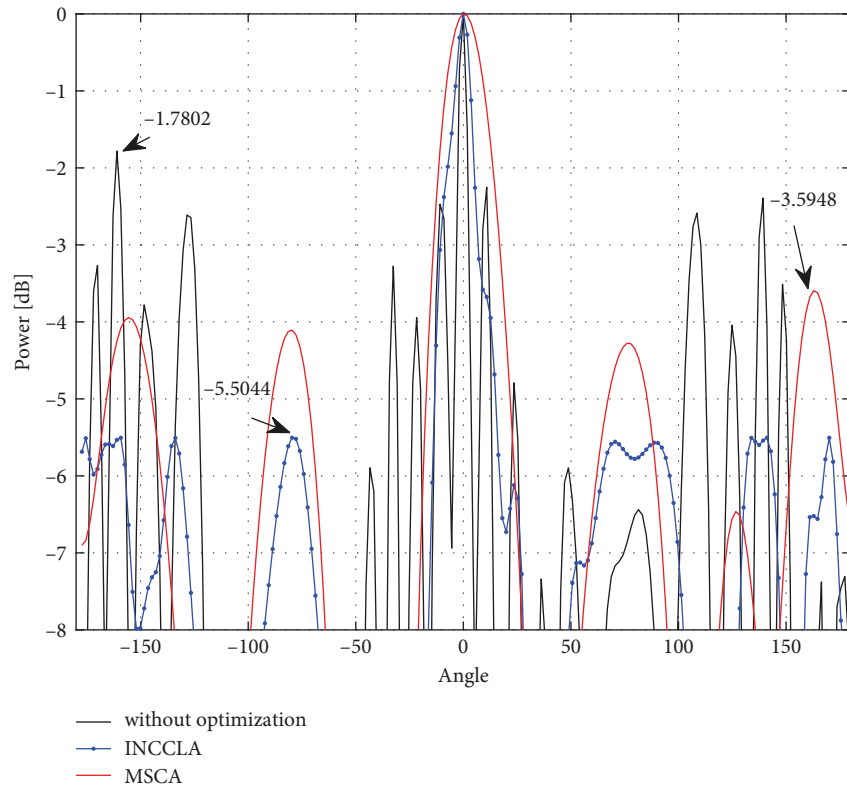


(a)

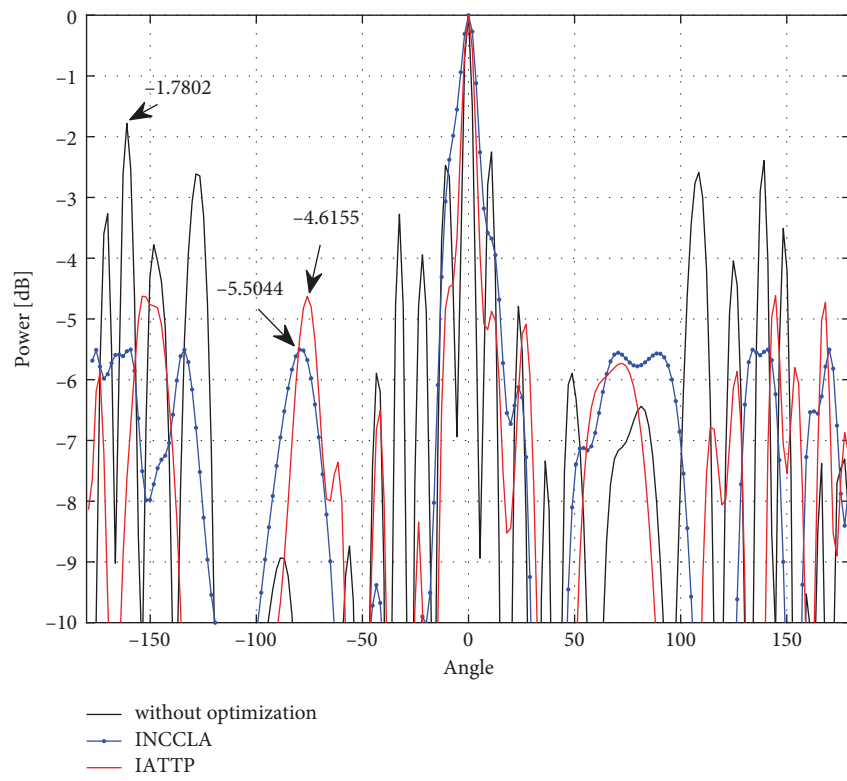


(b)

FIGURE 6: Continued.



(c)



(d)

FIGURE 6: Continued.

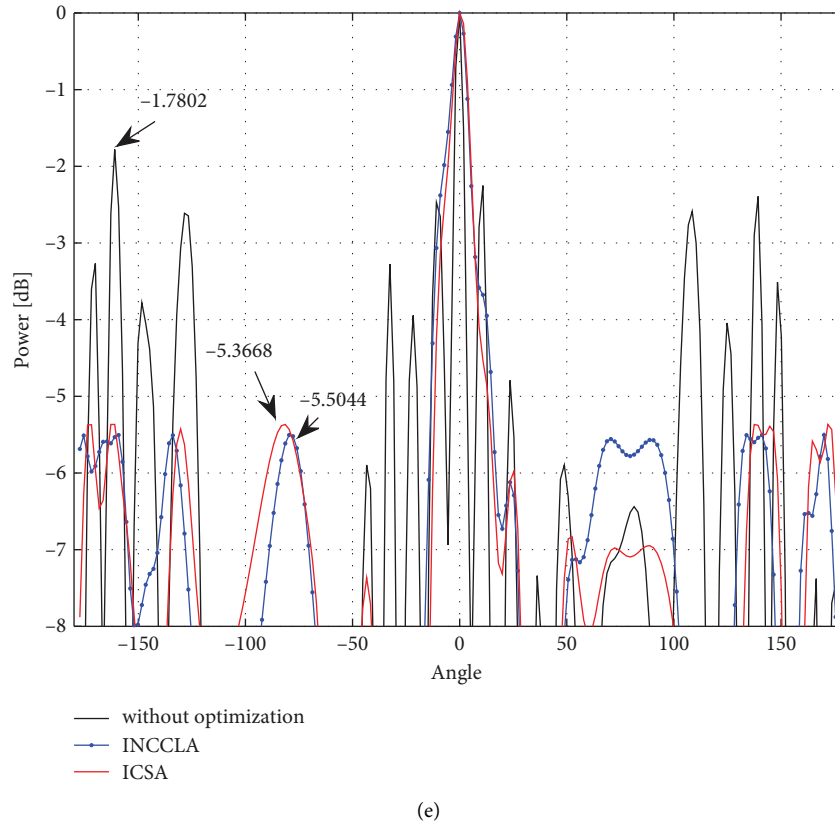


FIGURE 6: Beam diagram of INCCLA and each comparison algorithm in the cartesian coordinate system. (a)–(e) are the beam diagram of INCCLA with NCCLA, NSABC, MSCA, IATTP, and ICSA, respectively.

independently. The best median PSL obtained by each algorithm corresponding to the collaborative beam optimization scheme was selected for comparison. The beam diagram of INCCLA with each comparison algorithm in the Cartesian coordinate system is visually presented in Figure 6. The PSL obtained by each algorithm is labeled within the figure.

According to Figure 6, it can be seen that each algorithm achieves better collaborative beam optimization compared to the unoptimized algorithm, where the best PSL values obtained by INCCLA, NCCLA, NSABC, MSCA, IATTP, and ICSA are -5.5044 dB, -5.2306 dB, -5.2181 dB, -3.5948 dB, -4.1655 dB, and -5.3668 dB, respectively. Compared with the other five algorithms, the best PSL value obtained by INCCLA is the smallest and achieves the best collaborative beam optimization among the five algorithms. In summary, the proposed INCCLA also performs better in engineering applications.

5. Conclusions

This paper proposes an improved New Caledonian Crow Learning Algorithm (INCCLA) to improve further the convergence performance and the ability to escape from the local optimum of the NCCLA algorithm. First, INCCLA introduces cosine similarity in the parent selection phase. The selected parent can guide the juvenile crow to exploit different regions, maintaining the balance between population diversity and the convergence speed of the algorithm

in the evolutionary process. Second, INCCLA sets up a mixed learning mechanism of complete learning and incomplete learning, with the complete learning phase accelerating individual convergence and improving the convergence speed of the algorithm, and the incomplete learning phase increasing population diversity and enhancing the ability of the algorithm to escape from the local optimum. Finally, the juvenile crow reinforcement phase introduces weight factors and random perturbations to increase the global search and local exploration ability of the algorithm in the evolutionary process. The proposed parent reinforcement phase enhances the individual search capability of the parent and improves the overall performance of the algorithm. Experimental results on the CEC2013 and CEC2020 test suites show that the improvement strategy proposed in this paper effectively improves the overall performance of INCCLA, enabling the algorithm to maintain a balance between convergence speed and population diversity during the evolutionary process, and can achieve better results in most of the test functions. In addition, INCCLA is also more competitive in unimodal, multimodal, and composition problems compared with the other four optimization algorithms. INCCLA is also applied to the collaborative beamforming optimization problem, demonstrating the usefulness of INCCLA in engineering applications.

Although the INCCLA algorithm proposed in this paper has obvious advantages in terms of convergence accuracy, its

time complexity is slightly higher. In future work, the time complexity of the algorithm needs to be further reduced. In addition, it further broadens the application of the INCCLA algorithm in practical engineering fields.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Project of Scientific and Technological Innovation Development of Jilin in China under Grant 20210103090.

References

- [1] S. Aslan, H. Karaboga, and D. Karaboga, "Improved quick artificial bee colony (iqABC) algorithm for global optimization," *Soft Computing*, vol. 23, no. 24, pp. 13161–13182, 2019.
- [2] S. Shekhawat and A. Saxena, "Development and applications of an intelligent crow search algorithm based on opposition based learning," *ISA Transactions*, vol. 99, pp. 210–230, 2020.
- [3] D. Y. Tang, S. B. Dong, X. F. Cai, and J. Zhao, "A two-stage quantum-behaved particle swarm optimization with skipping search rule and weight to solve continuous optimization problem," *Neural Computing & Applications*, vol. 27, no. 8, pp. 2429–2440, 2016.
- [4] J. S. Liu, Y. N. Mao, X. Z. Liu, and Y. Li, "A dynamic adaptive firefly algorithm with globally orientation," *Mathematics and Computers in Simulation*, vol. 174, pp. 76–101, 2020.
- [5] A. Faramarzi, M. Heidarinejad, and S. Mirjalili, "Marine predators algorithm: a nature-inspired metaheuristic," *Expert Systems with Applications*, vol. 152, Article ID 113377, 2020.
- [6] W. G. Zhao, Z. X. Zhang, and L. Y. Wang, "Manta ray foraging optimization: an effective bio-inspired optimizer for engineering applications," *Engineering Applications of Artificial Intelligence*, vol. 87, Article ID 103300, 2020.
- [7] Y. J. Wang, P. Sun, X. M. Shi, and L. Zhang, "Green deployment method of micro base station for ultra-dense heterogeneous cellular networks based on constrained dolphin swarm algorithm," *IEEE Access*, vol. 8, pp. 5349–5362, 2020.
- [8] S. Gupta and K. Deep, "A novel random walk grey wolf optimizer," *Swarm and Evolutionary Computation*, vol. 44, pp. 101–112, 2019.
- [9] A. Sayed, M. Ebeed, Z. M. Ali et al., "A hybrid optimization algorithm for solving of the unit commitment problem considering uncertainty of the load demand," *Energies*, vol. 14, no. 23, p. 8014, Article ID 8014, 2022.
- [10] N. Habib Khan, R. Jamal, M. Ebeed, S. Kamel, H. Zeinoddini-Meymand, and H. M. Zawbaa, "Adopting Scenario-Based approach to solve optimal reactive power Dispatch problem with integration of wind and solar energy using improved Marine predator algorithm," *Ain Shams Engineering Journal*, vol. 13, no. 5, p. 101726, Article ID 101726, 2022.
- [11] M. Bastawy, M. Ebeed, A. Ali, M. F. Shaaban, B. Khan, and S. Kamel, "Optimal day-ahead scheduling in micro-grid with renewable based DGs and smart charging station of EVs using an enhanced manta-ray foraging optimisation," *IET Renewable Power Generation*, vol. 16, no. 11, pp. 2413–2428, 2022.
- [12] H. Samma, C. P. Mohamad Saleh, and J. Lim, "A new reinforcement learning-based memetic particle swarm optimizer," *Applied Soft Computing*, vol. 43, pp. 276–297, 2016.
- [13] G. Zhang, Y. M. Li, and Y. H. Shi, "Distributed learning particle swarm optimizer for global optimization of multimodal problems," *Frontiers of Computer Science*, vol. 12, no. 1, pp. 122–134, 2018.
- [14] J. W. Lu, J. Zhang, and J. N. Sheng, "Enhanced multi-swarm cooperative particle swarm optimizer," *Swarm and Evolutionary Computation*, vol. 69, p. 100989, Article ID 100989, 2022.
- [15] F. Wang, X. J. Wang, and S. L. Sun, "A reinforcement learning level-based particle swarm optimization algorithm for large-scale optimization," *Information Sciences*, vol. 602, pp. 298–312, 2022.
- [16] L. Z. Cui, G. H. Li, Y. L. Luo et al., "An enhanced artificial bee colony algorithm with dual-population framework," *Swarm and Evolutionary Computation*, vol. 43, pp. 184–206, 2018.
- [17] M. A. Awadallah, M. A. Al-Betar, A. L. Bolaji, E. M. Alsukhni, and H. Al-Zoubi, "Natural selection methods for artificial bee colony with new versions of onlooker bee," *Soft Computing*, vol. 23, no. 15, pp. 6455–6494, 2019.
- [18] X. Y. Zhou, Y. L. Wu, M. S. Zhong, and M. W. Wang, "Artificial bee colony algorithm based on multiple neighborhood topologies," *Applied Soft Computing*, vol. 111, p. 107697, Article ID 107697, 2021.
- [19] T. Y. Ye, W. J. Wang, H. Wang et al., "Artificial bee colony algorithm with efficient search strategy based on random neighborhood structure," *Knowledge-Based Systems*, vol. 241, p. 108306, Article ID 108306, 2022.
- [20] H. Wang, W. J. Wang, X. Y. Zhou et al., "Firefly algorithm with neighborhood attraction," *Information Sciences*, vol. 382–383, pp. 374–387, 2017.
- [21] Y. J. Sun, X. L. Wang, Y. H. Chen, and Z. J. Liu, "A modified whale optimization algorithm for large-scale global optimization problems," *Expert Systems with Applications*, vol. 114, pp. 563–577, 2018.
- [22] H. Zamani, M. H. Nadimi-Shahraki, and A. H. Gandomi, "CCSA: conscious neighborhood-based crow search algorithm for solving global optimization problems," *Applied Soft Computing*, vol. 85, p. 105583, Article ID 105583, 2019.
- [23] S. Gupta and K. Deep, "A memory-based grey wolf optimizer for global optimization tasks," *Applied Soft Computing*, vol. 93, p. 106367, Article ID 106367, 2020.
- [24] W. Long, M. Xu, J. J. Jiao, T. B. Wu, M. Z. Tang, and S. H. Cai, "A velocity-based butterfly optimization algorithm for high-dimensional optimization and feature selection," *Expert Systems with Applications*, vol. 201, p. 117217, Article ID 117217, 2022.
- [25] G. G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.
- [26] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. L. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.
- [27] S. M. Li, H. L. Chen, M. J. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020.

- [28] W. Al-Sorori and A. M. Mohsen, "New Caledonian crow learning algorithm: a new metaheuristic algorithm for solving continuous optimization problems," *Applied Soft Computing*, vol. 92, Article ID 106325, 2020.
- [29] A. W. Mohamed, A. A. Hadi, and A. K. Mohamed, "Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 7, pp. 1501–1529, 2020.
- [30] J. Z. Tu, H. L. Chen, M. J. Wang, and A. H. Gandomi, "The colony predation algorithm," *Journal of Bionics Engineering*, vol. 18, no. 3, pp. 674–710, 2021.
- [31] F. A. Hashim and A. G. Hussien, "Snake Optimizer: a novel meta-heuristic optimization algorithm," *Knowledge-Based Systems*, vol. 242, p. 108320, Article ID 108320, 2022.
- [32] J. J. Liang, B. Y. Qu, and P. N. Suganthan, "Problem definitions an evaluation criteria for the CEC2013 special session on real-parameter optimization," *Technical Report*, vol. 2012, 2013.
- [33] C. Yue, K. Price, and P. Suganthan, "Problem definitions and evaluation criteria for the CEC 2020 special session and competition on single objective bound constrained numerical optimization," *Zhengzhou Univ., Zhengzhou, China, Tech. Rep.*, Article ID 201912, 2019.
- [34] J. Derrac, S. García, D. Herrera, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011.
- [35] H. Wang, W. J. Wang, S. Y. Xiao, Z. H. Cui, M. Y. Xu, and X. Y. Zhou, "Improving artificial Bee colony algorithm using a new neighborhood selection mechanism," *Information Sciences*, vol. 527, pp. 227–240, 2020.
- [36] S. Gupta, K. Deep, S. Mirjalili, and J. H. Kim, "A modified sine cosine algorithm with novel transition parameter and mutation operator for global optimization," *Expert Systems with Applications*, vol. 154, p. 113395, Article ID 113395, 2020.
- [37] Y. P. Xiao, H. B. Chi, and Q. Q. Li, "An improved artificial tree algorithm with two populations (IATTP)," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104324, Article ID 104324, 2021.
- [38] J. Gholami, F. Mardukhi, and H. M. Zawbaa, "An improved crow search algorithm for solving numerical optimization functions," *Soft Computing*, vol. 25, no. 14, pp. 9441–9454, 2021.
- [39] S. Jayaprakasam, S. K. A. Rahim, and C. Y. Leow, "PSOGSA-Explore: a new hybrid metaheuristic approach for beam-pattern optimization in collaborative beamforming," *Applied Soft Computing*, vol. 30, pp. 229–237, 2015.

Research Article

Dynamic Optimization Modeling of Smart Tourism Information System Using VRGIS in Big Data Environment

Anfeng Xu ^{1,2} and Wenjun Zeng¹

¹Economics and Management School of Harbin University of Science and Technology, Harbin 150080, China

²Key Laboratory of Island Tourism Resource Data Mining and Monitoring, Ministry of Culture and Tourism, Sanya 572099, China

Correspondence should be addressed to Anfeng Xu; xuanfeng@hrbust.edu.cn

Received 21 July 2022; Revised 29 August 2022; Accepted 12 September 2022; Published 10 October 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Anfeng Xu and Wenjun 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.

The establishment of an intelligent, comprehensive, and all-encompassing information system for tourism management is the current trend in tourism informatization as a result of the continual development of modern information technology. Significant advancements in the field of VRGIS and its usage in research have been made as a result of the use of VRGIS to categorize, assess, plan, and manage tourism resources. The analysis of the recent development of VRGIS in tourism resource research is the first section of this work. This study examines and implements a mobile, computerized, and intelligent tourism service system that gives visitors a sense of the surrounding landscape using VRGIS. Three-dimensional mapping, environment detection, personal trajectory, and Weibo sharing are just a few of the system's many helpful features. While travelling, tourists can get services that are more intelligent and practical. The drawbacks of conventional geographic information systems include their reliance on sophisticated models, network limitations, and operational challenges. New software architecture is put in place to get rid of network restrictions, virtual reality peripherals are used to make operation more convenient, and system modeling is rebuilt using the TIN data model and model simplification. The results of experiments show improved refresh rates and peripheral expansion modules. The user experience is enhanced by this research.

1. Introduction

The tourism industry has become an essential driver of growth in a great number of nations, and the tourism resources that are available are an essential component of the material foundation on which the tourism sector is built and maintained [1, 2]. In order to provide a fundamental basis for the evaluation, planning, and reasonable protection of tourism resources, it is essential to recognize, classify, collect, and build a database of tourism resources. This is carried out in order to provide a fundamental basis. As a result of the one-of-a-kind qualities of tourism resources, the theory and methods of VRGIS are one-of-a-kind when it comes to spatial cognition and the design of classification systems, as well as the standardization of databases, as well as spatial analysis and evaluation analysis [3–5]. It is now an accepted

method and an integral part of the entire process of researching, developing, and making use of tourism resources that this practice has become a common place when carrying out research on tourism resources. The utilization of VRGIS in tourism research and application has important enlightenment significance for both the understanding of the theoretical and technical system and the guiding of the utilization of tourism resources[6].

Since the beginning of time, one of the most vital areas of research has been focused on the application of GIS in the process of developing and managing tourism resources. The study of tourism resources has made extensive use of GIS. In recent years, the tourism industry has begun to make use of geographic information systems, more commonly referred to as GIS. There is a strong connection between the theory and technology of GIS and the investigation and utilization

of tourism resources[7–9]. The vast majority of researchers working in this area have concentrated their attention on the function of GIS in the cultivation and application of tourism resources; however, a few of them have also carried out pertinent review research in this sector. There is still a lack of systematic application of GIS in the development and management of tourism resources, which leads to issues such as a lack of application and a lack of understanding of application fields. One solution to this problem is to increase the use of GIS in tourism resource development and management[10–12].

Due to the rapid advancement of information technology in the modern world, the current trend toward the informatization of tourism involves the extensive application of cutting-edge technologies such as VRGIS. Due to advancements in mobile Internet technology, high-speed networks are becoming more portable, which contributes to their growing popularity. The integration of smartphones with information search, communication, entertainment, and social networking provides users with access to dynamic information services and personalized recommendation services [13–17]. The traditional tourism information management system has a number of issues, such as a lack of usability flexibility, a one-dimensional information display, and a subpar user experience maintaining our forward momentum. Rapid advancements in virtual reality (VR) and augmented reality (AR) technologies in recent years have made it possible to show user-required information in multiple temporal and spatial dimensions, as well as to perform highly immersive interactive activities. New concepts for the future growth of the tourism industry have been inspired by these advances. VR offers customers a realistic and engaging visual experience through its visual sense [18, 19]. Virtual and physical aspects are combined in augmented reality (AR), which is based on VR. By including text, pictures, voiceover, and any other relevant information, you may help the user become more aware of and understand the real world. A new technology based on mobile platforms is starting to emerge as smart device performance keeps getting better: VRGIS technology. Additionally, mobile VRGIS is less confined by the space environment and is more versatile and portable than standard virtual-real integration. The VRGIS technology that is based on mobile platforms has been successfully used in a wide range of fields (including security, education, and urban management) [20, 21], and as a result, it provides an invaluable resource for resolving the issues that the tourism information service industry is currently experiencing. To accomplish the goals of a mobile VRGIS tourist system, mobile intelligent terminals must incorporate VRGIS technology [4, 22–24].

Traditional tourism is based on a two-dimensional map, but this type of vacation offers little in the way of experience or interaction, and there is room for improvement in the areas of information acquisition and spatial analysis. There are still some travel-related services that can only be accessed online, specifically via websites. Due to a lack of adaptability, real-time travel services are currently unavailable. Due to the fact that many modern systems only provide a single navigation or scene-reproduction function, they cannot be

user-friendly or interactive. This is because they only perform a single function. This study creates a VRGIS-based tourism system for mobile platforms with the intention of enhancing the system's usability and interactivity.

2. Background

The utilization of big data has made possible the opening of new lines of inquiry into the collection and analysis of data pertaining to conventional tourist destinations. Creating a tourism resource directory that is based on big data and developing a semantic rule extraction model for key attributes can significantly cut down on the amount of time spent out in the field. The tourism system performs a combined evaluation based on a particular weighting by utilizing a big data semantic analysis and statistical analysis, in addition to a big data expert scoring mechanism. Data from social media platforms that pertain to the travel of tourists, such as the latitude and longitude of each point of interest (POI), the content and travel notes of each POI, the number of travel comments received by each POI, and its score, are used to take part in the evaluation of tourism resources and to improve the scientific nature of public participation. Some researchers believe that the train collection head can be used to collect data about the popularity of tourist destinations within the comment network [7, 11]. This data can then be used to study the relationship between search popularity and comment popularity. The Internet is utilized in order to collect large amounts of information from a selected group of scholars, such as the number of comments and the overall rating of a tourist city, while analysis is utilized in order to determine the individual ratings of various types of tourism resources. After that, GIS software can be used to measure and assess the tourism industry's resources. Some researchers concentrate their attention on rural tourism resources and make use of geospatial big data in order to conduct a spatial classification of rural tourism within tourism development zones [4, 15, 17, 19, 21].

To account for the influence and contribution of these factors, it is necessary to consider the overall potential assessment of regional tourism resources, as well as spatial characteristics such as scale, scope, structure, spatial correlation, and pattern pedigree. It is necessary, when evaluating tourism resources, to improve comparability between different evaluation methods and to compare and optimize various algorithms and results. In addition, it is necessary to enhance the comparability of various evaluation methods.

It is important to evaluate the aggregation, structure, pattern, and distribution law of tourism resources in space when examining regional tourism resources. This includes factors such as the quantity and density, type and quality structure, and combination and spatial correlation of tourism resources within a specific spatial range, as well as other factors of a similar nature. Construction of a prototype from scratch: more than a dozen distinct GIS methods have been used to evaluate a region's tourism resources. Models for evaluating the availability of resources in gathering areas, agglomeration areas, and combination areas are included in these methods. These models consist of buffer and proximity

analysis, network and overlay analysis, and combination area analysis, among others. In tourism planning, some academics have utilized the GIS multiscale grid modeling technology to extract the spatial map of tourism resources, as well as significant points, lines, and areas of high-quality tourism resource areas. This has strengthened scientific planning.

The administrative unit scale and the tourist area scale are the two primary scales that are utilized when evaluating the tourism resources of a regional area. When carrying out spatial modeling, they only use statistical indicators as their basis for evaluating the tourism resources offered by the region. An examination of the manner in which tourism resources are dispersed geographically, as well as the spatial correlation between the resources' quality, will be carried out, along with an investigation into the methods for extracting the spatial map of the region's tourism resources and designing tourism resources for tourism products. There has not yet been any work carried out to create and implement potential value knowledge maps. To make matters even more confusing, no attempt has been made to take into account nonadministrative geographic areas, such as coastal zones or geomorphic units (such as mountains and valleys), in order to include tourism resources. This makes the situation even more difficult. If any consideration of the spatial relationship between tourism resources is given at all, it is in the form of a qualitative description. This is in stark contrast to the extremely limited attention that is paid to the issue. Combining traditional mathematical models with GIS data is the primary foundation upon which the evaluation method rests.

3. System Design

The system design is as follows.

3.1. Architecture Design. The system implements route planning and virtual roaming using three-dimensional landscape maps and geospatial data from tourist attractions. Users can do this to get a firsthand view of the scene and learn more about the route and beautiful sights before they journey there. The virtual environment, images, and text labels are flawlessly integrated into the real world seen by tourists through the use of location-based service technology and augmented reality technology. This is carried out to present visitors with a fresh viewpoint on cognition and an interactive application experience, two things that conventional software cannot offer. Figure 1 illustrates the system's architecture.

The user layer is the topmost layer. Some of the system's most crucial operations, such as virtual roaming and route planning, are made possible by the usage of a three-dimensional landscape map as a map base. The system is responsible for doing these tasks. Thanks to AR technology, the intelligent service module can recognize its surroundings and show a 3D model of an object. At the same time, a heat map depicting the movement of individuals through the picturesque region is created utilizing GPS positioning and data mining techniques. Users can share social media items such as their own footprints, those of their friends, and

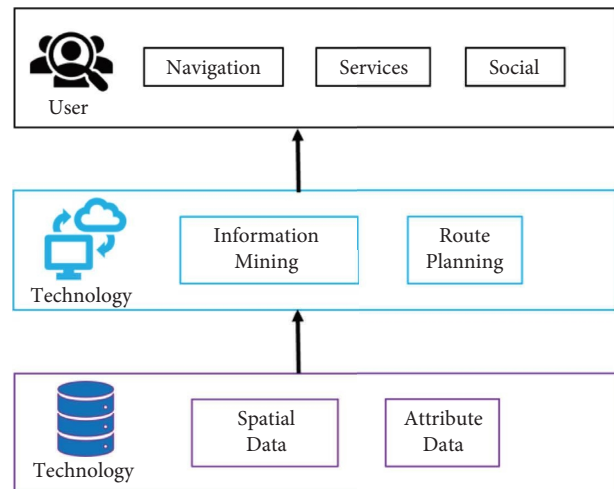


FIGURE 1: Structure of intelligent tourism management system.

connections to other apps such as Weibo through the social module, for the most part. The footprints of friends are another illustration of this kind of stuff.

Technical elements make up the second layer of this structure. In essence, the system's technology consists of two distinct components. With the use of location, multisensor tracking, and registration technology, spatial data may be obtained and analyzed. Opportunities for spatial analysis and spatial data mining arise as a result. GIS spatial analysis and data scheduling are used to execute three-dimensional virtual navigation and route planning on the map, which is based on a three-dimensional landscape map.

The data layer is the topmost layer. The system data centre is made up of two distinct parts: a geospatial database and a GIS spatial database of picturesque sites. The road network, trajectory, and position of scenic locations are all included in the GIS spatial database for scenic spots. One feature that can be used in data is text; additional examples include photographs and videos.

3.2. Geographic Information Modeling. This study makes use of it, and a panoramic camera is used to capture all of the images in an area as textures so that fitting processing can be carried out later. The main perspective of the observer is encompassed by the panoramic image, which enables the observer to experience what it is like to be immersed in the area and enhances the feeling of being immersed. A representation of the panoramic image observation model can be seen in Figure 2.

This research makes use of a digital elevation model, also known as a DEM, to model the uneven surface and the buildings because it is impossible to describe the area using a straightforward two-dimensional plan. In order to accomplish the goals of this investigation, a model consisting of irregular triangles was utilized. Because each point in the TIN model needs to be stored with its elevation coordinates and topology, the TIN model is more realistic than its DEM counterpart, which is one reason why it is more reliable in data storage. The representation of a region's topology as it

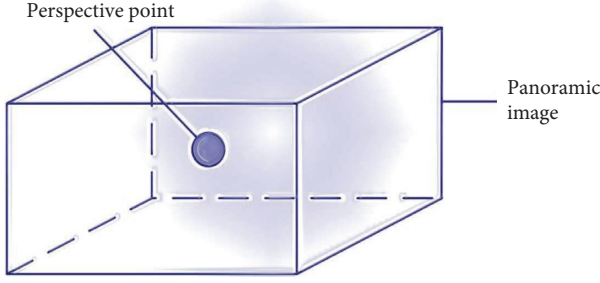


FIGURE 2: Panoramic image observation model.

appears in the TIN model is shown in Figure 3. This representation keeps track of every triangle, face, and point in the area.

An extension of the logistic regression classification known as the Softmax regression classification method is utilized in this study in order to better solve the multi-classification problem. This is due to the fact that the classification of human actions is a multiclassification that is mutually exclusive.

The area in Figure 3 is marked as follows:

$$ar = \{a, b, c, d, e, f, g, h\}. \quad (1)$$

Vertices in the graph are marked as follows:

$$ve = \{1, 2, 3, 4, 5, 6, 7, 8\}. \quad (2)$$

Then, we obtain the triangle file as shown in Table 1.

By eliminating insignificant points, lines, and surfaces from the model's geometric pattern, it is possible to cut down on the number of faces the model has. By utilizing the vertex deletion method and the progressive mesh simplification method, this can reduce the amount of running data required by the model, while simultaneously increasing the refresh rate. When deciding whether or not to remove the inner central vertex, one of the most important factors to take into account is how this will affect the image's position in the plane. Here is the formula you need to use to calculate the results:

$$F = |k(\text{Point}_i - C)|, \quad (3)$$

$$k = \frac{K}{|K|},$$

$$K = (\text{Point}_i - \text{Point}_{i+1})(\text{Point}_i - \text{Point}_j).$$

The calculation method of C is related to the simplified triangle a_i :

$$C = \frac{\sum x_i a_i}{\sum a_i}, \quad (4)$$

$$x_i = \frac{\text{Point}_5 + \text{Point}_1}{2},$$

$$a_i = \Delta \text{Point}_5 \text{Point}_1.$$

The sensor rotation matrix is as follows:

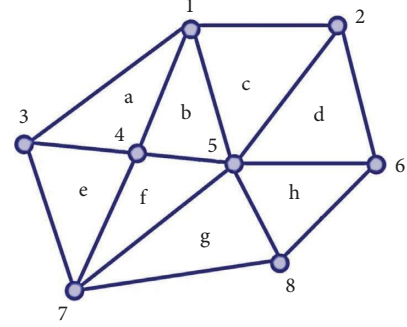


FIGURE 3: Topology diagram.

TABLE 1: Detail of triangle file.

Area point	Vertex			Collar triangle		
A	1	3	4	—	e	B
B	1	4	5	a	f	C
C	1	2	5	—	d	B
D	2	5	6	c	h	—
E	3	4	7	a	f	—
F	4	5	7	b	g	E
G	5	7	8	d	—	H
H	5	6	8	d	—	G

$$\begin{bmatrix} S(0) & S(1) & S(2) \\ S(3) & S(4) & S(5) \\ S(6) & S(7) & S(8) \end{bmatrix}. \quad (5)$$

The system makes use of tile map technology and displays the three-dimensional map of the landscape in Mercator projection. This is carried out so that the map can be seen in its full three-dimensional glory. First, a model of the Wutai Mountain scenic area is constructed with the help of virtual reality in order to generate an accurate landscape map. After that, you must tile the images after cutting the landscape map into squares, comparing and correcting them with the overall map that you downloaded from Baidu, and then finally tiling them. Second, the indexed and published cut map tiles are arranged using the WMS interface standard. The server side will return the tile data that corresponds to the received data range once it has parsed the data that was transmitted by the mobile terminal. The map can now be completely scheduled by the mobile terminal as a consequence of the real-time splicing and rendering that has taken place. The coordinates are then transformed, as the last step. In order to ensure that the actual latitude and longitude are used for navigation, a formula needs to be used to perform a transformation on them before they can be projected onto a map of a three-dimensional landscape.

Using VRGIS spatial analysis and the Freud shortest path algorithm, this system implements the navigation function. In addition to the conventional two-dimensional vector map navigation, it provides users with a mode for virtual roaming. In addition, the system dynamically adjusts the rotation angle of the map so that it corresponds with the user's current direction and range on the mobile terminal

screen beneath the top view of the three-dimensional landscape map, which is monitored in real time by the mobile sensor. In the context of the three-dimensional landscape map, this is visible beneath the map's top view.

This system employs the outdoor 3D registration technology of mobile terminals in order to realize the authenticity and real-time capabilities of the mobile virtual reality system. This technology is based on the mobile augmented reality module's positioning recognition. This technology enhances the real-time capabilities, stability, and robustness of 3D environment registration, along with the authenticity and fusion of the virtual-real combination scene. It focuses primarily on the spatial positioning of the camera and the placement of the virtual object in actual space. Based on these details, the rendering of the virtual object is completed.

4. Results

Utilizing a distributed system architecture, the system deploys gateway services, map data services, and scenic spot information data services on multiple Internet cloud servers. As a direct result, the system will be able to accommodate the exponential increase in user numbers. Distributed system operation is greatly aided by cohesion and transparency that are at a high level. The network can be used to establish a distributed cluster framework that grants each service node a high level of autonomy. The cluster framework has fewer moving parts, which makes it easier to maintain and simpler to operate. As a direct result, from the user's perspective, the distribution of each node to the user's application is completely obscured. The distributed network architecture eliminates the need for users to be concerned with complex relationship structures and synchronization requirements when accessing data that is distributed across multiple locations. Using the gateway service, a master control server will distribute a large number of service requests to individual servers when it receives a large number of service requests. This will decrease the amount of computational stress the system experiences.

This mode displays a three-dimensional landscape map of the entire Wutai Mountain scenic area. Users may zoom in, move around, and access more information in this mode. The landscape map may be rotated while you are moving around, and the labels for each picturesque spot will do the same. The user's perspective remains unchanged from the frontal angle when they enter roaming mode. The designation of the beautiful area serves as the starting point for route planning, as shown in Figures 4 and 5. Additionally, it is possible to plan routes to a variety of beautiful places, and if desired, the tour route can be researched beforehand. In particular, the system we have created has descriptions of attractions in both Chinese and English.

In the environment recognition module, users can click on a specific scenic spot on the map, and the system will automatically pop up the specific picture and introduction in Chinese and English of the scenic spot. Users can also upload photos of the place, as well as make reviews and ratings.

In order to produce information-rich queries, prosperous scenic spots, key monitoring areas, and important



FIGURE 4: 3D map display.

landmark buildings in the vicinity of the scenic spots are used to conduct specific queries. These queries are based on the scenic spots that are important to the user's interests and are used to conduct specific queries. Through the standardized processing of queries such as landmark queries, business district queries, catering queries, entertainment place queries, traffic route queries, fire protection facility attribute queries, water conservation facility attribute queries, and business data queries, it is possible to integrate the business management of 3D visualizations of scenic spots into 3D cities, as shown in Figure 6.

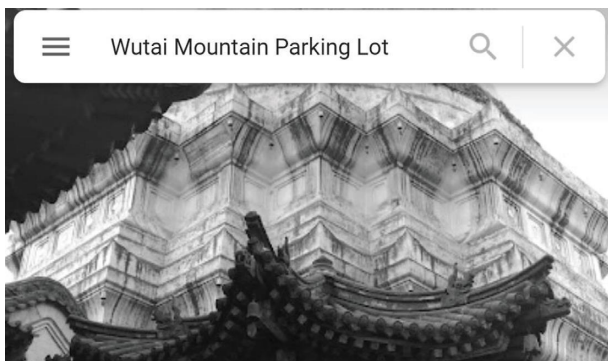
In the social module, users can share their route maps, tourist photos, etc. The tourism system also supports other contact APP logins, saving users the trouble of registering separately, as shown in Figure 7.

5. Discussion

Participation by the tourism industry and the general public in VRGIS projects, as well as the number of tourism systems employing this technology, are on the rise. According to the findings of previous studies, academics who work in the field of tourism and geographic information systems (GISs) appear to be the primary sources of information used to analyze and evaluate tourism resources. The latter emphasizes the evaluation's content in addition to its outcomes. When evaluating or analyzing tourism resources, the analytic hierarchy process (AHP) places a greater emphasis on spatial thinking patterns and technological innovations than the national standard method, which is the more prevalent technique. In recent years, an increasing number of VRGIS researchers have focused their efforts on tourism resource-related research. This type of research examines census technology for tourism resources, tourism resource analysis and evaluation methods, tourism resource development planning, and technology for analyzing big data in tourism.



FIGURE 5: Route navigation map.



Wutai Mountain Parking Lot

FIGURE 6: Specific classic search module.



FIGURE 7: Share module.

5.1. Developing and Perfecting the Research Methodology. As the growth of tourism resources' transitions from a key development stage to a comprehensive development stage, the rise of global tourism, and the iterative update of intelligent and portable mobile terminal equipment, an increasing number of industries, businesses, and members of

the general public will become involved in the recognition of tourism resources.

The method system can be innovatively enhanced with the aid of VRGIS and a tourism knowledge model. Future tourism resource big data research hotspots will include scientific tourism resource spatial cognitive models, tourism resource graph knowledge models, tourism resource development case models, and tourist destination selection behavior models. There will be a new and improved method for evaluating the available tourism resources, which will be based on a set of semantic rules derived from spatial thinking, as well as relevant maps.

In Section 1 of this study, we will examine the rapid growth of VRGIS in tourism resource research over the past several years. Mount Wutai, a Buddhist holy site, is used as an example to illustrate how VRGIS can be used to develop a mobile, informaticist, intelligent, and personalized tourism service system that provides tourists with a better understanding of the surrounding landscape.

5.2. An Original Perspective and Method of Interacting with the Outside World. The system consists of a three-dimensional landscape map, recognition of the surrounding environment, tracking of a user's trajectory, and the capacity to share information via Weibo. On the road, tourists now have access to services that are both more knowledgeable and more convenient. Traditional geographic information systems have several drawbacks, including operational challenges that are difficult to use, complex model structures, and limited network capacities. System modeling is reconstructed with the aid of the TIN data model and model simplification; a new software architecture is implemented to eliminate network constraints; virtual reality peripherals are utilized to make operation more straightforward and convenient. Experiments conducted for this study demonstrate that this work enhances the user experience by

increasing the number of peripheral expansion modules and the system's refresh rate.

Data Availability

The data used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by (1) Social Science Foundation of Heilongjiang Province "research on the digital platform ecosystem of the tourism industry in Heilongjiang province" (21GLB064), (2) Scientific Research Project of High-Quality Development of Tourism, Culture, Radio, Film, and Sports Industry in Hainan Province (HNLW(ZC)22-23) "practical exploration of the concept of metaverse and its impact on culture and tourism development," and (3) key laboratory of island tourism resource data mining and monitoring, Ministry of Culture and Tourism (KLITRDMM 2022–20).

References

- [1] Y. Liao, "Research and implementation of the intelligent tourism system based on VRGIS," in *Proceedings of the international conference on cyber security intelligence and analytics*, pp. 130–135, Haikou, China, February 2019.
- [2] M. Beck, "Real-time visualization of big 3D city models," *International Archives of the Photogrammetry Sensing and Spatial Information Sciences*, vol. 34, no. 2, 2003.
- [3] W. Höhl, "Official survey data and virtual worlds—designing an integrative and economical open source production pipeline for xR-applications in small and medium-sized enterprises," *Big Data and Cognitive Computing*, vol. 4, no. 4, p. 26, 2020.
- [4] M. N. Kamel Boulos, Z. Lu, P. Guerrero, C. Jennett, and A. Steed, "From urban planning and emergency training to Pokémon Go: applications of virtual reality GIS (VRGIS) and augmented reality GIS (ARGIS) in personal, public and environmental health," *International Journal of Health Geographics*, vol. 16, no. 1, pp. 7–11, 2017.
- [5] X. Liu, Y. Zhang, W. Cao, and Y. Zhu, "Design and realization of a VRGIS-based digital agricultural region management system," in *Proceedings of the International Conference on Computer and Computing Technologies in Agriculture*, pp. 446–455, Springer, Berlin, Heidelberg, October 2010.
- [6] T. Templin, D. Popielarczyk, and M. Gryszko, "Using augmented and virtual reality (AR/VR) to support safe navigation on inland and coastal water zones," *Remote Sensing*, vol. 14, no. 6, p. 1520, 2022.
- [7] S. M. C. Loureiro, J. Guerreiro, and F. Ali, "20 years of research on virtual reality and augmented reality in tourism context: a text-mining approach," *Tourism Management*, vol. 77, Article ID 104028, 2020.
- [8] S. Borović and I. Marković, "Utilization and tourism v of geothermal waters in Croatia," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 52–63, 2015.
- [9] T. Bahaire and M. Elliott-White, "The application of geographical information systems (GIS) in sustainable tourism planning: a review," *Journal of Sustainable Tourism*, vol. 7, no. 2, pp. 159–174, 1999.
- [10] R. N. Okech, "Effective utilization of human resources in Kenya: case of tourism entrepreneurial opportunities," *Journal of Human Resources in Hospitality & Tourism*, vol. 6, no. 2, pp. 43–57, 2007.
- [11] J. Mango, E. Çolak, and X. Li, "Web-based GIS for managing and promoting tourism in sub-Saharan Africa," *Current Issues in Tourism*, vol. 24, no. 2, pp. 211–227, 2021.
- [12] S. Afnarius, F. Akbar, and F. Yuliani, "Developing web-based and mobile-based GIS for places of worship information to support halal tourism: a case study in bukittinggi, Indonesia," *ISPRS International Journal of Geo-Information*, vol. 9, no. 1, p. 52, 2020.
- [13] T. Huang, J. C. Xi, and Q. S. Ge, "Spatial differentiation and integration optimization of an urban agglomeration tourism system under the influence of high-speed railway network evolution," *Applied Spatial Analysis and Policy*, vol. 12, no. 2, pp. 349–376, 2019.
- [14] H. T. T. Hoang, Q. H. Truong, A. T. Nguyen, and L. Hens, "Multicriteria evaluation of tourism potential in the central highlands of vietnam: combining geographic information system (GIS), analytic hierarchy process (AHP) and principal component analysis (PCA)," *Sustainability*, vol. 10, no. 9, p. 3097, 2018.
- [15] A. Zhang, Y. Yang, T. Chen, J. Liu, and Y. Hu, "Exploration of spatial differentiation patterns and related influencing factors for National Key Villages for rural tourism in China in the context of a rural revitalization strategy, using GIS-based overlay analysis," *Arabian Journal of Geosciences*, vol. 14, no. 2, pp. 83–15, 2021.
- [16] S. Kang, G. Lee, J. Kim, and D. Park, "Identifying the spatial structure of the tourist attraction system in South Korea using GIS and network analysis: an application of anchor-point theory," *Journal of Destination Marketing & Management*, vol. 9, pp. 358–370, 2018.
- [17] D. Omarzadeh, S. Pourmoradian, B. Feizizadeh, H. Khallaghi, A. Sharifi, and K. V. Kamran, "A GIS-based multiple eco-tourism sustainability assessment of West Azerbaijan province, Iran," *Journal of Environmental Planning and Management*, vol. 65, no. 3, pp. 490–513, 2022.
- [18] J. Zhang, W. Feng, T. Yuan, J. Wang, and A. K. Sangaiah, "SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking," *Applied Soft Computing*, vol. 118, Article ID 108485, 2022.
- [19] J. Zhang, J. Sun, J. Wang, Z. Li, and X. Chen, "An object tracking framework with recapture based on correlation filters and Siamese networks," *Computers & Electrical Engineering*, vol. 98, Article ID 107730, 2022.
- [20] E. Eftekhari and M. Mahdavi, "Land suitability assessment using ANP in a GIS environment for tourism development site (case study: lavasan-e Kuchak Rural District, Tehran Province, Iran)," *Journal of Tourism Hospitality Research*, vol. 7, no. 1, pp. 5–17, 2019.
- [21] M. Lazoglou and D. C. Angelides, "Development of a spatial decision support system for land-use suitability assessment: the case of complex tourism accommodation in Greece," *Research in Globalization*, vol. 2, Article ID 100022, 2020.
- [22] K. Prueksakorn, J. C. Gonzalez, J. Keson, S. Wongsai, N. Wongsai, and P. Akkajit, "A GIS-based tool to estimate carbon stock related to changes in land use due to tourism in

- Phuket Island, Thailand,” *Clean Technologies and Environmental Policy*, vol. 20, no. 3, pp. 561–571, 2018.
- [23] J. Dey, S. Sakhre, V. Gupta et al., “Geospatial assessment of tourism impact on land environment of Dehradun, Uttarakhand, India,” *Environmental Monitoring and Assessment*, vol. 190, no. 4, pp. 181–210, 2018.
- [24] C. Lincaru, S. Pirciog, D. Atanasiu, C. Stroe, V. Ciucă, and A. Grigorescu, “Patterns of mainly tourism sectors at local level by employee’s characteristics using gis multivariate clustering analysis-Romania case study,” *Regional Science Inquiry*, vol. 12, pp. 261–290, 2020.

Research Article

A Novel Affective Analysis System Modeling Method Integrating Affective Cognitive Model and Bi-LSTM Neural Network

Zhuqing Yang ¹, Liya Zhou,² and Zhengjun Jing³

¹Jiangsu Vocational College of Information Technology, Wuxi, Jiangsu 214153, China

²Changshu Binjiang Vocational and Technical School, Changshu, Jiangsu 215512, China

³Jiangsu University of Technology, Changzhou, Jiangsu 213001, China

Correspondence should be addressed to Zhuqing Yang; yangzq@jsit.edu.cn

Received 6 September 2022; Revised 19 September 2022; Accepted 20 September 2022; Published 7 October 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Zhuqing 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.

The severity of mental health issues among college students has increased over the past few years, having a significant negative impact on not only their academic performance but also on their families and even society as a whole. Therefore, one of the pressing issues facing college administrators right now is finding a method that is both scientific and useful for determining the mental health of college students. In pace with the advancement of Internet technology, the Internet has become an important communication channel for contemporary college students. As one of the main forces in the huge Internet population, college students are at the stage of growing knowledge and being most enthusiastic about new things, and they like to express their opinions and views on study life and social issues and are brave to express their emotions. These subjective text data often contain some affective tendencies and psychological characteristics of college students, and it is beneficial to dig out their affective tendencies to further understand what they think and expect and to grasp their mental health as early as possible. In order to address the issue of assessing the mental health of college students, this study makes an effort to use public opinion data from the university network and suggests a college student sentiment analysis model based on the OCC affective cognitive model and Bi-LSTM neural network. In order to label three different types of positive, negative, and neutral sentiment on the microblog text of college network public opinion, we first design a sentiment rule system based on the OCC affective cognition elicitation mechanism. In order to effectively and automatically identify the sentiment state of college students in the network public opinion, this study uses a Bi-LSTM neural network to classify the preprocessed college network public opinion data. Finally, this study performs comparison experiments to confirm the validity of the Bi-LSTM neural network sentiment recognition algorithm and the accuracy of the OCC sentiment rule labeling system. The findings show that the college student sentiment recognition effect of the model is significantly enhanced when the OCC sentiment rule system is used to label the college network public opinion data set as opposed to the naturally labeled data set. In contrast to SVM and other classification models like CNN and LSTM, the Bi-LSTM neural network-based classification model achieves more satisfactory classification results in the recognition of college opinion sentiment.

1. Introduction

In the modern era, issues related to mental health have emerged as an increasingly prevalent public concern associated with significant risks to society [1]. College students, particularly in recent years and especially as a result of the influence of COVID-19, are being subjected to an increasing amount of pressure in the forms of higher education, employment, and competition, all of which contribute to the

emotional problems and mental health problems that students experience. College students frequently struggle with a variety of mental health issues, including anxiety, depression, inferiority complex, and interpersonal sensitivity. What's more, a significant number of them have entertained the idea of ending their own lives [2]. There is no doubt that this will have serious adverse effects not only on individuals but also on families and even on society as a whole. If the emotional tendencies and psychological state of the students

can be detected early on, the school will be able to provide timely and specific assistance to students who are struggling, which will reduce the amount of damage caused. As a consequence of this, it is extremely valuable to discover an efficient method to identify college students who have issues related to their mental health. People's use of the Internet as a medium for the transmission of information and communication is becoming increasingly widespread as the Internet continues to gain in popularity. College students make up a sizeable portion of the population of people who use the Internet in China. Due to the anonymity and equality offered by microblogging platforms, college and university students are increasingly turning to them as a means of conveying their feelings, elaborating on their positions, and articulating their requirements. Adolescents and college students are especially susceptible to the influence of Internet public opinion because they make up the largest segment of Internet users. This has a negative impact on the students' emotional tendencies as well as their mental health. Only by mining and analyzing the sentiment tendencies of college students in college network public opinion can we understand college students' mental health status in a timely manner and provide targeted treatment and prevention [3]. This is because only by doing so can we grasp college students' sentiment tendency in an all-encompassing manner. As a consequence of this, the analysis of the sentiment of college students in network public opinion needs to be strengthened in order to guarantee the psychological well-being of college students and to preserve the harmonious and stable growth of colleges and universities.

However, due to the large number of subjective texts that are uploaded to the network platform, relying solely on traditional artificial means to analyze these texts will not only require a significant amount of manual labor and a significant amount of time, but it will also significantly reduce the accuracy of recognition. The theory of machine learning is constantly evolving, which has led to an increase in the number of academics who are applying machine learning algorithms to the study of sentiment recognition of university online public opinion. These scholars have found that their efforts have been fruitful [4, 5]. An opinion evolution system model based on complex agent networks was constructed by the literature [6] in order to analyze the influence of various factors within the model on the development of public opinion. This model was constructed in accordance with the characteristics and evolution rules of university public opinion generation. After that, simulation experiments are used to demonstrate the model's viability for use in actual situations. Literature [7] addressed the issue of sparse features in microblog text by employing a labeled LDA model to model microblog text. As a result, the semantic information contained within the text was enriched, and the text's ability to be classified was enhanced. In addition to this, it makes use of technology that aligns words in order to train the translation model, after which it applies this trained model to the university network public opinion analysis. Literature [8] performs an analysis of the characteristics of the text utilizing statistical technology in the form of a word cloud in order to obtain the high-frequency words

that were used in user comments. The next step is for them to narrow down the scope of high-frequency words and teach a naive Bayesian classifier in order to finish the text's classification of emotions. At long last, a display of the evolution map of public opinion in colleges and universities as it has developed within the context of microblogs is presented. However, the affective analysis technique that utilizes machine learning depends on the corpus domain, and the contextual knowledge of the context cannot be effectively used, and its classification performance is also affected to some extent. Then, the deep learning techniques improve this problem. It can automatically learn text affective information from a large number of samples and automatically perform feature representation, which provides new ideas for analysis of university opinion and classification of college students' sentiment. Literature [9] makes use of the extensive data set of IMDB film reviews and organizes its textual data into categories that are either positive or negative. After that, based on the data that have been preprocessed, they use a technology called Word2Vec to express the text, and after that, they use an LSTM neural network to analyze the sentiment of the film review text. The results of the test demonstrate the viability and effectiveness of the LSTM model for the analysis of the emotions contained within film reviews.

In light of the information presented above, the authors of this study devise a model for sentiment analysis using an OCC model and a Bi-LSTM neural network, approaching the problem from the perspective of affective cognition theory and deep learning technology. They then use this model to investigate the opinions expressed by college students participating in university online public opinion. The following are some of the novelties that emerged from our research: (1) at the current stage of research on the analysis of the sentiments of online public opinion, the sentiment categories of data sets frequently rely on manual prior labeling, and there is a lack of clear rules or systems. This study contributes to the research by incorporating the traditional OCC affective cognition evaluation model. As a result, the research on the recognition of sentiments in online public opinion has become more standardized. A standardized sentiment rule system is constructed based on the OCC affective cognition elicitation mechanism. In order to construct a dataset, three different types of positive, negative, and neutral sentiment annotations are applied to the crawled microblog texts in university online public opinion. (2) In this study, the characteristics of short texts of university online opinion are combined, and the Bi-LSTM sentiment analysis algorithm is used to train and obtain a sentiment recognition model for college students. The study was conducted by combining the characteristics of university online opinion short texts. The new model is not only capable of resolving the data sparsity problem that plagues traditional machine learning, but it can also improve recognition accuracy by making effective use of feature sequences of textual contextual information. (3) In the final part of this research project, comparison experiments are carried out in order to validate the correctness of the OCC sentiment rule labeling system as well as the scientific merit

of the Bi-LSTM neural network sentiment classification algorithm. The experimental results prove that, compared with the natural annotation method, the method based on the OCC sentiment rule system greatly improves the effect of the model on college students' sentiment recognition. What's more, the classification model on the basis of Bi-LSTM neural network achieves more desirable classification results in university opinion sentiment recognition compared with the classification models such as SVM and other deep learning networks like CNN and LSTM. As a result, the utilization of deep learning technology to investigate the feelings held by college students in the context of university-based online public opinion is of utmost significance. It is able to identify the emotional state of college students participating in online public opinion in an effective and automated manner, which makes the monitoring and management of unexpected public opinion much simpler. Concurrently, school teachers are in a position to provide timely attention and regulation in response to the pessimistic sentiment of college students.

2. Related Basic Theory

2.1. Sentiment Classification. The research on sentiment classification based on online public opinion has become a hot topic as a direct result of the meteoric rise in the amount of data that is stored online as well as the rising demand for the monitoring and management of online public opinion. As a distinct task within the realm of text classification, sentiment analysis shares some parallels with text classification but also highlights some key distinctions. The former focuses on subjective factors as the research object, while the latter centers on objective content as its primary concern. The fundamental task of sentiment classification is to classify texts into two or more types, positive or negative, based on the emotional information that they contain. This task is accomplished by analyzing the texts' content and looking for patterns of positive or negative sentiment. In its most basic form, it can be understood as a categorization of the attitudes and opinions expressed by the publisher [10], a process that is also known as opinion mining in some circles. Researchers have done a significant amount of work in the field of sentiment classification research and have proposed a few research methods that are representative of the field. From these methods, one can derive three more general categories: sentiment dictionary methods [11], machine learning methods [12], and deep learning methods [13]. Figure 1 depicts the processes that each of the three approaches uses to classify the respondents' feelings.

In light of what has been seen in Figure 1, one can reach the following conclusions: the same is that the first step of these three methods is to preprocess the text. That is, data cleaning, word segmentation, and deactivation of Chinese text. The difference lies in the fact that the methods that are based on a sentiment dictionary make use of a previously constructed sentiment dictionary in order to annotate each sentiment symbol in the text, and finally, the formula is designed to calculate the sentiment tendency value of the full text. This is where the

distinction lies. Selecting and extracting feature items from the text is the central step of the machine learning methods. These feature items are then used to train a text sentiment classification model. In the final step, the trained model is evaluated for its accuracy in classification using a test set through which it is first put through its training. In contrast, the deep learning methods encode the word vector after preprocessing the text, then use deep neural networks for feature extraction, and finally calculate the probability of each sentiment class by using the softmax function to output the sentiment class. These methods are referred to collectively as "deep learning." Because the methods are based on a variety of different principles, the practical applications of each of them have their own unique set of benefits and drawbacks. The methods for classifying emotions that are based on sentiment dictionaries are straightforward and easy to put into practice. The construction of a dictionary, on the other hand, necessitates specialized domain knowledge, which results in the methods having poor universality and a limited capacity for generalization. The intelligence of sentiment classification is improved by machine learning methods, but because these methods rely on the corpus domain, contextual knowledge of the context cannot be effectively used, and the classification performance is also impacted to some degree. On the other hand, the methods of sentiment classification that are based on deep learning have a high rate of accuracy and can be applied in a wide variety of contexts. When compared to more conventional machine learning models, it is capable of automatically extracting features on its own. As a consequence of this, the method of sentiment classification based on deep learning is utilized throughout the study.

2.2. Text Representation. Text is an unstructured or semi-structured form of data organization, which means that computers are unable to directly process it [14]. Text consists of a collection of characters that have been arranged in a certain order. Therefore, if we want to automate the processing of natural language with the assistance of computers, we will first need to convert text into structured data, also known as text representation. This step is necessary because we cannot automate the processing of natural language without it. The representation of text can, in general, be broken down into two categories: discrete representation and distributed representation.

The method of discrete representation known as one-hot coding is common [15]. It is a statistical-based processing method. It regards a word as a symbol, and the dimension of the word vector is the size of the whole dictionary. For each word in the vocabulary, set its corresponding position to 1 and the rest to 0. This method, despite being straightforward and user-friendly, suffers from two major drawbacks, which are as follows: first, because each word in the dictionary is represented by a high-dimensional vector, it is simple to create a dimensional catastrophe when the dictionary is large. This is because each word has its own high-dimensional vector. Second, due to the fact that every pair of

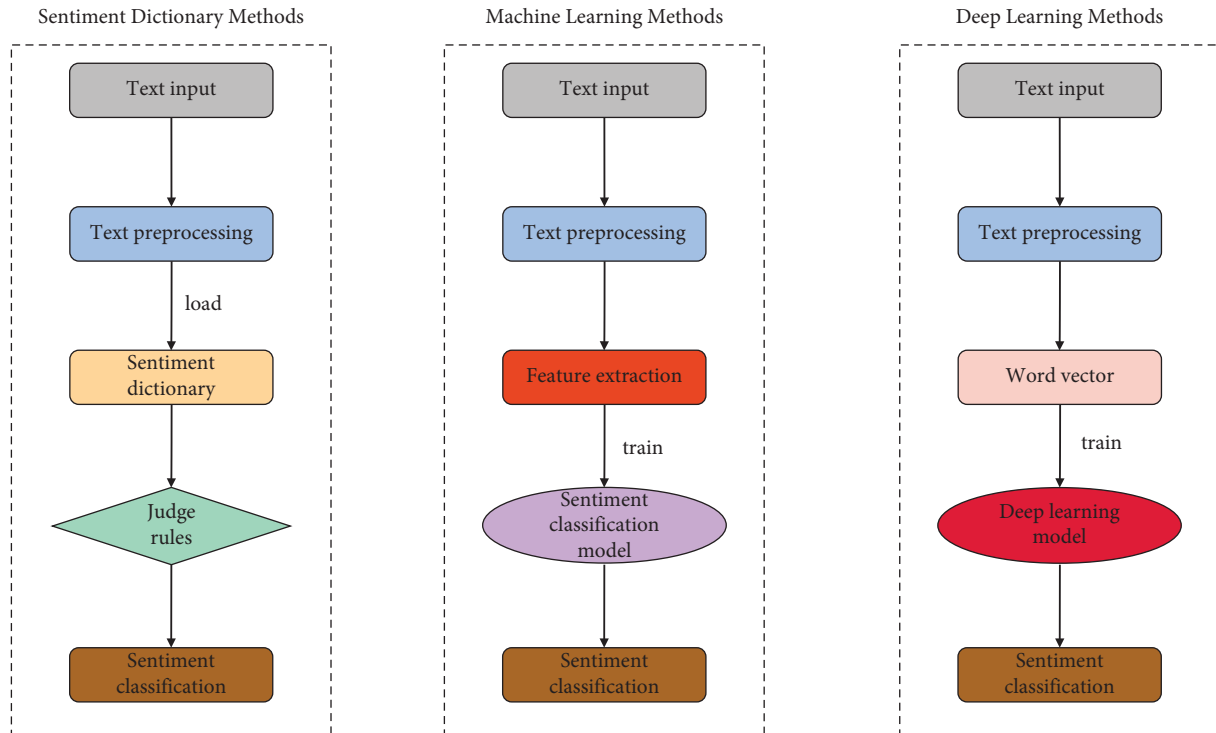


FIGURE 1: Flow of sentiment classification of three representative methods.

words is independent of the other, this technique is unable to capture semantic information, which can easily result in the problem known as the “semantic gap.”

Word2Vec is a method of representative distributed representation [16] that Google developed based on the conventional one-hot model and improved upon it. It is capable of mapping features from a high-dimensional space to a low-dimensional space, thereby resolving the issue of dimensional disaster caused by sparse data in the one-hot model and mining the semantic relationships between words. Word2Vec is essentially a straightforward model of a neural network. Following the completion of the training phase, the text content will be converted into a vector with K dimensions. In addition, the similarity of vectors can be utilized in order to convey the similarity of text semantics. Word2Vec performs well in both the classification of text and the classification of sentiment. The one-hot vector serves as the input for the Word2Vec model. After that, the vector is passed on to the neural network model for training. The parameters of the neural network model are continuously adjusted, and the weight matrix is modified as it is being trained in order to obtain distributed vectors. Word2Vec primarily incorporates two distinct variants of the CBOW and Skip-gram models. Figure 2 presents an illustration of the structure diagrams for both of these models.

When the structure diagrams of the two models shown in Figure 2 are compared and analyzed, we find that CBOW is able to predict the target words given knowledge of the contexts in which the words are found. In order to obtain the word vector associated with the context, it connects the mapping layer directly to the softmax node of the output layer. Additionally, all inputs are projected to the same

mapping layer. It does not take into account the word order of the context in which the target word is being used. In contrast to the CBOW theory, the Skip-gram model makes use of the context’s target words to make predictions about it. Because of this, the word vector for each word in the context contains the position information that corresponds to that word [17]. However, it leads to a longer training time for the Skip-gram model than the CBOW model. The latter is suitable for training larger datasets, and the former is suitable for a smaller amount of data, and it will make the word vectors more accurate. For the purpose of text vector training in this paper, we make use of the Skip-gram algorithm.

3. Evaluation Model Based on Bi-LSTM Neural Network

From the perspectives of affective cognition and deep learning, this research creates a model for recognizing the emotions of college students by employing the OCC model and the Bi LSTM neural network. These two models are used in conjunction with one another. The real feelings that are implied behind students’ text modality data can be mined with the assistance of this model, and the students’ psychological shifts over the course of a certain amount of time can be taken into comprehensive consideration in order to achieve a rapid and accurate identification of students’ sentiment tendencies. The fundamental structure of the model is shown in Figure 3.

The primary components of the sentiment recognition model developed as a result of this research are shown graphically in Figure 3. (1) The gathering of data: the primary

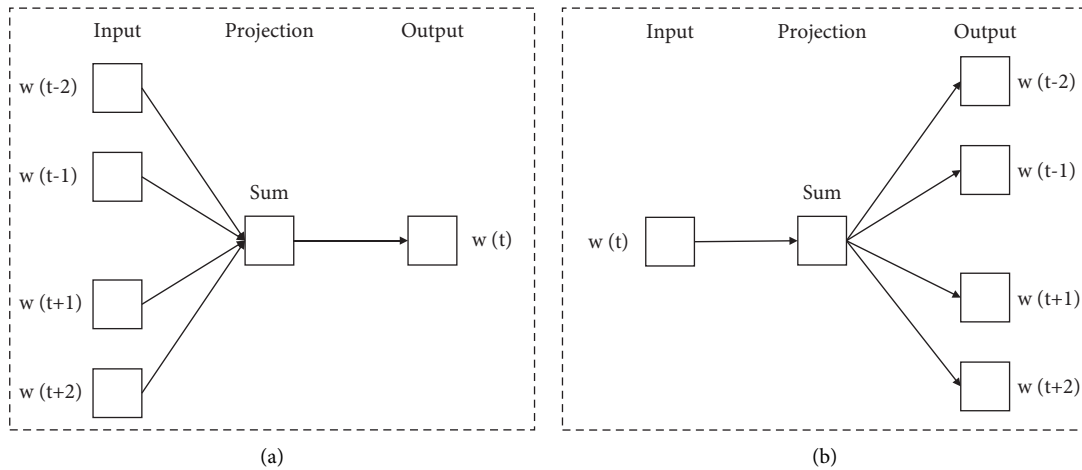


FIGURE 2: Structure of CBOW and Skip-gram model. (a) CBOW model structure. (b) Skip-gram model structure.

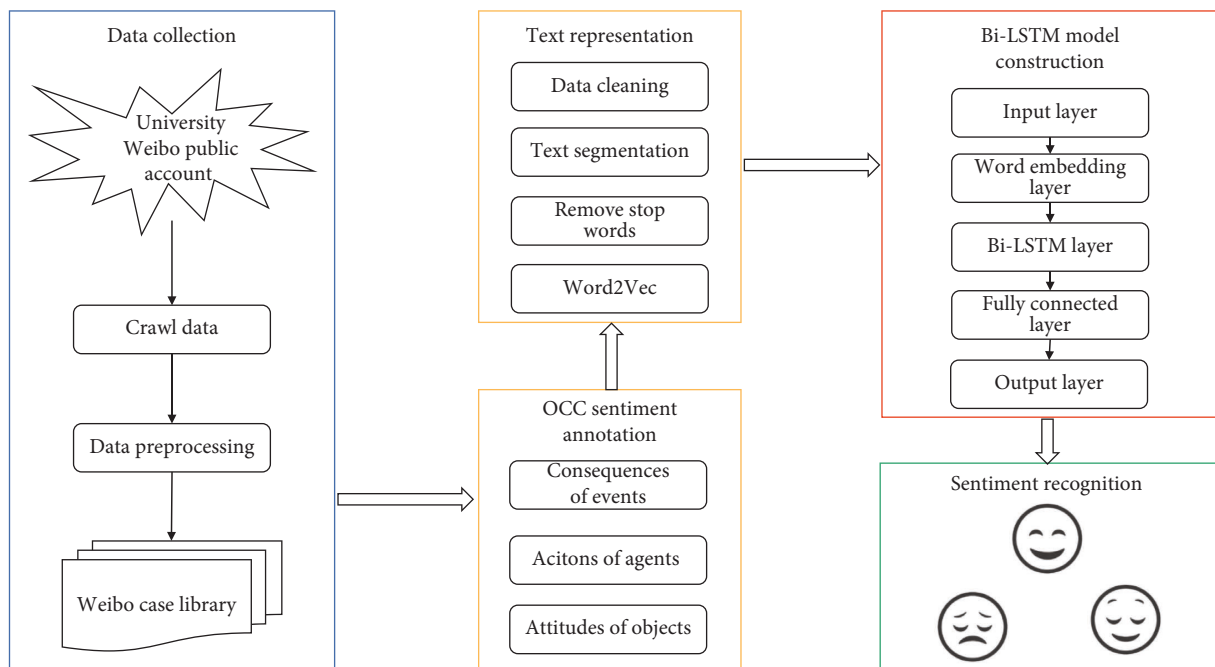


FIGURE 3: Sentiment analysis model of college students based on OCC model and Bi-LSTM neural network.

purpose of this section is to implement the sorting, filtering, and cleaning of the data that was gathered by crawlers in order to get the data ready for the subsequent construction of the model. (2) A reasonable sentiment rule system has been established on the basis of the OCC affective cognition model in order to label the text data of online public opinion with three types of positive, negative, and neutral emotions. These emotions are as follows: (3) following the pre-processing step, the text data are still unstructured or semistructured data, which means that computers are unable to recognize them directly. Because of this, we have decided to incorporate the Skip-gram model into the Word2Vec method for the purpose of text vector training. (4) At the end of the process, the Bi-LSTM model is used as a classifier. The final sentiment recognition model is obtained after adjusting

the hyperparameters through training, and the positive and negative feelings of college students are evaluated.

3.1. Sentiment Annotation Based on OCC Model. When it comes to expressing one’s views in the context of human interaction, the use of emotion is an essential component. It is necessary for people to investigate and express their emotions using the sentiment model. The OCC model is the most classical affective cognitive model in cognitive psychology. It is also a psychological sentiment model that is widely used in today’s society [18]. It provides a classification scheme for sentiment and provides a scientifically referable basis for standardizing the sentiment labeling system of online public opinion. When it comes to modeling

sentiment, the OCC model is the most prominent example. It provides a classification scheme for sentiment as well as a scientifically referable basis for standardizing the sentiment labeling system of online public opinion. As a result of this, the focus of our research is on implementing the OCC model for sentiment annotation of online opinion texts. This allows us to take into account the closed-loop principle and finer levels of sentiment granularity. Its high computability lays the groundwork for investigating the sentiment of university-based online opinion information texts, and its mechanism of affective cognitive elicitation is also an important support for investigating the factors that contribute to the formation of sentiment.

The basic sentiment types of college students' online public opinion are determined to be, after screening and combining with the actual situation, the following eight basic emotions: happy, pity, admiration, reproach, gratitude, anger, love, and hate. These are the emotions that have been

adopted as the basic sentiment types. It is not necessary to place an excessive amount of emphasis on particular negative sentiment, however, because the research is more inclined to explore the sentiment tendencies in the online public opinion of college students. As a result, it divides the eight OCC basic sentiment types of online public opinion into three distinct categories: positive, negative, and neutral. That is to say, four different types of positive sentiment, such as happiness, admiration, gratitude, and love, are mapped onto the positive sentiment category, whereas four different types of negative sentiment, such as pity, reproach, anger, and hate, are mapped onto the negative sentiment category. As can be seen in Figure 4, the comment text that does not fit into any of these eight categories of sentiment is considered to have a neutral level of sentiment.

A 9-dimensional sentiment space is constructed in this paper, as shown in Figure 4, and a sentiment variable is assigned to each web text, as shown in formula (1).

$$S = \{s_{\text{happy}}, s_{\text{pity}}, s_{\text{admiration}}, s_{\text{reproach}}, s_{\text{gratitude}}, s_{\text{anger}}, s_{\text{love}}, s_{\text{hate}}, s_{\text{other}}\}. \quad (1)$$

In formula (1), $s[0, 1]$ denotes the value of each dimension of sentiment.

Then, according to the OCC model mapping of the web text, three different formulas for expressing sentiment, as in formulas (2)–(4).

$$S(\text{Positive}) = S(\text{happy}) \cup S(\text{admiration}) \cup S(\text{gratitude}) \cup S(\text{love}), \quad (2)$$

$$S(\text{Negative}) = S(\text{pity}) \cup S(\text{reproach}) \cup S(\text{anger}) \cup S(\text{hate}), \quad (3)$$

$$S(\text{Neutral}) = 1 - S(\text{positive}) - S(\text{negative}). \quad (4)$$

3.2. Sentiment Recognition Based on Bi-LSTM Model.

Textual information is necessary for humans to communicate their feelings and thoughts to one another, and it also plays a significant role in the external representation of a person's mental conditions. It is essential, according to the consensus of opinion expressed online, to accurately determine the state of a college student's mental health by analyzing the mental condition of the student and the sentiment expressed in any content that they post [19]. Even though text is frequently organized as sequential data, the tendency of text can be mined for sentiment effectively based on semantic comprehension if we are able to capture information about the contextual setting of sentences.

This research develops an online opinion sentiment rule that, as was mentioned in Section 3.1, is able to accurately identify the sentiment categories that are present in online opinion texts. This rule is based on the affective cognitive model that was developed by OCC. On the other hand, in today's world, the number of online opinion texts that we need to perform sentiment recognition is in the tens of thousands, hundreds of thousands, or even millions or hundreds of millions. This is because the Internet is filled

with people sharing their thoughts. The effectiveness of sentiment classification will be poor if we only manually annotate each web opinion microblog document using the OCC model. Researchers now have a viable option for dealing with large amounts of data, thanks to the rapid development of deep learning. Because of this, the research uses a Bi-LSTM neural network to develop a sentiment recognition model for online public opinion. This is done so that the researchers can quickly and effectively identify the sentiment tendency of college students.

The long short-term memory (LSTM) network is well-known for its distinct gating structure and memory units, which can help avoid problems associated with gradient disappearance and long-term reliance [20]. However, when it learns the features of the text sequence, the information can only propagate in one direction and cannot make good use of the text context information. When applied to the processing of text data, the Bi-LSTM method has the ability to obtain feature sequences that contain text context information. In addition to the information that is unique to it, it includes the information that is associated with the entirety of the text data. Therefore, it contributes to the text's

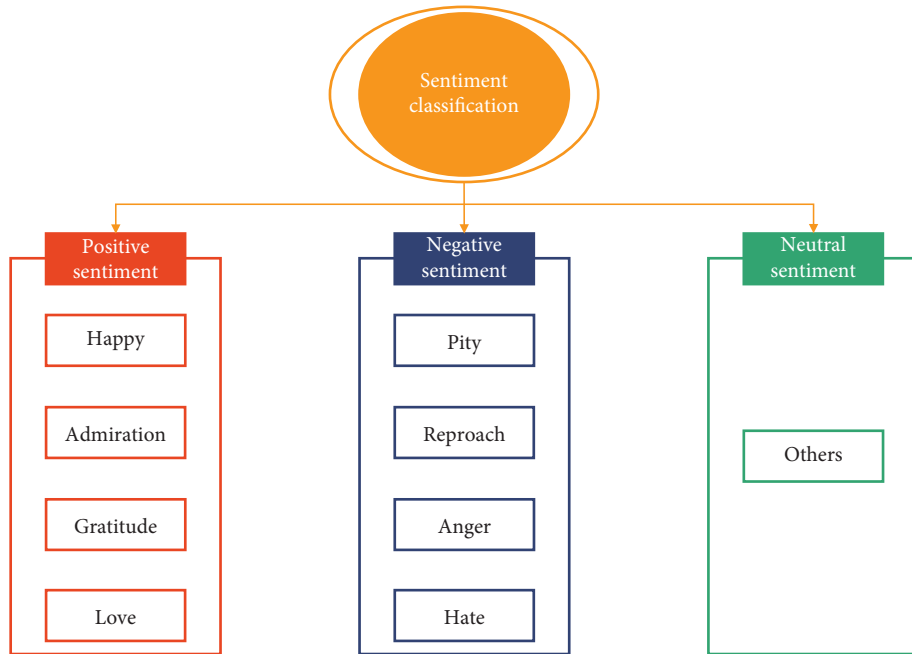


FIGURE 4: Sentiment types mapping based on OCC model.

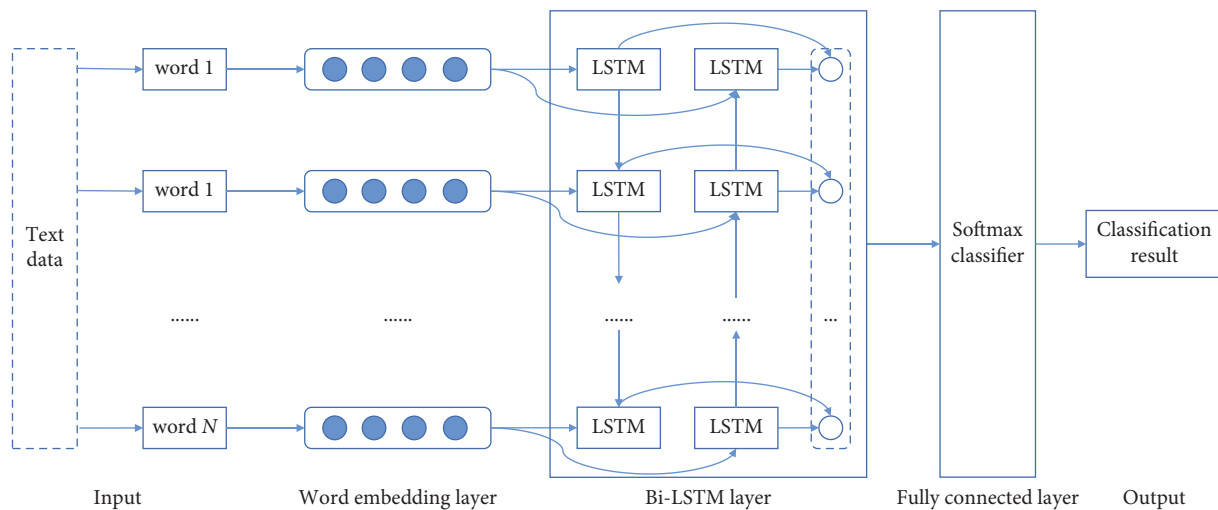


FIGURE 5: Sentiment analysis model based on Bi-LSTM neural network.

increased capacity for differentiation [21]. Because of this, our research proposes a three-layer neural network architecture consisting of a word embedding layer, a Bi-LSTM layer, and a full connectivity layer that spans the contextual interval to learn the sentiment information implied by the sentences. This will allow for a comprehensive analysis of the semantics of sayings as well as the acquisition of useful representations of sentiment feature representations. Figure 5 presents a visual representation of the model’s flow for recognizing sentiment.

This study employs the Skip-gram model to train the word vector of the text in the word embedding layer because the microblog text exhibits the characteristics of

colloquialism and short and succinct expression. As a result, the word vector of the text in this layer is able to more accurately and appropriately represent the text vector.

Following that, two LSTM networks with different timing are connected to the same output in the Bi-LSTM layer [22]. In this situation, the forward LSTM is utilized to extract the text’s top-level information, and the backward LSTM is used to get the text’s bottom-level information. Assume that r_t, f_t, b_t stand for the text feature representation at time t , the forward hidden state at time t , and the backward hidden state at time t , respectively. r_t needs to be computed by integrating f_t and b_t in a spliced manner. And f_t is calculated from the input s_t at time t and the hidden state

information at time $t - 1$. b_t is calculated from the input s_t at time t and the hidden state information at time $t + 1$. The details are shown in formulas (5)–(7).

$$r_t = f_t \oplus b_t, \quad (5)$$

$$f_t = \delta(U \times s_t + V \times f_{t-1} + \beta), \quad (6)$$

$$b_t = f(U' \times s_t + V' \times f_{t+1} + \beta'), \quad (7)$$

where \oplus denotes the integration of both in a spliced manner. δ denotes the LSTM nonlinear function, U , V , U' , and V' indicate the weights of the function, respectively. β and β' represent the bias of the function.

Last but not least, the fully connected layer takes as its input the text feature representation that was obtained from the layer before it, classifies it through the use of a softmax function, and then outputs the computed sentiment tendency values.

4. Experimental Testing and Analysis

In order to test the efficacy and reliability of the model designed in this paper in identifying the sentiment of college students' online public opinion, the sentiment analysis model based on OCC and Bi-LSTM neural network is compared with other representative sentiment analysis models, such as SVM, CNN, and LSTM, respectively. The purpose of this comparison is to test the effectiveness and reliability of the model designed in this paper.

4.1. Experimental Data. The students at a university located in Jiangsu Province are going to serve as the subjects of this study. In order to carry out the experiment, the comments made by university students on topics that were deemed to be pertinent were crawled from the official microblog account that the university maintains. Both manual annotation and the OCC sentiment rule theory, which was presented in Section 3.1, are utilized as respective foundations for the application of sentiment annotation to the dataset. In this paper, the data that were collected by the crawler are filtered and cleaned up to leave 9600 pieces of data. The members of the project team who understood the OCC sentiment rule system and those who did not understand the OCC model are asked to label the microblog information texts with sentiment classification, respectively, in order to form the online opinion dataset for the university.

4.2. Data Processing. In many cases, the quality of the data has an effect on the experimental findings of the sentiment analysis performed on college students. In the meantime, the raw web text data that are crawled from the public microblogs of universities are not suitable for direct utilization in the context of sentiment analysis. Because of this, we need to run these raw data through a series of preprocessing operations such as data cleaning, word separation, and deactivation before we can use them for sentiment

TABLE 1: Parameter settings in the Skip-gram model.

Parameters	Values
Window size	5
Minimum word frequency	5
The dimension of output word vector	300

analysis of text. These operations include cleaning the data, separating the words, and deactivating the data.

In addition to the preprocessing operation performed on the microblog text dataset, the Word2Vec tool must be used for the word vector training in this particular paper. Because the Skip-gram model is constantly adjusted to the target words based on the prediction of the context, we use it in our research to train and obtain the word vector files. Although the number of predictions is higher, the learning effect is also significantly more beneficial. Table 1 presents the training process's various parameters in an illustrated format.

4.3. Evaluating Indicators. Common evaluation indicators for sentiment classification include accuracy, precision, recall, and $F1$ -measure. Sentiment classification is a form of text classification. Because of this, the purpose of this paper is to evaluate the effect of the experimental model for sentiment classification using the aforementioned four common evaluation criteria, and Figure 6 provides an illustration of the detailed evaluation index system.

4.4. Results and Analysis

4.4.1. Validation of OCC Sentiment Annotation. The natural labeled dataset was used as the comparison dataset and as input to the Bi-LSTM sentiment recognition model for performance testing respectively. The experimental results of the comparison are illustrated in Table 2, and the purpose of this was to verify the effect that the OCC sentiment rule labeled dataset had on the effect of the sentiment recognition model.

The experimental comparison results are presented in Table 2, and they demonstrate that when compared with the manual natural annotation method, the implementation of the OCC sentiment rule system for the sentiment annotation of college online opinion datasets results in a significant increase in the effectiveness of the model. It is possible that this is because many college and university students do not use emotion words when expressing their feelings on social network platforms like microblogs. This is something that is common among students at those institutions. For this situation, natural annotation methods often cannot determine the sentiment attributes of these texts, so the annotators may give annotation results based on the prevailing environment or mood. However, this results in less standardized sentiment annotation and eventually affects the recognition performance of the model. In order to solve issues of this nature, the OCC has developed sentiment rules that provide a rational and standardized annotation system. Because of this, the recognition performance of the model is improved, and it also explains the reasons for affective

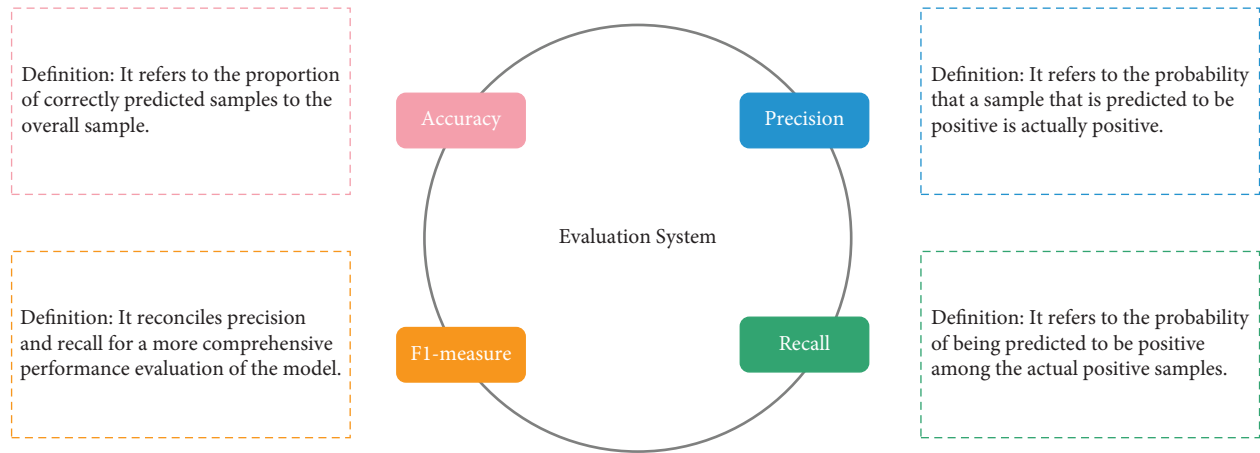


FIGURE 6: Evaluation index system.

TABLE 2: Effect of different sentiment labeling methods on sentiment recognition results.

Sentiment labeling methods	Accuracy	Precision	Recall	F1-measure
OCC sentiment annotation	86.27	84.02	85.67	84.84
Artificial sentiment annotation	78.16	76.24	76.51	76.38

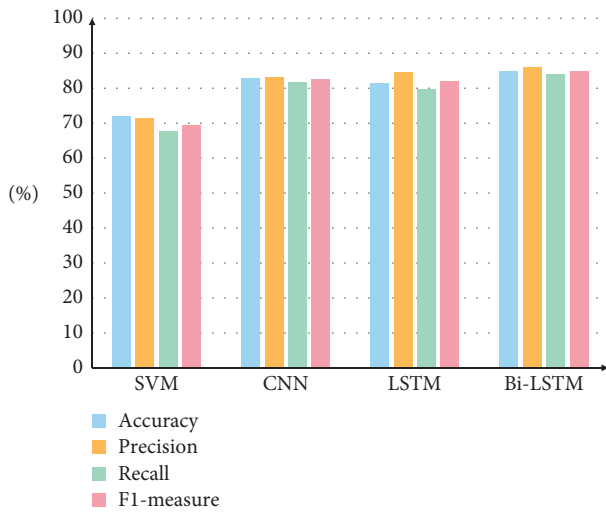


FIGURE 7: Emotion recognition results of different methods.

cognition from the perspective of cognitive psychology. Additionally, it makes the sentiment labeling of university online opinion datasets more accurate and reasonable.

4.4.2. Comparison Experiments with SVM, CNN, and LSTM Evaluation Models. Other comparative experiments are designed in this paper with the intention of completely confirming the model’s efficacy in the sentiment recognition of university online public opinion. On the university public opinion dataset that was constructed for the study, comparison experiments are carried out utilizing a Bi-LSTM model in conjunction with SVM, CNN, and LSTM models. Figure 7 provides a visual representation of the findings from the experiments.

The overall classification performance of the neural network methods is significantly better than that of the SVM method when using the same sentiment labeled dataset and the same word vector training, as shown in Figure 7. This is in comparison to the traditional machine learning method, which is illustrated in the figure. Among these models, the Bi-LSTM-based sentiment classification model achieves more satisfactory classification results in all four evaluation indicators, and its overall classification performance is significantly better than that of the SVM, CNN, and LSTM models. Although the LSTM model has long sequence processing capability, the experimental results show that the accuracy of the CNN model is slightly higher than that of the LSTM model. In addition, in comparison to the LSTM model, the Bi-LSTM algorithm is able to obtain feature sequences containing textual contextual information when processing text data. These feature sequences contain both their own unique information as well as the correlation information of the whole text data, which enhances the discriminability, resulting in more satisfactory classification results.

5. Conclusion

As a result of the Internet’s ever-increasing popularity in today’s society, a substantially greater number of individuals take part in online conversations. College students are willing to express their feelings, elaborate on their positions, and reflect their demands through the use of social platforms because they are active participants in the online platform. Because of this, relying on the textual data in university online public opinion and utilizing appropriate evaluation models to analyze the sentiment tendencies of college students hidden in these data can help universities grasp the ideological dynamics and mental health status of college

students in a timely manner, discover some existing problems and prevent them before they happen, and provide better assistance to schools and student management as much as possible. In this paper, with the assistance of relevant data that was crawled from the public platform of college microblogs, we develop a sentiment analysis model for college students on the basis of the OCC model and the Bi-LSTM neural network from the perspective of affective cognition and deep learning. The goal of this model is to identify college students' sentiment tendency in university online public opinion. The comparative results show that the model for sentiment recognition that was constructed during the research has achieved ideal results in sentiment recognition of college students, and it has certain feasibility. This is demonstrated by the fact that the model was able to achieve these ideal results.

According to the findings presented above, the research on the sentiment recognition of college students conducted by this study has yielded satisfactory results, and the performance of the sentiment recognition model has been significantly improved. Having said that, there are a few areas that could use some additional research and development: (1) in terms of data collection, this study only collected data from the public microblogging platforms of colleges and universities. And the data sources after preprocessing are not rich enough in terms of data volume and data dimension although they have higher credibility. In the future, it is necessary to further expand the data sources and improve the comprehensiveness of the data. (2) Meanwhile, this study only conducts sentiment analysis research on text data in social platforms. In the follow-up, we can explore the sentiment recognition method of multi-source data that combines image and text data, and find a model with stronger generalization ability to better realize the sentiment classification of college students. (3) At last, based on the sentiment classification model of college students, this paper only studies the problem of positive, negative and neutral three classifications. And people's emotions are rich and diverse, so it is hoped that a more fine-grained multi-sentiment recognition model can be studied in the future.

Data Availability

The labeled dataset used to support the findings of this study can be obtained 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 Higher Vocational Education High-level Professional Group Construction Project, under Grant Su Jiaozhi Letter [2021] no. 1, Jiangsu Province Higher Vocational Education Industry-Education Integration Platform Construction Project, under Grant Su Jiaozhi Letter [2019] no. 26, Natural Science

Research Project of Jiangsu Province Colleges and Universities, under Grant 18KJD510011, Jiangsu Province High-Level Key Professional Construction Project Funding, under Grant Su Jiaogao [2017] no. 17, Innovation Research on Construction of Ideological and Political Education Work System in Colleges and Universities in the Era of Big Data, under Grant 2022SJQTX0033, Research on the Student-Centered "Introduction to New Generation Information Technology" Public Course Flipped Hybrid Teaching Model, under Grant 2020-AFCEC-055, and Project of the Jiangsu Education System Party Construction Research Association, under Grant 2021JSYDJ01017.

References

- [1] F. T. Maalouf, B. Mdawar, L. I. Meho, and E. A. Akl, "Mental health research in response to the COVID-19, Ebola, and H1N1 outbreaks: a comparative bibliometric analysis," *Journal of Psychiatric Research*, vol. 132, pp. 198–206, 2021.
- [2] J. M. van der Wal, C. D. van Borkulo, M. K. Deserno et al., "Advancing urban mental health research: from complexity science to actionable targets for intervention," *The Lancet Psychiatry*, vol. 8, no. 11, pp. 991–1000, 2021.
- [3] F. M. Plaza-Del-Arco, M. D. Molina-Gonzalez, L. A. Urena-Lopez, and M. T. Martin-Valdivia, "A multi-task learning approach to hate speech detection leveraging sentiment analysis," *IEEE Access*, no. 9, pp. 112478–112489, 2021.
- [4] H. A. Bouarara, "Sentiment analysis using machine learning algorithms and text mining to detect symptoms of mental difficulties over social media," *International Journal of Information Systems and Social Change*, vol. 12, no. 2, pp. 1–15, 2021.
- [5] M. S. Kalaivani and S. Jayalakshmi, "Sentiment analysis on micro-blog data using machine learning techniques-A Review," *IOP Conference Series: Materials Science and Engineering*, vol. 1049, no. 1, Article ID 012012, 2021.
- [6] H. Jiang, "Innovation of ideological and political education in colleges and universities from the perspective of network public opinion," *Journal of Contemporary Educational Research*, vol. 5, no. 7, pp. 26–30, 2021.
- [7] B. Ozyurt and M. A. Akcayol, "A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA," *Expert Systems with Applications*, vol. 168, Article ID 114231, 2021.
- [8] R. A. Ariyanto and N. Chamidah, "Sentiment analysis for zoning system Admission policy using support vector machine and naive bayes methods," *Journal of Physics: Conference Series*, vol. 1776, no. 1, Article ID 012058, 2021.
- [9] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, "Co-LSTM: convolutional LSTM model for sentiment analysis in social big data," *Information Processing & Management*, vol. 58, no. 1, Article ID 102435, 2021.
- [10] F. M. Al-Kharboush and M. A. Al-Hagery, "Features extraction effect on the accuracy of sentiment classification using ensemble models," *International Journal of Science and Research*, vol. 10, no. 3, pp. 228–231, 2021.
- [11] R. Duan, Z. Huang, Y. Zhang, X. Liu, and Y. Dang, "Sentiment classification algorithm based on the cascade of BERT model and adaptive sentiment dictionary," *Wireless Communications and Mobile Computing*, vol. 2021, no. 5, Article ID 8785413, 8 pages, 2021.
- [12] K. Gulati, S. Saravana Kumar, R. Sarath Kumar Boddu, K. Sarvakar, D. Kumar Sharma, and M. Nomani,

- “Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to COVID-19 pandemic,” *Materials Today Proceedings*, vol. 51, no. 1, pp. 38–41, 2022.
- [13] A. Mabrouk, R. P. D. Redondo, and M. Kayed, “Deep learning-based sentiment classification: a comparative survey,” *IEEE Access*, no. 8, pp. 85616–85638, 2020.
- [14] M. García, S. Maldonado, and C. Vairetti, “Efficient n-gram construction for text categorization using feature selection techniques,” *Intelligent Data Analysis*, vol. 25, no. 3, pp. 509–525, 2021.
- [15] A. H. Osman and O. M. Barukub, “Graph-based text representation and matching: a review of the state of the art and future challenges,” *IEEE Access*, no. 8, pp. 87562–87583, 2020.
- [16] A. K. Sharma, S. Chaurasia, and D. K. Srivastava, “Sentimental short sentences classification by using CNN deep learning model with fine tuned Word2Vec,” *Procedia Computer Science*, vol. 167, pp. 1139–1147, 2020.
- [17] H. A. Al-Muzaini and A. M. Azmi, “Impact of stemming and word embedding on deep learning-based Arabic text categorization,” *IEEE Access*, no. 8, pp. 127913–127928, 2020.
- [18] F. Riaz, S. Jabbar, M. Sajid, M. Ahmad, K. Naseer, and N. Ali, “A collision avoidance scheme for autonomous vehicles inspired by human social norms,” *Computers & Electrical Engineering*, vol. 69, pp. 690–704, 2018.
- [19] D. Lili, S. Lei, and X. Gang, “Public opinion analysis of complex network information of local similarity clustering based on intelligent fuzzy system,” *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 2, pp. 1693–1700, 2020.
- [20] R. K. Behera, M. Jena, S. K. Rath, and S. Misra, “Co-LSTM: convolutional LSTM model for sentiment analysis in social big data,” *Information Processing & Management*, vol. 58, no. 1, Article ID 102435, 2021.
- [21] P. Ts and P. Shrinivasacharya, “Evaluating neural networks using Bi-Directional LSTM for network IDS (intrusion detection systems) in cyber security,” *Global Transitions Proceedings*, vol. 2, no. 2, pp. 448–454, 2021.
- [22] P. Bahad, P. Saxena, and R. Kamal, “Fake news detection using Bi-directional LSTM-recurrent neural network,” *Procedia Computer Science*, vol. 165, pp. 74–82, 2019.

Research Article

Comparative Analysis of the Performance of Complex Texture Clustering Driven by Computational Intelligence Methods Using Multiple Clustering Models

Jincheng Zhou,^{1,2,3} Dan Wang ,^{2,3,4} Lei Ling,^{2,3} Mingjiang Li,^{1,2} and Khin-Wee Lai ⁵

¹School of Computer and Information, Qiannan Normal University for Nationalities, Duyun 558000, China

²Key Laboratory of Complex Systems and Intelligent Optimization of Guizhou, Duyun 558000, China

³Key Laboratory of Complex Systems and Intelligent Optimization of Qiannan, Duyun 558000, China

⁴School of Mathematics and Statistics, Qiannan Normal University for Nationalities, Duyun 558000, China

⁵Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

Correspondence should be addressed to Dan Wang; wdan@sgmtu.edu.cn

Received 6 August 2022; Revised 9 September 2022; Accepted 12 September 2022; Published 29 September 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Jincheng 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.

Traditional texture cluster algorithms are frequently used in engineering; however, despite their widespread application, these algorithms continue to suffer from drawbacks including excessive complexity and limited universality. This study will focus primarily on the analysis of the performance of a number of different texture clustering algorithms. In addition, the performance of traditional texture classification algorithms will be compared in terms of image size, clustering number, running time, and accuracy. Finally, the performance boundaries of various algorithms will be determined in order to determine where future improvements to these algorithms should be concentrated. In the experiment, some traditional clustering algorithms are used as comparative tools for performance analysis. The qualitative and quantitative data both show that there is a significant difference in performance between the different algorithms. It is only possible to achieve better performance by selecting the appropriate algorithm based on the characteristics of the texture image.

1. Introduction

Pattern recognition research places a significant emphasis on the study of image classification. The labeling of the image samples with the appropriate categories after they have been categorized is its responsibility. The characteristics of picture samples serve as the foundation for classification [1], with texture being one of the most important features to use when attempting to characterize image detail information [2]. Shape, color, texture, and other low-level features as well as more complicated high-level feature information are some examples of the characteristics of picture samples. As a consequence of this, the investigation of texture image feature extraction and classification method is of significant significance, both in terms of theory and practice.

Images with a texture can accurately reflect the surface characteristics of the objects or scenes they depict, and they are a visual element that is both common and important. Because of this, the extraction and recognition of texture pattern characteristics that are included in texture images has always been an important study direction [3], particularly in the fields of image understanding, pattern recognition, and computer vision. The primary goal of texture classification is to extract discriminative texture features from texture images. Once these features have been extracted, some type of distance measurement and classifier are applied to determine the category of the texture image. The most important aspect of research that goes into texture image recognition is the process of extracting the features of textured images. Even with powerful classifiers, it can be difficult to obtain decent recognition results if the derived

texture features lack the ability to differentiate between different types of textures. Extensive research has been done on texture images by academics, primarily in the areas of texture feature extraction, texture image segmentation, texture image classification, recognition, and others. The process of identifying textures requires the extraction of important features from the textures [4]. The accuracy of the classification of texture images is significantly influenced by the performance of the texture characteristics. Photos of textures that have desirable textural qualities are simpler to categorize. Bad texture features, on the other hand, make it difficult to classify texture images, which not only take a long time to classify but also have a poor classification effect. This is because bad texture features make it difficult to distinguish between different types of textures. In addition to the properties of the texture, the classification methods that are used have a significant impact on the result of the classification of texture images [5]. Methods that are effective in classification cut down on the amount of time spent on the process while simultaneously raising both classification efficiency and accuracy [6]. Therefore, the extraction of features from texture images and the development of classification schemes for those images are essential components of texture image analysis.

Over the course of the past half century, academics from both the United States and other countries have engaged in a substantial amount of research on various textural aspects. As our knowledge of texture images expands, a number of different algorithms for the extraction of texture picture features are presented and have gained widespread use. The gray-level co-occurrence matrix, also known as GLOM [6] is the most representative approach to the process of extracting features from texture pictures. The extracted texture feature has a strong effect on classification when applied to photographs with regular textures. There will be some correlation and co-occurrence between the two pixels at some distance because the texture image is generated by the periodicity change of gray distribution in spatial position. This means that there will be some distance between the two pixels. The GLCM algorithm analyzes an image's texture in order to characterize its features by locating gray-level co-occurrences in space.

The Markov Random Field (MRF) [7–9] and the Fractal Dimension [10–12] are two approaches to texture segmentation that have stood the test of time. Because the MRF technique considers the texture picture to be a two-dimensional random process and assumes that each pixel's gray-value is solely related to that of the surrounding pixels, the texture image is modeled as a two-dimensional MRF model [13, 14]. This is because the MRF technique assumes that each pixel's gray value is solely related to that of the surrounding pixels. Because the parameters of the MRF model can explain both the intensity and the direction of the texture image, the features of the texture image correspond to the parameters of the MRF model. The result of this is that the process of extracting features from a texture image is the same as the process of estimating the model parameters in the MRF approach, and the estimated parameters can be used to characterize the features of the texture image. To put

it another way, the MRF technique utilizes the two-dimensional random process that is commonplace in conventional signal analysis in order to describe the characteristics of image textures [15]. The primary objective of the fractal dimension approach, which is classified as a method of structural analysis, is to investigate the structure and morphology of texture textures at various scales in texture photographs. Many times, rather than using an integer, a fraction will be used to express the fractal dimension. Researchers have shown that the roughness of an image correlates with its fractal dimension. When the corresponding image is more jagged, the fractal dimension will be larger; conversely, when the matching image is smoother, the fractal dimension will be smaller. Fractal dimension can be used to represent the roughness of a texture image because texture pictures exhibit self-similarity and varied roughness at different scales. This means that fractal dimension can be used as the characteristic parameter for the classification and segmentation of texture images [16]. Another well-known approach to extracting features from texture images is known as the Gabor filtering method. It is analogous to the wavelet transform method of investigation. Both of them are capable of carrying out analyzes of images at multiple resolutions. As a consequence of this, another name for it is the Gabor wavelet approach. The Gabor filtering method takes into account the textural properties of different scales as narrow-band signals. The method then extracts the features of these narrow-band signals using filter-banks that have varying center frequencies and bandwidths. Over the past few years, the Local Binary Pattern (LBP) and its improved algorithm have become the method of choice for extracting texture picture features the majority of the time. The LBP technique is the method that is utilized the most frequently for the purpose of texture feature extraction because the extracted texture features not only have excellent texture classification accuracy, but also have simple theory, efficient implementation, and invariance to monotonous lighting changes. The fundamental concept behind the LBP algorithm is to encode the local neighborhood features and then calculate the histogram of these encoded values as the texture image's feature description. This process is repeated until all of the local neighborhood features have been encoded.

As the technology for texture image classification improves, more difficult and complex texture image processing tasks are assigned, but the feature extraction methods that have been around for a long time have noticeable bottlenecks. In the field of texture picture feature extraction, the method that is capable of self-learning has started to become more popular. The concept of deep learning refers to a model of neural networks that simulates the layered extraction of human brain functions. It employs a multi-layer network topology and learns from a huge number of training examples, which enables it to develop a very complicated model and complete an extremely challenging task. As a result, it can automatically extract discriminative characteristics relevant for classification from samples. Although deep learning has made significant advancements in texture classification, the traditional method has reached a high level

of maturity after decades of development and is frequently used in engineering. This is in contrast to the situation with deep learning, which has made significant advancements in texture classification.

This study will focus primarily on the analysis of the performance of a number of different texture clustering algorithms. In addition, the performance of traditional texture classification algorithms will be compared in terms of image size, clustering number, running time, and accuracy. Finally, the performance boundaries of various algorithms will be determined in order to determine where future improvements to these algorithms should be concentrated. Before moving on to the actual process of extracting the texture features, the structure arrangement of this paper begins by performing a Gabor filtering pre-processing on the grayscale image. After that, the texture image is clustered and segmented utilizing a number of different clustering algorithms; the NMI and RI indicators are calculated by comparing the obtained label value with the ground-truth label; and finally, the results of the texture segmentation are compared and analyzed qualitatively and quantitatively.

2. Basic Feature Analysis of Texture Image

Texture images can be found anywhere in the environment of day-to-day life, and it is not difficult to obtain or identify them, as demonstrated in Figure 1. These features of the texture pattern can describe the fundamental characteristics of an object's surface or structure, and they are a very important aspect of human visual perception that contributes to cognitive function. Natural texture, artificial texture, and a combination of the two are the most common types of textures found in real-world images of textures. Many images, such as cloth, wood, and forests, among others, have many similar repeated patch-units that are regularly distributed. These kinds of images are frequently referred to as texture images, and the repeated regular distribution of similar units is frequently referred to as the texture's features.

The textural aspects of an image sometimes take the form of patterns such as spots, grids, stripes, rings, and other similar designs. These patterns are used to characterize the spatial distribution and spatial relationship between the gray levels of an image. It is a regular and thorough reflection of a huge number of patch-units that have properties that are either comparable or identical to one another. Generally speaking, the surface characteristics of many items have very distinct differences in texture. For instance, the texture of the forest is more granular than the texture of the farmland, the texture of the farmland is less visible than the texture of the lake, and the rate of color change is more gradual. The resolution of an image has a significant impact on the characteristics of its texture. For instance, high-resolution pictures are able to portray the details of small ground objects quite well, but the features of the textures themselves are not readily apparent. Images of low resolution make the details of textures and patterns much more apparent. Texture characteristics, in comparison to geometric features and

gray features, hold a greater amount of information. It has found widespread application in a variety of domains. As a result, the categorization and recognition of texture images have emerged as significant tools for the human visual system to use in its quest to understand its surroundings. The extraction of features from a texture image is both an essential step and an important part of the process of comprehending and recognizing a texture image.

It is challenging for scholars to provide an exact and consistent description of texture as well as a mathematical model of texture as illustrated in Figure 2. This is because there is a vast diversity and complexity of texture patterns or patch-units contained in texture photographs. As of recent times, there is no definition of texture that is generally recognized by everyone. Despite the absence of a standardized definition of texture and a mathematical model for describing it, there is a general agreement among people on the following qualities of texture images: textons are the fundamental unit of visual perception in texture images, and their appearance is characteristically repetitive. Textons can appear in a texture image in either a regular or random pattern, depending on the type of texture being created. Because of this, the existence and repeatability of texton is considered to be the most important quality of a texture image. When viewed from this angle, a texture image can be understood to represent the culmination of a process in which texton is arranged or distributed in accordance with a set of predetermined guidelines. The detection of texture textons and their repeatability has become the main study content of texture image feature extraction and classification because diverse texture textons and arrangement rules can combine to generate thousands of distinct texture images. In addition, texture textons typically depict regional aspects of the texture images. Therefore, in order to properly grasp texture images, we should not think of texture textons as a point process but rather as regional features that fall within a particular range. This is the approach we should take when analyzing and comprehending texture images.

Because each of the examples that are shown in Figures 1 and 2 only contains a single texture, it is very simple to analyze the properties of each of the textures using these figures. Actual photographs, on the other hand, exhibit a diverse range of surface qualities. Mountains, rivers, farmlands, and lakes are all examples of different types of textured regions that can be found in remote sensing photographs. Nontexture regions are the parts of an image that have only minor or no shifts in grayscale, whereas texture regions are the parts of an image that have significant shifts in grayscale. Texture regions are referred to as the parts of an image that have significant shifts in grayscale. It is difficult to extract different types of texture features using methods such as threshold segmentation and others due to the fact that this is the case. In Figure 3, you can see a picture that contains multiclass textures, as well as the findings of the ground-truth classification that correspond to those results (b). If one wishes to successfully realize the texture description of a texture region, the texture clustering approach is an absolute necessity. A texture feature is a feature that has the ability to successfully differentiate between various different textures.

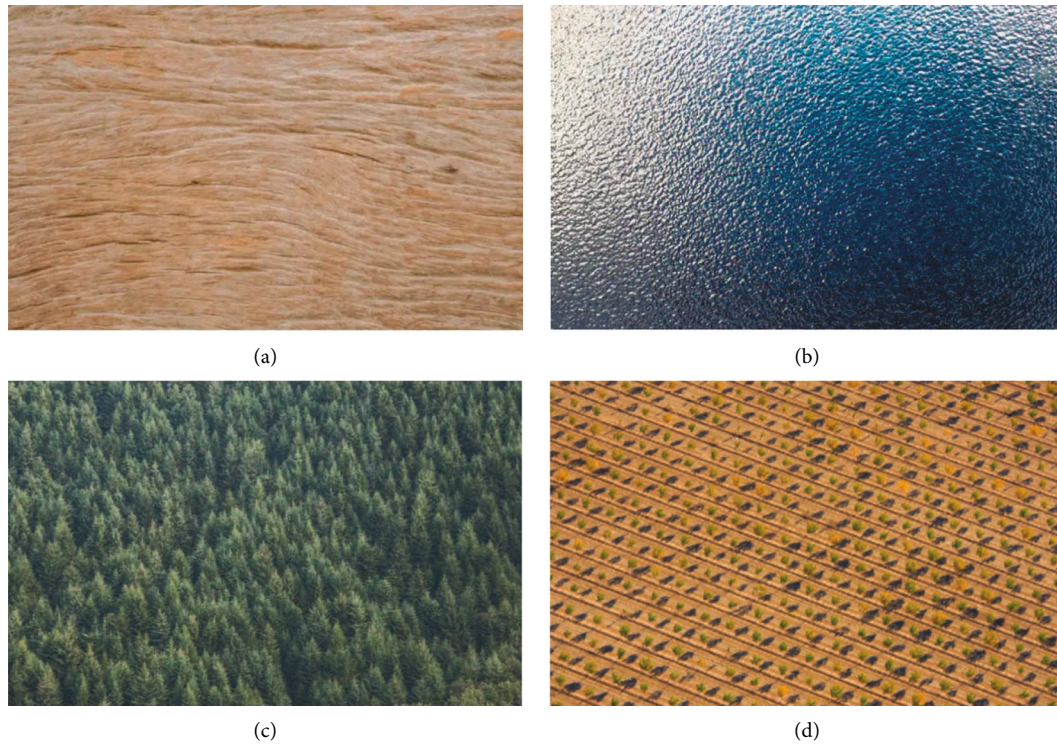


FIGURE 1: The sample of single-texture images. (a) Wood; (b) Lake; (c) Forest; and (d) Farmland.

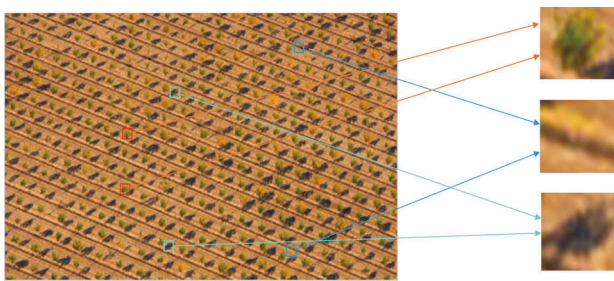


FIGURE 2: Texture feature and its textons.

The quality of its extraction has an effect on the accuracy of the results obtained from texture classification, texture identification, and any other subsequent texture image processing. This is because the accuracy of these outcomes is dependent on the extracted texture. A good texture feature should have the advantages of small computation, small feature size, and strong discrimination capacity, and it should also be able to be employed in engineering. These are the characteristics of a good texture feature. The traditional method of texture classification has progressed to a more developed stage and is widely applied in engineering.

After decades of development, the traditional method has reached a high level of maturity and has seen widespread application in the field of engineering. Researchers have made significant strides in the area of texture feature extraction and have proposed a number of different methods to extract texture features. These methods include fuzzy subspace clustering, K-mean, meanshift, Gaussian mixture

model, LBP, fractal model, and wavelet. The research on texture feature extraction methods is booming and has great prospects, but the traditional methods have limitations as a result of the hazy definition of texture and the high complexity of texture. This is due to the fact that texture can be difficult to define. This paper will focus primarily on the analysis of the performance of various texture clustering algorithms, as well as the comparison of that performance to the performance of traditional texture classification algorithms in terms of image size, clustering number, running time, and accuracy. Additionally, the performance boundaries of various algorithms will be determined in order to locate the focus for subsequent improved algorithms. In the following section, we will discuss a number of traditional texture clustering algorithms in order to compare and contrast their performance with other algorithms in the experimental section.

3. Typical Texture Clustering Algorithms

3.1. K-Means. The K-Means algorithm is a method of unsupervised learning and clustering that is founded on the concept of partitioning. The Euclidean distance is the standard indicator that is utilized in the process of measuring the degree of similarity between different data samples. The distance between the data objects has an adverse effect on the similarity, which is expressed as a negative proportion. The greater the degree of resemblance, the closer together the points fall. The procedure requires the starting number of clusters, k , as well as the initial cluster centers to be specified in advance. The position of the cluster center is

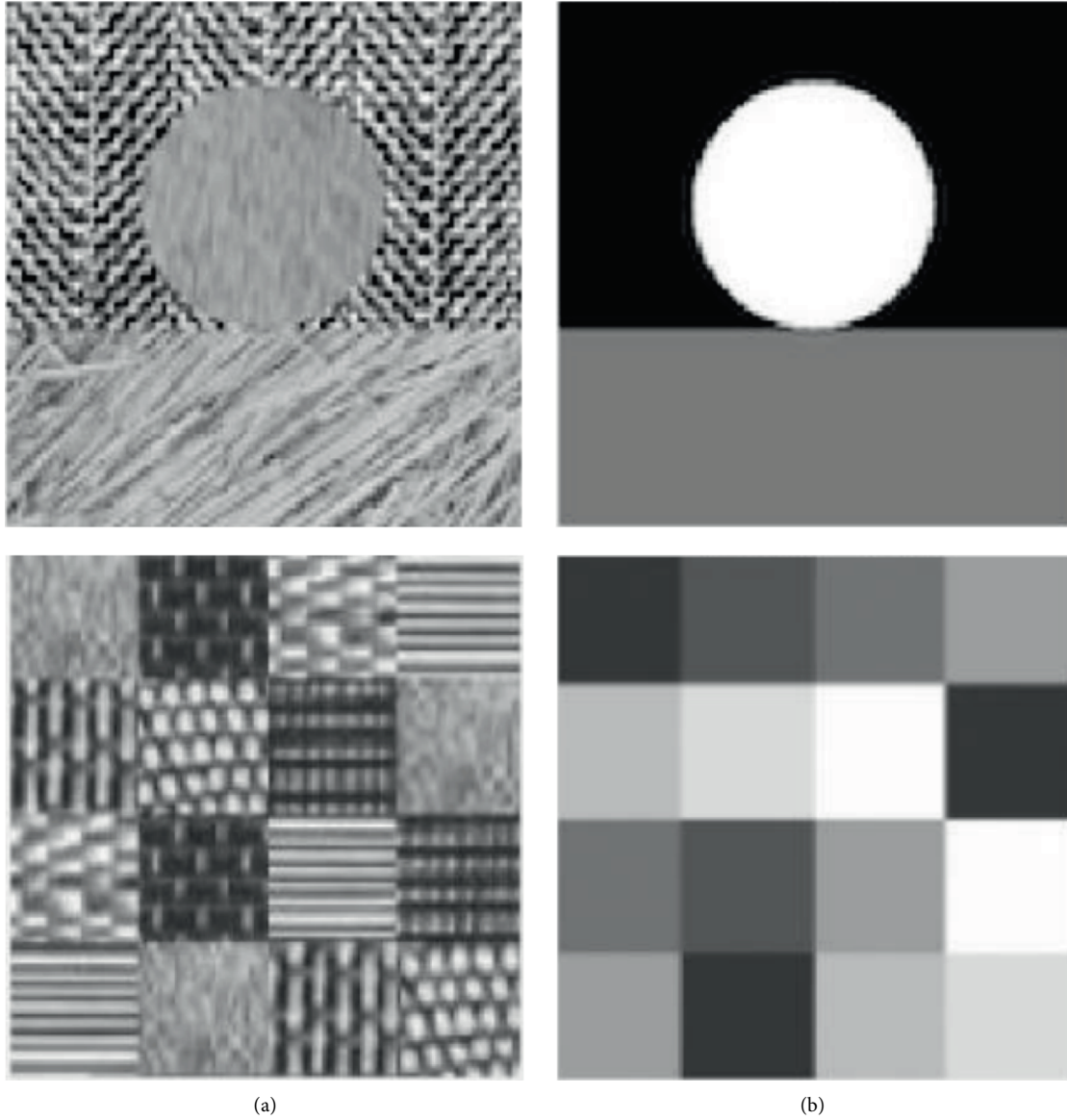


FIGURE 3: Texture images with multiple categories and its ground-truth. (a) Multi-texture images. (b) Ground-truth.

continuously updated, and the sum of squared error (SSE) of the clusters is continuously lowered, all in accordance with the degree to which the data item and the cluster center are comparable to one another. The clustering process is complete and the ultimate outcome is attained when either the SSE ceases to experience any changes or the objective function converges.

The flow diagram for K-mean cluster is shown in Figure 4.

$$d(x, C_i) = \sqrt{\sum_{j=1}^m (x_j - C_{ij})^2}, \quad (1)$$

where x is sample data; C_i is the i -th cluster center; m is the dimension of the data sample; x_j and C_{ij} is the j -th attribute value of x and C_i .

3.2. Fuzzy C-Means. One of the most popular unsupervised clustering algorithms is called the K-means algorithm. This is due to the fact that it is both efficient and simple to use. It employs an iterative methodology and serves a sizable user population. On the other hand, the K-means algorithm is a challenging approach to clustering that calls for the number of cluster categories and classification groups to be determined in advance. This method has several limitations that must be taken into consideration in the event that there is uncertainty regarding the classification categories. The Fuzzy C-means (FCM) approach is a type of unsupervised data clustering that utilizes membership in order to determine which category a data sample is a part of in order to classify the sample. In contrast to the k-means algorithm, the fuzzy clustering method, also known as FCM, is not as much of a fuzzy clustering method as it is a hard one. When using fuzzy clustering, it is not necessary for each data sample to be distinctly arranged within

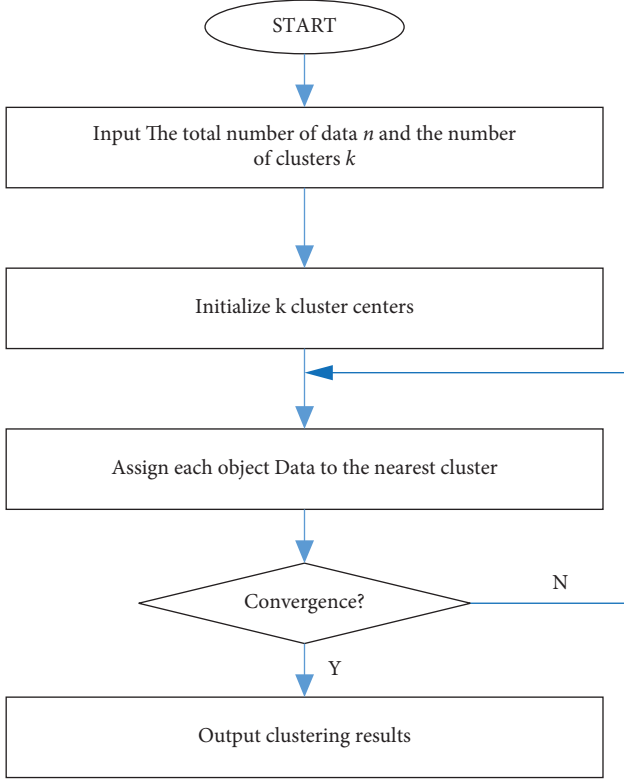


FIGURE 4: Flow diagram for K-mean cluster.

a single category. On the other hand, as shown in Figure 5, the category can be classified in an infinite variety of different ways. When it comes to dealing with uncertain or hazy categories, the fuzzy clustering method, also known as the FCM, is one of the most prominent lines of clustering development and offers a number of benefits. The following equations can be used to express clustering based on the FCM:

$$\min J(u, v) = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - v_j\|^2, \quad (2)$$

subjected to. $\sum_{j=1}^C u_{ij} = 1, u_{ij} \geq 0,$

where m is the membership factor, usually taken as 2; $\|x_i - v_j\|^2$ represents the Euclidean distance from the current data point to the center point; x_i is the current pixel point, v_j is the cluster center point, N is the total number of data points, C is the contour curve, and (2) indicates that the sum of membership degrees of all categories is 1.

The process of texture image segmentation can be regarded as the type of sample data belongs to which kinds of textures. For texture images, the pixels have uncertain characteristics. Therefore, it is not appropriate to use the hard clustering method to divide the categories of pixels. FCM can better use the image information and apply the fuzzy relationship to the texture image, which can have a better and more accurate cluster & segmentation effect.

3.3. Gaussian Mixture Model. The general expression of Gaussian mixture model can be rewritten as:

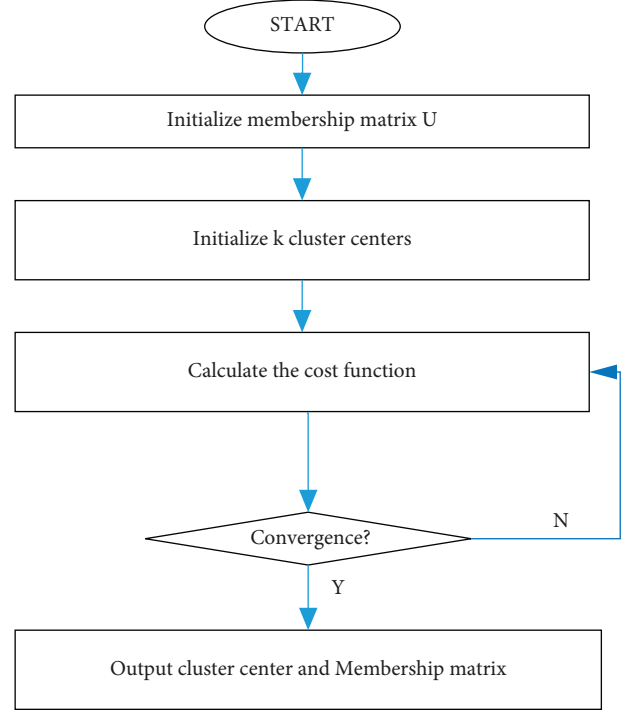


FIGURE 5: Flow diagram for Fuzzy C-mean cluster.

$$P(x) = \sum_{k=1}^K \alpha_k \varphi(x | \theta_k), \quad (3)$$

where K represents the number of Gaussian distributions that make up the mixed distribution, i.e., K clusters; α_k represents the mixing coefficient of the k -th cluster, $\alpha_k > 0$ and $\sum_{k=1}^K \alpha_k = 1$; $\varphi(x | \theta_k)$ represents the Gaussian density, where $\theta_k = (u_k, \sigma_k^2)$ is denoted as follows:

$$\varphi(x | \theta_k) = (2\pi)^{-d/2} (\sigma_k^2)^{-1/2} \exp\left(-\frac{(x - u_k)^T}{2\sigma_k^2} (x - u_k)\right), \quad (4)$$

u_k, σ_k^2 represent the mean and covariance matrix of the k -th cluster, respectively; d represents the data dimension.

Gaussian mixture model adopts EM algorithm to estimate parameters. EM algorithm consists of E-step and M-step. In other words, it maximizes the log likelihood function of incomplete data to estimate parameters of Gaussian mixture model.

The EM algorithm gradually improves the parameters of the model by utilizing the EM steps in a randomized order. This results in a steady increase in the likelihood probability of the parameters as well as the training samples, and the program finally ends at the maximum point.

3.4. Mean-Shift. Mean-shift is an approach for density clustering that can be utilized as a way for segmenting texture images. Mean-shift is referred to as "shifting the mean." The density gradient is utilized in mean-shift in order to estimate the parameters of the samples, and the kernel function is utilized in order to weight the samples. The

nonparametric kernel density estimation approach that Silverman suggested offers a methodical demonstration. Mean-shift makes use of the common kernel function principle in order to propose a kernel-based density estimation algorithm. This algorithm assigns weights to samples within each bandwidth in such a way that the contribution of the offset to the mean-shift vector varies depending on the distance between the sample and the offset point. In other words, the kernel-based density estimation algorithm uses the common kernel function principle.

Given n sample points $x_i, i = 1, 2, \dots, n$ in the d dimensional space R^d , the mean-shift vector at the points can be written as follows:

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x), \quad (5)$$

The Mean-shift algorithm is extended to the following form:

$$M_h(x) = \frac{\sum_{i=1}^n G((x_i - x)/h)w(x_i)(x_i - x)}{\sum_{i=1}^n G((x_i - x)/h)w(x_i)}, \quad (6)$$

where $w(x) \geq 0$ is the weight of sample point x ; $G(x_i)$ is a unit kernel function, $G(x) = g(\|x\|^2)$; h is the bandwidth of the kernel function.

3.5. Fuzzy Subspace Clustering. In order to cluster high-dimensional data sets, a new approach called fuzzy subspace clustering (FSC) has been devised. This algorithm was inspired by fuzzy clustering and LAC. The FSC algorithm locates the subspace clusters in which every dimension of the initial data is linked to each cluster in terms of the probability or weight. In a dimension, the weight that is to be attributed to the dimension is proportional to the cluster density. The higher the cluster density, the bigger the weight. In other words, each cluster has an association with all dimensions of the data that was initially collected.

One of the subcategories of clustering methods is referred to as fuzzy subspace clustering. In addition to having strong denoising capabilities, it is capable of transforming high-dimensional data into useful subspaces for clustering. The following is an expression that can be used to describe the objective function of fuzzy subspace clustering:

$$J_{\text{FSC}} = \sum_{k=1}^K \sum_{i=1}^d w_{ki}^T (u_{kj}(x_{ji} - v_{ki})^2 + \epsilon_0),$$

$$\text{subjected.to. } u_{kj} \in \{0, 1\}, \sum_{k=1}^K u_{kj} = 1; 0 < \sum_{j=1}^n u_{kj} < N,$$

$$0 \leq w_{ki} \leq 1; \sum_{i=1}^d w_{ki} = 1, \quad (7)$$

where K , N , and d represent the number of clusters, the number of samples and the characteristic dimension of samples, respectively; ϵ_0 is a small regularization constant. In order to reflect the characteristics of subspace clustering, the weight vector $w_k, k = 1, 2, \dots, K$ for each cluster is designed,

where $w_{ki} \in w_k$ indicates the contribution of the i th feature to the k -th cluster.

Given training data set $D_{tr} = (x_i, y_i)$, $x_i \in R^d$, $y_i \in R, i = 1, 2, \dots, N$, $X = \{x_1, x_2, \dots, x_N\}$ is divided into k classes by FSC algorithm, which is corresponding to K fuzzy rules. Each feature selected by each fuzzy rule corresponds to a fuzzy subset A_i^k . If the Gaussian function is used as the membership function, the corresponding Gaussian function parameters for A_i^k can be estimated as follows:

$$c_i^k = \sum_{j=1}^N u_{kj} x_{ji}^k / \sum_{j=1}^N u_{kj},$$

$$\delta_i^k = h \sum_{j=1}^N u_{kj} (x_{ji}^k - c_i^k) / \sum_{j=1}^N u_{kj}, \quad (8)$$

where x_{ji}^k denotes the i th feature selected by the j -th sample $x_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T$ from the k -th fuzzy rule; u_{kj} represents whether the j -th sample belongs to the K -th cluster.

3.6. Maximum Entropy Clustering Algorithm. The Maximum Entropy Clustering Algorithm (also known as MEC) is one of the clustering algorithms that is considered to be among the most representative. The mathematical representation of it is straightforward, and the physical significance of what it means is unmistakable. It is an algorithm for clustering that is frequently employed by academics. The MEC clustering algorithm has better denoise than the classical fuzzy C-means clustering, which makes it possible to obtain better clustering and brings the segmentation results closer to the ground-truth results. This is especially useful in the segmentation of texture images that contain noise. The expression of the function that the MEC algorithm uses is as follows:

$$\min_{U,V} \left(\sum_{i=1}^C \sum_{j=1}^N u_{ij} \|x_j - V_i\|^2 + \lambda \sum_{i=1}^C \sum_{j=1}^N u_{ij} \ln u_{ij} \right), \quad (9)$$

$$\text{subjected.to. } 0 \leq u_{ij} \leq 1 \sum_{i=1}^C u_{ij} = 1, 1 \leq i \leq C, 1 \leq j \leq N.$$

4. Comparative Performance Analysis

Traditional methods have been refined over the course of many decades, which has resulted in their high level of maturity and widespread application in engineering. The primary objective of this study is to evaluate and assess a wide range of existing typical texture classification algorithms, evaluate and assess their performance with regard to image size, clustering number, running time, and accuracy, and determine the performance boundaries of various algorithms in order to determine the focus for subsequent improved algorithms.

4.1. Data Sources for Texture Classification. In this study, the Brodatz texture images found in the public database serve as the basis for feature extraction and classification exercises involving a variety of texture clustering techniques. 112

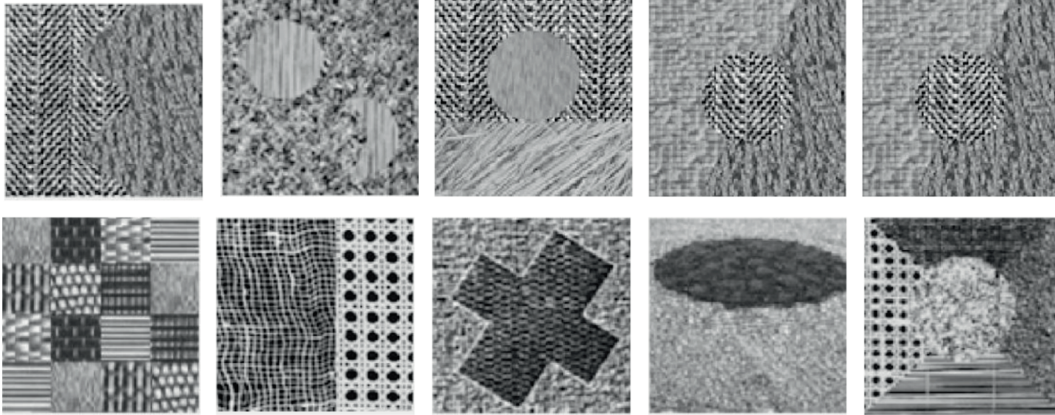


FIGURE 6: The sample of ten representative Brodatz texture images.

distinct grayscale texture images are included in the Brodatz texture dataset. Each image has a resolution of 640 pixels by 640 pixels and 8 bits. Figure 6 depicts the 10 representative texture images that were chosen for the study, and they are available for download at <http://www.uis.no/tranden/brodatz.html>. This selection was made to make the analysis more manageable. The Brodatz texture image that we have chosen is a composite image that is made up of seven individual basic texture images, and the size of the composite image has been resized to match the size of the individual basic texture photographs. In order to create an accurate representation of the environment in which the real dataset was collected, Gaussian noise with varying standard deviations was applied to each texture image.

For the purpose of the experiment, each texture image with a resolution of 640 by 640 pixels needs to be segmented into sixteen nonoverlapping subimages of either 100 by 100 or 124 by 124 pixels. To put that another way, the experiment requires a library of texture images with 640 different subimages, but we only chose 10 representative photos to analyze. In accordance with the experimental steps, the feature of each sub-image is extracted. In the first step, a Gabor filtering preprocessing operation is carried out on the grayscale image, and this operation extracts the features of the image's texture.

4.2. Experimental Setup

4.2.1. Comparison Algorithm. After decades of growth, the traditional technique has reached a high level of maturity and has seen widespread use in the field of engineering. In the area of texture feature extraction, researchers have made significant strides, leading to the development of new algorithms such as K-means clustering algorithm, Fuzzy C-Means algorithm, Gaussian mixture model, Mean-Shift, Fuzzy Subspace clustering, and Maximum Entropy Clustering Algorithm. All of these algorithms were proposed by the researchers. The theory behind several algorithms as well as the technique of putting them into practice was presented in Section 2 of this study. It is important to note that all of the comparison algorithms used for this research employ open source MATLAB code, and the default values for their

parameters. This makes quantitative analysis of following trials more convenient. The operational system is a 32 bit version of Windows 10, and the programming environment is MATLAB 7.10.0.499. The experimental hardware consists of an Intel Core i5-7240 CPU, which has a basic frequency of 3.40 GHz and 4 GB of memory (R2017a).

4.2.2. Evaluation Criteria. In order to quantitatively evaluate the performance of different comparison texture clustering algorithms, the experiment adopts running time, Rand index (RI) and Normal Mutual Information (NMI) for quantitative analysis. The calculation equation of RI and NMI is shown in the following equations:

$$\text{NMI} = \frac{\sum_{i=1}^c \sum_{j=1}^c (N_{ij} \log N \cdot N_{ij}) / N_i \cdot N_j}{\sqrt{\sum_{i=1}^c N_i \log N_i / N} \sqrt{\sum_{j=1}^c N_j \log N_j / N}}, \quad (10)$$

$$\text{RI} = \frac{f_{00} + f_{11}}{N(N-1)/2}. \quad (11)$$

4.3. Performance Analysis for Different Texture Clustering Algorithms. In our experiment, we begin by applying Gabor preprocessing to the texture image in order to reduce the amount of interference from the surrounding noise. Based on the preprocessed data, we next employ a variety of approaches to extract features and cluster texture. With the Gabor filter, you can accurately depict and identify various textures. Using a Gaussian kernel function in the spatial domain, a two-dimensional Gabor filter is equivalent. Because of the multiplicative convolution feature, the Fourier transform of the Gabor filter impulse response is a convolution of the harmonic function Fourier transform and the Gaussian function Fourier transform. In order to construct the filter, two parts must be used: a real one and an imaginary one that are orthogonal to one another. As a result, there are three stages to the texture clustering process. The first step in the clustering process is to perform picture preprocessing using the Gabor method to reduce the amount of noise in the image. Extracting texture feature vectors from each pixel in an image, and then using these vectors to

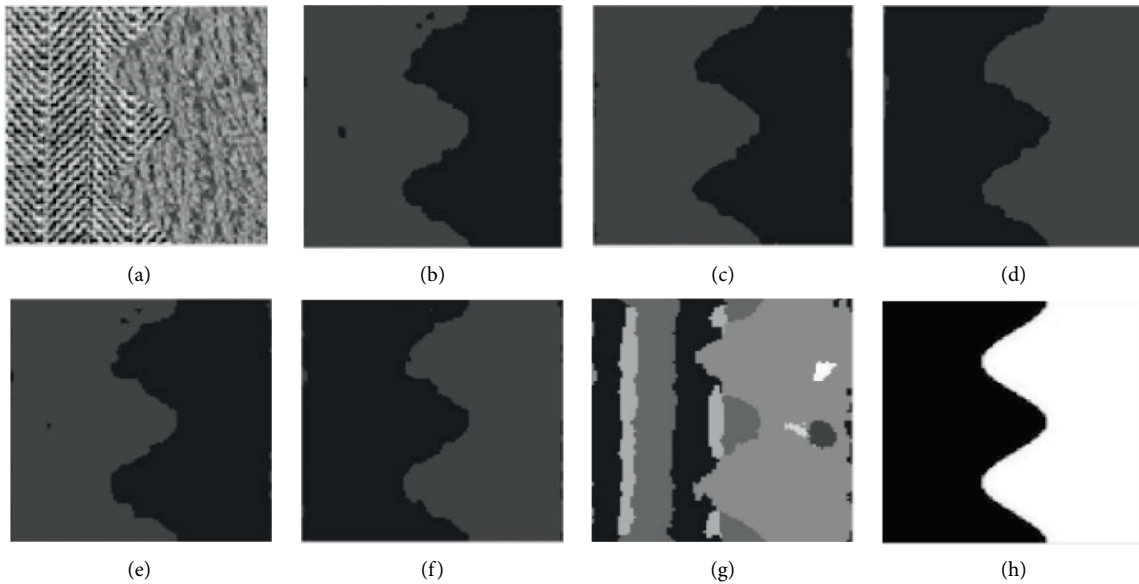


FIGURE 7: Clustering results of different algorithms in texture image t21. (a) Raw image. (b) FCM. (c) FSC. (d) GMM. (e) MEC. (f) K-means. (g) Mean-shift. (h) Ground-truth.

generate picture features, is the second phase. The final clustering results are obtained in the third stage, which involves applying a variety of clustering approaches. The texture feature vectors of each pixel in the image are processed using these techniques. It is possible to produce a clustered image by separating pixels into multiple clusters and then using the texture clustering to display the resulting image with different gray values for each of the different clusters. As part of our evaluation of the various texture clustering algorithms, we will undertake both qualitative and quantitative data analysis.

In comparative experiment, we selected 10 representative images for analysis. However, due to space limitations, this section only selects the clustering results of 6 texture images for qualitatively discussing the results. Figures 7–11 are the clustering results of different algorithms for different texture images.

Figure 7 is the clustering result of the texture image t21. It can be clearly seen from the raw image that the texture image should be divided into two categories, but except that the results of mean-shift are messy, FCM, FSC, GMM, MEC, and K-means can get more accurate results. The boundary of FSC and GMM is smooth, while other results are rough and misclassified. The essence of the two-dimensional K-means model is that it draws a circle with the center of each cluster. In other words, the center of the circle is the maximum Euclidean distance from the center of a cluster to the center of another cluster as the radius. That means it truncates the training set by a circle. Moreover, K-means requires that the shape of these clusters must be circular. Therefore, the cluster fitted by K-means model is very different from the actual data distribution (maybe ellipse), and multiple circular clusters are often mixed and overlapped with each other. In general, K-means has two shortcomings, which makes its fitting effect on many data sets (especially low

dimensional data sets) unsatisfactory: the shape of texture is not flexible enough, the fitting results are quite different from the actual results, and the accuracy is limited.

Figure 8 is a clustering result of the texture image t22. It can be seen that the raw image has three textures, and the difference is small, GMM and K-means can get more accurate results, while FCM, FSC, MEC, and mean-shift have poor clustering results. There are a large number of misclassification phenomena in the background region of FCM. From the results, it can be clearly seen that FCM algorithm divides the upper and lower triangular regions very disorderly. MEC and mean-shift classify the two regions into one class in the texture image.

Figure 9 is a clustering result of the texture image t31. It is obvious in the result that although FCM algorithm has segmented three categories in texture clustering, there is a problem in the segmented region. The upper rectangle and the lower rectangle are not accurately segmented. However, compared with the segmentation results of FSC and MEC algorithm, the segmentation accuracy of FCM algorithm is improved. From the results in Figure 9(f), it can be seen that the k-means algorithm directly classifies the regions with very low density as the background, while the FCM algorithm clearly divides the regions with low density. Therefore, FCM algorithm can deal with uneven density region well.

Figure 10 is a clustering result of the texture image. From the clustering results, it can be seen that FSC and MEC algorithms classify the edges of two adjacent classes into two classes, K-means algorithm segments the edges and background together, and FCM algorithm classifies the edges and squares into one class. So FCM algorithm can also deal with the uneven edge region.

As can be seen from the clustering results in Figure 11, although the FSC and MEC algorithms segment the texture of the middle region, they do not accurately segment the

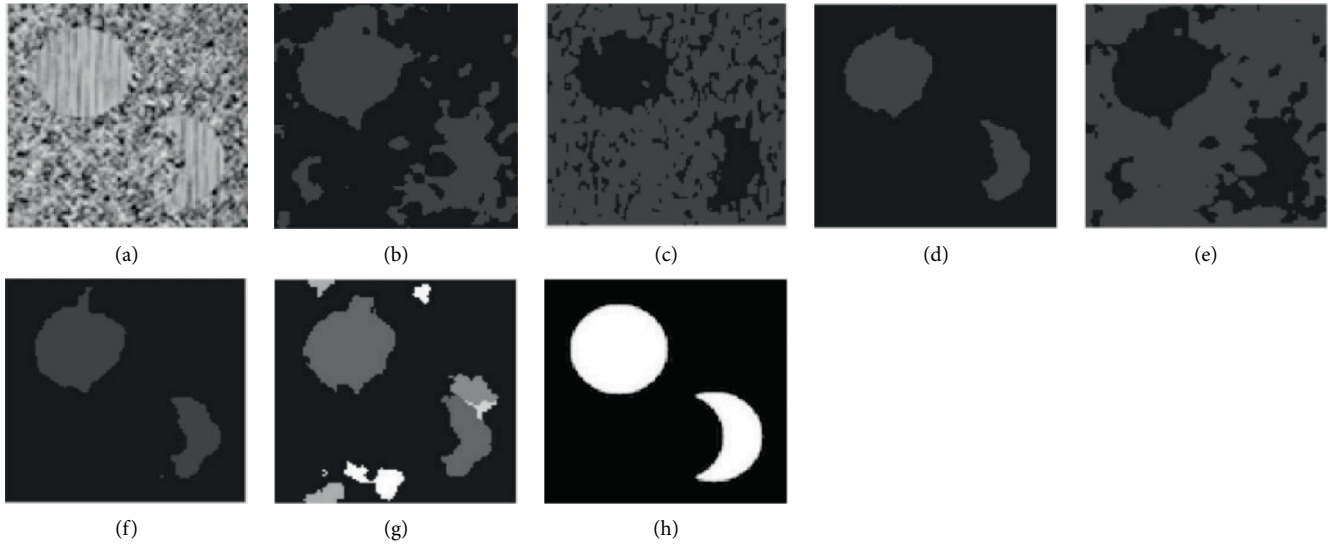


FIGURE 8: Clustering results of different algorithms in texture image t22. (a) Raw image. (b) FCM. (c) FSC. (d) GMM. (e) MEC. (f) K-means. (g) Mean-shift. (h) Ground-truth.

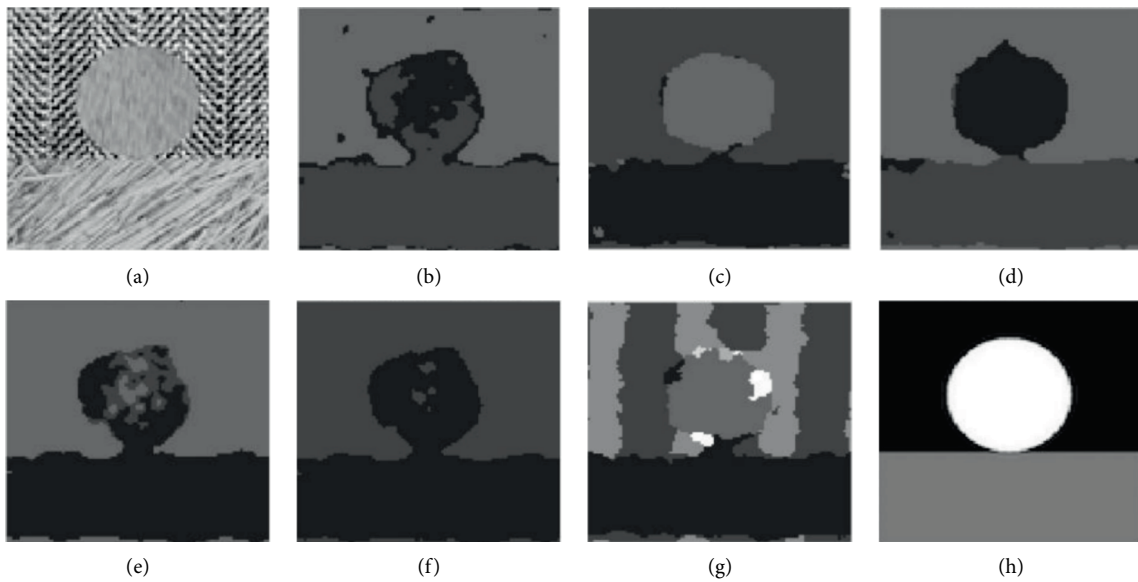


FIGURE 9: Clustering results of different algorithms in texture image t31. (a) Raw image. (b) FCM. (c) FSC. (d) GMM. (e) MEC. (f) K-means. (g) Mean-shift. (h) Ground-truth.

texture of the edge region. Figure 11 is a square region with multiclass textures. From the ground-truth clustering results, FCM, FSC, GMM, and K-means can be used in some grid regions, but the results are not ideal. In particular, the category of boundary area is not accurate, resulting in unclear boundary. Mean-shift divides several different textures into the same category.

According to the above clustering results, compared with the image segmentation results of K-means algorithm, FCM, FSC, and mean-shift algorithm, GMM algorithm can reduce the number of wrong segmentation regions and obtain good segmentation results. Gaussian mixture model (GMM) can be regarded as an optimization of K-means model. It is not only a common technical means in industry, but also a

generative model. Gaussian mixture model tries to find the mixed representation of multi-dimensional Gaussian probability distribution, so as to fit arbitrary shape data distribution.

In order to quantitatively evaluate the performance of different comparison texture clustering algorithms, the experiment adopts running time, Rand index (RI) and normal mutual information (NMI) for quantitative analysis. Tables 1–3 show the comparative analysis for running time, rand index, and normalized mutual information. Table 1 shows the effect of image size on the running-time results. Two different sizes of texture images are selected in the experiment. For texture image with a size of 124×124 , the running time of FCM is 22.85, while the image processing

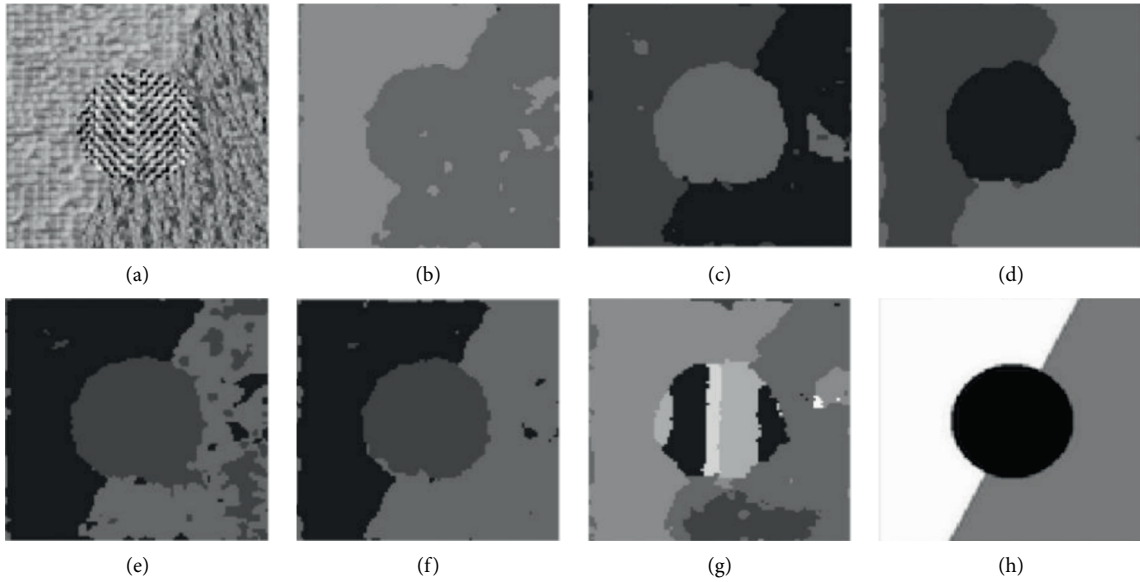


FIGURE 10: Clustering results of different algorithms. (a) Raw image. (b) FCM; (c) FSC. (d) GMM. (e) MEC. (f) K-means. (g) Mean-shift. (h) Ground-truth.

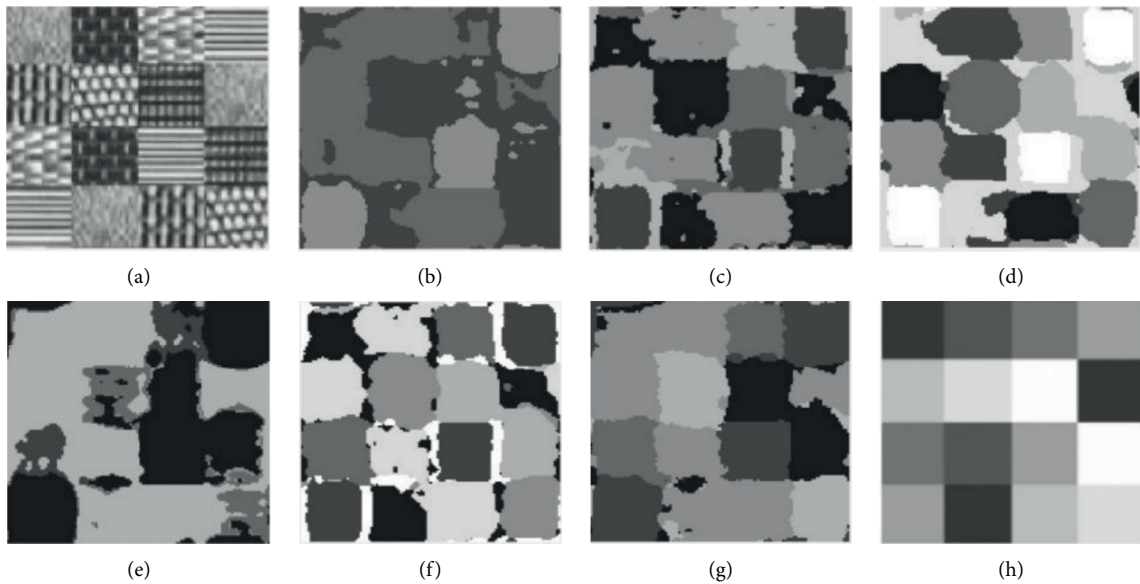


FIGURE 11: Clustering results of different algorithms. (a) Raw image. (b) FCM. (c) FSC. (d) GMM. (e) MEC. (f) K-means. (g) Mean-shift. (h) Ground-truth.

TABLE 1: Comparative analysis for running time.

Images	Size	Cluster	FCM	FSC	GMM	MEC	Kmeans	Mean-shift
T21	100 × 100	2	34.51	6.32	7.50	3.26	3.55	15.93
T22	100 × 100	2	9.11	5.59	8.10	3.42	3.45	28.42
T31	100 × 100	3	21.99	5.64	16.84	4.39	5.94	9.77
T32	100 × 100	3	18.69	7.74	8.29	4.07	4.58	15.45
T41	100 × 100	4	18.69	10.40	58.70	6.87	6.99	11.74
D1	124 × 124	7	22.85	23.34	192.58	17.77	17.16	24.93
Z2	100 × 100	5	34.13	14.24	62.53	13.12	13.30	28.61
Z3	100 × 100	2	23.65	4.34	10.39	3.74	3.44	13.62
Z4	100 × 100	2	13.87	4.21	9.93	3.93	4.06	16.45
Z5	100 × 100	2	7.62	3.25	20.93	2.73	3.42	4.80

TABLE 2: Comparative analysis for Rand Index (RI).

Images	Size	Cluster	FCM	FSC	GMM	MEC	Kmeans	Mean-shift
T21	100×100	2	0.92	0.95	0.97	0.96	0.96	0.76
T22	100×100	2	0.74	0.73	0.91	0.74	0.91	0.84
T31	100×100	3	0.84	0.88	0.91	0.81	0.78	0.80
T32	100×100	3	0.79	0.91	0.92	0.84	0.93	0.85
T41	100×100	4	0.61	0.91	0.92	0.79	0.91	0.88
D1	124×124	7	0.66	0.72	0.76	0.62	0.76	0.71
Z2	100×100	5	0.71	0.75	0.85	0.66	0.74	0.80
Z3	100×100	2	0.72	0.72	0.74	0.72	0.70	0.70
Z4	100×100	2	0.76	0.73	0.77	0.80	0.81	0.59
Z5	100×100	2	0.90	0.91	0.87	0.91	0.94	0.93

TABLE 3: Comparative analysis for normalized mutual information (NMI).

Images	Size	Cluster	FCM	FSC	GMM	MEC	Kmeans	Mean-shift
T21	100×100	2	0.76	0.85	0.88	0.85	0.86	0.52
T22	100×100	2	0.39	0.31	0.67	0.40	0.66	0.49
T31	100×100	3	0.56	0.70	0.74	0.55	0.53	0.55
T32	100×100	3	0.58	0.73	0.78	0.61	0.79	0.61
T41	100×100	4	0.34	0.74	0.75	0.51	0.71	0.64
D1	124×124	7	0.31	0.36	0.50	0.30	0.48	0.42
Z2	100×100	5	0.43	0.48	0.51	0.33	0.48	0.44
Z3	100×100	2	0.36	0.45	0.49	0.35	0.33	0.37
Z4	100×100	2	0.48	0.45	0.51	0.49	0.52	0.34
Z5	100×100	2	0.70	0.72	0.68	0.76	0.80	0.77

time of 100×100 is 18.69. In general, the running time of 124×124 pixels image is longer than that of 100×100 for any algorithm. From this, we can conclude that the size of the image determines the length of the running time. The larger the size of image, the slower the running time.

In this experiment, a variety of experiments are carried out on the number of clusters. Five synthetic texture images with two clusters, two images with three clusters, one image with four clusters, one image with five clusters, and one image with seven clusters are selected. For example, there are seven clusters in D1. As for the RAND index, when the number of clusters is more, the value of the RAND index is inversely proportional. In other words, the more the number of clusters, the lower the value of the RAND index. In normalized mutual information, we can also see that when the number of clusters is greater, the value of normalized mutual information is inversely proportional. The more the number of clusters, the lower the value of NMI as shown in Tables 1-2.

In texture image D1 with the size of 124×124, it can be seen that GMM algorithm and K-means algorithm have better effects. For the three evaluation indicators, GMM segmentation has the best effect. It can be seen from Table 1 that MEC algorithm has the fastest running time, but the GMM algorithm has the best clustering results for Rand index and normalized mutual information, but MEC is not the worst. Therefore, the correlation between running time and clustering performance index is not strong. By comparing result of five images with two clusters and result of two images with three clusters, it is concluded that when the image texture is relatively independent, it will be easier to segment. When a texture in image contains another texture, it will cause large errors and cause misclassification.

Through the data analysis in Tables 2 and 3, each algorithm can find that there is a proportional relationship between Rand index and normalized mutual information, which is the embodiment of the clustering performance. When one value is larger, the other is larger, and the value of rand index is larger than that of normalized mutual information.

5. Conclusion and the Future Work

Traditional texture cluster algorithms have seen widespread application in engineering, but despite this, there are ongoing challenges associated with their high complexity and limited universal applicability. This paper focuses primarily on the analysis of the performance of various texture clustering algorithms. It also compares the performance of traditional texture classification algorithms in terms of image size, clustering number, running time, and accuracy. Finally, it determines the performance boundaries of various algorithms in order to locate the focus for subsequent improved algorithms. For the purposes of the experiment, a number of traditional clustering algorithms have been chosen to serve as benchmarks for evaluating overall performance. Both qualitative and quantitative findings point to the fact that the performances of various algorithms are quite distinct from one another. Better performance can only be achieved by selecting the appropriate algorithm in accordance with the characteristics of the image's texture. In the future, we are going to conduct an analysis of the texture segmentation algorithm that is based on deep learning. In addition, we are going to conduct an in-depth comparison of the conventional algorithm and the intelligent algorithm in

order to determine the conditions under which they can adapt and the boundaries of their performance.

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 declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61862051), the Science and Technology Foundation of Guizhou Province (No. [2019]1299), the Top-notch Talent Program of Guizhou province (No. KY[2018]080), the Natural Science Foundation of Education of Guizhou province (No. [2019]203) and the Funds of Qiannan Normal University for Nationalities (Nos. qnsy2018003, qnsy2019rc09, qnsy2018JS013, and qnsyrc201715).

References

- [1] A. Vailaya, M. Figueiredo, A. Jain, and Hong-Jiang Zhang, "Image classification for content-based indexing," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 117–130, 2001.
- [2] M. Ohi, Y. Li, Y. Cheng, and T. Walz, "Negative staining and image classification — powerful tools in modern electron microscopy," *Biological Procedures Online*, vol. 6, no. 1, pp. 23–34, 2004.
- [3] S. Bandyopadhyay and U. Maulik, "Genetic clustering for automatic evolution of clusters and application to image classification," *Pattern Recognition*, vol. 35, no. 6, pp. 1197–1208, 2002.
- [4] M. N. Robert, "Image classification by a two-dimensional hidden Markov model," *IEEE Transactions on Signal Processing*, vol. 12, no. 9, pp. 988–992, 2000.
- [5] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 19, no. 6, pp. 1657–1663, 2010.
- [6] R. Manthalkar, P. K. Biswas, and B. N. Chatterji, "Rotation invariant texture classification using even symmetric Gabor filters," *Pattern Recognition Letters*, vol. 24, no. 12, pp. 2061–2068, 2003.
- [7] T. Kasetkasem, M. K. Arora, and P. K. Varshney, "Super-resolution land cover mapping using a Markov random field based approach," *Remote Sensing of Environment*, vol. 96, no. 3–4, pp. 302–314, 2005.
- [8] R. Paget and I. D. Longstaff, "Texture synthesis via a noncausal nonparametric multiscale Markov random field," *IEEE Transactions on Image Processing*, vol. 7, no. 6, pp. 925–931, 1998.
- [9] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2570–2583, 1992.
- [10] X. Chen, B. Neubert, Y. Q. Xu, and S. B. Kang, "Sketch-based tree modeling using Markov random field," *ACM Transactions on Graphics*, vol. 27, no. 5, pp. 345–356, 2008.
- [11] X. Liu, D. L. Langer, M. A. Haider, Y. Yang, M. N. Wernick, and I. S. Yetik, "Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class," *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 906–915, 2009.
- [12] F. Zhou, Q. Chen, B. Liu, and G. Qiu, "Structure and texture-aware image decomposition via training a neural network," *IEEE Transactions on Image Processing*, vol. 29, pp. 3458–3473, 2020.
- [13] 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.
- [14] M. V. Boland and R. F. Murphy, "A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells," *Bioinformatics*, vol. 17, no. 12, pp. 1213–1223, 2001.
- [15] G. H. Liu, L. Zhang, Y. K. Hou, Z. Y. Li, and J. Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010.
- [16] M. Li, J. Zhou, D. Wang, P. Peng, and Y. Yu, "Application of clustering-based analysis in MRI brain tissue segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2022, p. 1, 2022.

Research Article

A Mental Health Management and Cognitive Behavior Analysis Model of College Students Using Multi-View Clustering Analysis Algorithm

Danhui Dong  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 18 August 2022; Revised 4 September 2022; Accepted 6 September 2022; Published 27 September 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Danhui Dong 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.

In this new era that is full of social changes, ongoing economic transformation, an abundance of information resources, and a fast pace of life, the pressure that people feel to compete with one another is also increasing day by day. Because of the vast differences in people's states of consciousness and worldviews, interpersonal relationships have become increasingly difficult to navigate. Students in higher education institutions will eventually emerge as the dominant demographic in society. Their mental health has a significant bearing on all aspects of life, including learning and future growth. An objective condition that must be met in order to guarantee that the next generation of talent will have a high level of overall quality is the improvement of the mental health of college students (CSMH) in the new era. One component of public health is the emotional well-being of students in higher education. The state of the public's health is consistently ranked among the most urgent problems facing modern society. However, there is not much hope for the Chinese CSMH. In order to effectively manage their mental health, a variety of educational institutions, including colleges and universities, have proposed a large number of management strategies for CSMH. The vast majority of these strategies are not targeted, and they do not offer a variety of management strategies that are based on the many different psychological states. It is necessary to first be able to accurately predict the mental health status of each individual college student in order to achieve the goal of improving the mental health management of students attending colleges and universities. This study proposes using a multi-view K-means algorithm, abbreviated as MvK-means, to analyze the CSMH's data on mental health. This is possible because the data can be obtained from multiple perspectives. This paper presents a multi-view strategy as well as a weight strategy in light of the fact that each point of view contributes in its own unique way. Different weight values should be assigned to each view's data, which will ultimately result in an improved evaluation effect of the model. The findings of the experiments indicate that the model that was proposed has a beneficial impact on the analysis of the data pertaining to the mental health of college students.

1. Introduction

Students have a unique opportunity to shine on the college stage. College students have reached a point where their physical development is comparable to that of adults, but their mental health has not reached this level of development yet. The atmosphere that the students have been exposed to up to this point has been one that is comparable to the actual social atmosphere. When college students are forced to balance the demands of school, life, and work, their mentalities will shift in

a variety of different ways. A number of psychological issues could arise if the various pressures the students are subjected to are not alleviated in a timely and appropriate manner. If these issues are not resolved in a timely manner, it will lead to other psychological issues, which will interfere with the student's ability to study and function in daily life while in college as well as in the future. Students at a wide variety of colleges and universities have been involved in a disturbingly high number of cases of self-harm and suicide in recent years. Some students are unhappy as a result of a disagreement with a fellow student

who resides in the same dorm or attends the same class, and finally, as a consequence of a few insignificant occurrences, there is a wounding incident. Some students are subjected to a variety of pressures, and because they are unable to concentrate on it for an extended period of time, they are at risk for committing suicide. All of these predicaments can be traced back to the mental health issues that are prevalent among college students. The psychological well-being of students in higher education falls under the umbrella of the field of public health. Because both the nation as a whole and each individual family must invest a significant amount of time and resources to educate a college student, it would be a terrible waste if something were to happen to him or her before he or she could integrate into society and appreciate the significance of his or her own life. The problems with CSMH have now become a public health concern. As a result of this, it is clear that educational institutions, such as colleges and universities, ought to pay attention to CSMH [1–3] and efficiently manage the data pertaining to the mental health of college students. Problems with mental health have a negative impact not only on the sound physical and mental development of individuals but also on the growth of the nation and society as a whole. As a consequence of this, one of the most important responsibilities of modern college education is to pay attention to and investigate the psychological state of college students, to promptly diagnose and treat any mental illnesses that may be present in college students, and to provide appropriate direction to college students so that they can overcome psychological barriers and treat mental illnesses.

The research that is carried out on the mental health management of college students has the potential to enrich and improve the system that manages mental health. Countries from Europe and the Americas quickly established themselves as pioneers in the field of school mental health management as early as the eighteenth century. It is possible to attribute the rapid progress that has been made in the management of students' mental health in developed countries like the United States and Europe to the fact that the government in these countries is the primary driving force behind these advancements. On the other hand, both the connotation and the expansion of mental health management are extremely complex, and the research results frequently lack unity and the ability to be applied because they are based on hazy concepts. This is because the concepts on which they are based are not well understood. Due to the fact that the starting point for this research was so low, the research on the mental health management of Chinese college students got off to a very sluggish start in China. As a result, the country's research on the topic is far behind schedule. The content of the research has many instances of repetition, and the ideas and methods that were utilized in the research are not in a position to adequately reflect the robust scientific nature of the research. This, in turn, causes inconsistencies in the findings of the research. The majority of the research that is being done on the mental health management system of college students is research that is being done on the mental health services that are being provided to college students. This is because the mental health management system of college students is the focus of

the majority of the research that is being done. On the other hand, there is no research being done on the administration of college students' mental health from the point of view of public health. The scope of studies that are relevant is not overly broad, and the majority of the investigations are limited to relatively small samples. These studies have not demonstrated a fundamental understanding of the development law and characteristics of CSMH, nor have they suggested any effective public health management measures to improve CSMH. In addition, these studies have not suggested any concrete actions that could be taken to improve CSMH. Researchers are required to comprehend the structure of mental health management from a variety of perspectives in order to support it with varying degrees of theoretical results. This is necessary because the structure of mental health management is so open. Investigation into this topic has the potential to contribute significantly to the theoretical underpinnings of open systems approaches to the management of mental health. Research in this area is very important due to the fact that the management of the mental health of college students is an important part of the system that manages public health. The research that is being done on the mental health management system of college students is helpful in the development of the mental health management system on a more local level. The idea of a management system for mental health can be segmented into a variety of different levels and organizational structures. Due to the wide variety of cultural settings in which various strategies for managing mental health first emerged, these strategies can take on a wide variety of forms across different cultures. As a direct consequence of this, the research topic of mental health management incorporates an important element of ethnic flavor. It is especially true for a country that is composed of more than one ethnic group with each ethnic group having its own culture, and this is true even more so for the country as a whole. In addition, research on the mental health management system in China is still in its infant stages, and the majority of the studies that have been carried out thus far are of a substandard quality. At this time, the majority of them continue to merely copy foreign theory and practical experience without making any attempts at innovation or transformation in order to produce a distinctive theoretical system that is able to accommodate their very own growth. This is the case despite the fact that there are numerous opportunities for them to develop their own unique perspectives and perspectives that are grounded in their own experiences. In light of this, research into community-based mental health care management in China can be of assistance in working toward the goal of establishing mental health management at the local level.

The majority of CSMH's data collection is done through the use of questionnaires and the collection of physiological data. The traditional approach to managing mental health focuses primarily on the fundamental processes of adding new data, removing old data, and making modifications to existing data. A comprehensive data mining and analysis is not carried out using this method. As a result, it does not have much of an impact on the way CSMH is managed. It is essential to organize and analyze data regarding the mental

health of students in order to realize improvements in the efficacy of mental health education provided in schools. The psychological data of college students can be mined in-depth using a variety of methods, such as classification algorithms [4–6], regression algorithms [7–9], the cluster analysis method [10–12], big data technology [13–15], and so on. These techniques are able to perform an accurate analysis of the issues concerning the psychological information of the students. The potential and implicit information in the data can be obtained by relying on the platform of the psychological management system, which is combined with data mining technology. This allows for a large amount of data to be analyzed, which in turn helps obtain potential and implicit information. This kind of implicit information will provide major universities with a corresponding reference basis and effective solutions. Wearable technology and social networking software are utilized in the data collection process for reference [16]. The information is then entered into a system that makes predictions about people’s mental health, which allows for the prediction of the mental health of each individual student. The purpose of the research cited in reference [17] is to identify people who have psychological issues by conducting an analysis of social data using natural language processing methods and machine learning algorithms. Algorithms that are used in machine learning are utilized in reference [18] to identify mental health states and to make predictions regarding illness. Over 50 previous studies on the application of machine learning in mental health assessment were analyzed for this study. The author provides a summary of the current state of research on mental health that is based in machine learning techniques, using the studies mentioned above as support. Deep learning algorithms are used in reference [19] to mine the social data of individuals in order to determine the individuals’ mental health status. The mental health state evaluation method that is based on the deep learning algorithm has the potential to have an accuracy of evaluation that is close to 90%. In reference [20], typical machine learning algorithms are utilized in the assessment of a person’s mental health status, and the results of the assessments obtained by a variety of conventional machine learning algorithms are compared to one another.

The primary focus of the research work that was described above is the examination of data pertaining to health that was obtained from various social networking platforms. These data include some information that is either inaccurate or unqualified, both of which will have an effect on the final results of the mental health assessment. In addition, the majority of the studies referred to above conduct evaluations of psychological states by making use of professional evaluation tools, machine learning algorithms, and deep learning algorithms. There is room for advancement in terms of the accuracy of the evaluation that was obtained as a result of the method that was utilized. In light of the arguments that were presented earlier, the purpose of this paper was to propose a model for the management of the mental health of college students that is based on an algorithm for multi-view clustering. Specifically, the paper will discuss how this model will work. The following is an outline of the most important contributions made by this study: (1)

When analyzing students’ mental health, we use both the data on the students’ basic information as well as the data from the questionnaires. (2) An algorithm called MvK-means has been proposed with the intention of being used for the purpose of training a mental health management model. The default behavior of the traditional multi-view algorithm is to give equal weight to each view. This is one of the available settings. When dealing with real data, it is not appropriate to use this strategy, which presumes that each view has the same quality by default. This is due to the fact that the quality of the real data varies greatly from case to case. In light of this, a weighting strategy has been implemented in this piece of writing in order to effectively control the significance of each viewpoint. (3) Feed the collected data into the trained model in order to obtain information regarding the mental health status of college students who are enrolled in the educational institution. As a direct result of the information contained in this report, CSMH is being managed in an efficient manner. Experiments have shown that the model for the management of college students’ mental health that is based on the multi-view clustering algorithm that is used in this paper is capable of producing better results. This paper utilizes the model.

2. College Students’ Mental Health Management from the Standpoint of Public Health

2.1. College Students’ Mental Health Standards and Management. The idea of mental health standards is presenting a dynamic development trend as a direct result of the progress that has been made in human society, the economy, and science. This is the case because the concept of mental health standards is presenting a dynamic development trend. The ten mental health standards that were proposed jointly by Maslow and Mittelman are among the standards that have the greatest amount of public recognition. The standards provide a comprehensive breakdown of the various aspects of security, such as an accurate assessment of one’s own capabilities, the development of an ideal personality, the mastery of one’s feelings, the enhancement of one’s connections with others, and the promotion of personal development. “Physical, intellectual, and emotional mediation, adapting to one’s environment, having a sense of well-being, giving full play to one’s abilities at work, and living a productive life,” was how the International Conference on Mental Hygiene defined mental health in 1946. At the time, this was considered to be the standard for mental health. Over the past few years, academics in China have increased the amount of research they have done on the CSMH and have proposed new standards for the CSMH. College students should have the following mental health standards, according to the general consensus among all students: correct three views; personality integrity; positive mood; positive and harmonious interpersonal relationships; objective self-evaluation; positive sense of competition; love of life; willingness to learn; and the courage to pursue life value. Defining the mental health standards that college students

should achieve is the first and most important step toward improving the level of mental health management among college students.

Management is the process of effectively planning, organizing, leading, and controlling the resources owned by an organization within a specific environment in order for the organization to achieve the goals that have been set for the organization. These goals have been set in order for the organization to be successful. The term “CSMH management” refers to the related plans that have been formulated by higher education management institutions and schools to improve the mental health status of college students and improve the mental health level of college students according to the current situation of CSMH. The goal of these plans is to both improve the mental health status of college students and improve the mental health level of college students. The mental health of college students is going to be prioritized in the implementation of these plans. We can establish a special management department to provide psychotherapy and management for these college students who have mental health issues. The content, personnel, mechanism, mode, and measures of mental health management are the primary focal points of investigation in the studies that are currently being conducted on the topic of the management of the mental health of Chinese college students. These studies are being carried out in order to shed light on the topic of the management of the mental health of Chinese college students. There are not many studies that concentrate on the particular objectives or staged objectives of management, regulatory means, or specific evaluations. Therefore, the establishment of CSMH organization and management mechanism and the improvement of the relevant management system are issues that need to be accelerated in the academic circles. Also, these are questions that need to move much more quickly through the academic community. This study was conceived with the intention of both improving the mental health management system and bringing attention to the significant role that management plays within the mental health management system. The aforementioned circumstances served as the impetus for the development of this study.

2.2. The Principles of Mental Health Management of College Students. An in-depth analysis and application of CSMH data serve as the bedrock of efficient CSMH management. After data on the mental health of college students have been collected through a variety of channels, an intelligent model is used to analyze the data and extract potentially informative information from the data in order to evaluate the psychological state of the students. The findings of the evaluation are incorporated into the development of intervention strategies for the treatment of mental health issues experienced by college students. This strategy has the potential to improve the effectiveness and efficiency with which universities manage their student bodies. Specifically, it has the potential to improve the effectiveness of universities. The following is an illustration of the principle for the management of the mental health of college students, which can be found in Figure 1:

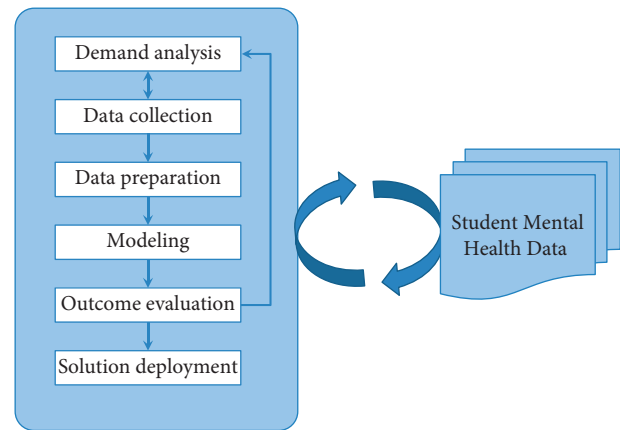


FIGURE 1: The CSMH management principle diagram.

The CSMH management process is depicted as being a component of the data mining workflow in the figure. The entirety of the procedure is connected by five links, which are as follows: demand analysis, data collection, data sorting, modeling, and the results of data evaluation.

- (1) The purpose of the stage of demand analysis is primarily to determine what useful information will be unearthed from the mental health data of college students during the course of this research. A well-defined objective serves as the foundation of demand analysis. The primary goal of this study was to gain a comprehensive understanding of college students' mental health status by analyzing data on college students' mental health. Students who may have psychological issues receive early warning, and professional psychological counselors will timely and effectively guide them according to the warning information they receive. After the objective has been defined, the necessity of this research can be summed up as the identification of psychological issues, either imminent or ongoing, that are present in college students. Once a diagnosis has been established, the appropriate medication can be prescribed based on the most successful approaches to the management of mental health.
- (2) The information data are the raw material, and because the raw material is of such high quality, the final product will also be of such high quality. As a result, the data collection process ought to be as exhaustive, in-depth, and precise as is humanly possible. For the purpose of collecting data on CSMH, it is possible to start with the fundamental information of the students, social data, psychological questionnaires, and so on. Collect data related to CSMH from a variety of perspectives in order to enrich the information collected.
- (3) Data sorting. In addition to the collection of raw data, preprocessing of that data is necessary. When it comes to the raw data that were collected, some of them are invalid data, some of them are incomplete data, and some of them are redundant data. The presence of

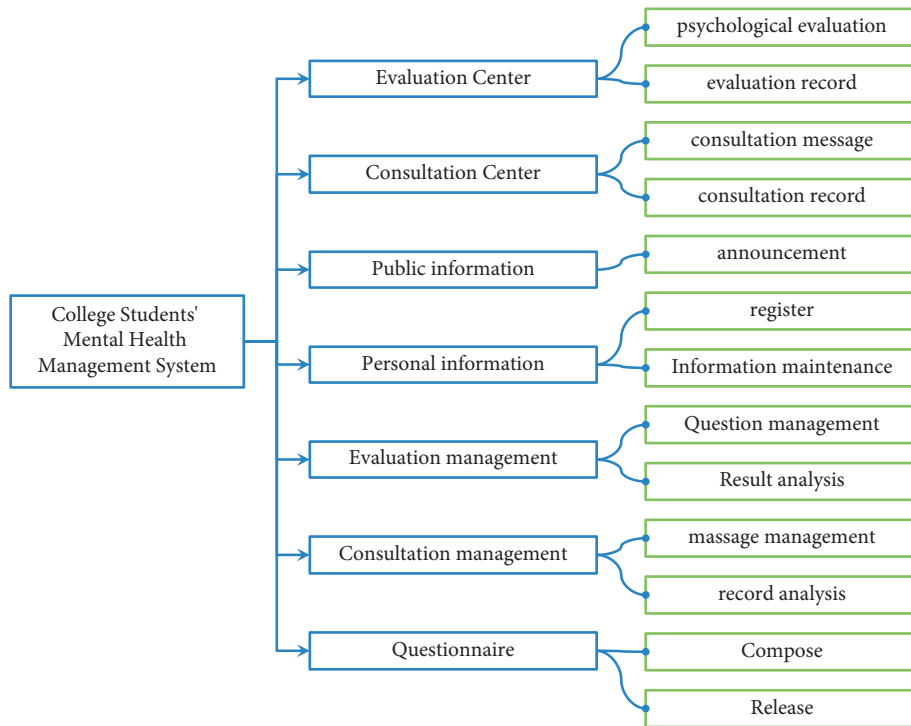


FIGURE 2: Architecture of CSMH management system.

these data will have a significant impact on the data mining process's outcome. As a result, the importance of organizing the data cannot be overstated.

- (4) Create an evaluation model. Models for data mining can be trained and created based on a selection of algorithms. A multi-view clustering algorithm was utilized for the purposes of this study. The execution steps of the algorithm consist of training a mental health assessment model using the training dataset. This model is then used to evaluate patients.
- (5) Result evaluation. The trained model is then given the test dataset to analyze, which allows the performance of the trained model to be evaluated. Calculated based on the evaluation index is the model's overall performance rating in terms of its evaluation. If the evaluation results are very positive, this indicates that the trained model is performing very well. The evaluation results that are produced by the model can provide college administrators with assistance in effectively managing college students.

2.3. Architecture of CSMH Management System. A complete mental health management system for college students should have psychological evaluation center, evaluation management, counseling center, counseling management, public information, personal information, and other functions. Among them, the evaluation center has the functions of collecting test evaluation data, psychological evaluation, and evaluation record management. Evaluation management includes evaluation data management, combined evaluation management, and evaluation result

management. The consultation center includes consultation message, online consultation, and appointment consultation. Consulting management includes appointment management, appointment scheduling, and message management. Public information includes viewing announcement information, mental health guide, public information management, and psychological article management. Personal information mainly includes registration, login, password retrieval, and personal information maintenance. Other functions mainly include file management and information management. Figure 2 depicts the system's architecture.

2.4. Classical K-Means Clustering Algorithm. On all of the data, the K-means algorithm is used to perform distance-based clustering. The primary objective of this algorithm is to determine the cluster center points for each classification with as much specificity as is feasible. Assume there are k different categories and n total data points. Calculate the distance between each sample and each class center, then choose k data at random to serve as the initial cluster center, calculate the distance between each sample and each class center, and divide the remaining data into classes as close to the class center as possible. It is necessary to continuously calculate the distance between each sample and a new round of cluster centers throughout the entire process, which is an iterative update process. The more iterations that are performed, the better the cluster centers that are obtained, up until the point where the cluster centers are no longer variable, consequently gaining possession of the last available class center. Following the selection of the class center,

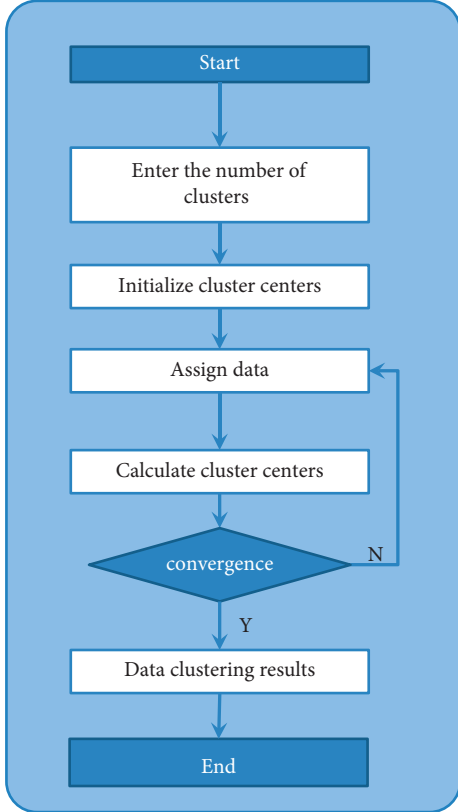


FIGURE 3: K-means algorithm flow chart.

the members of each class are chosen. Figure 3 depicts the detailed steps of the K-means clustering model [21], which are as follows:

The objective function of the K-means algorithm is as follows. The Euclidean distance is used in the function to calculate the distance between each data point and the cluster center.

$$J = \sum_{i=1}^k \sum_{j=1}^n w_{ij} d_{ij}^2, \quad (1)$$

where k is the number of categories in the dataset and n is the total number of samples in the dataset. $d_{ij} = \|x_i - z_j\|$ is used to calculate the Euclidean distance. x_i represents the i th sample and z_j represents the j th class center. w_{ij} represents the class to which the data point belongs.

$$w_{ij} = \begin{cases} 1, \\ 0. \end{cases} \quad (2)$$

If w_{ij} is 1, it means that the j th data point belongs to the class S_i . When w_{ij} is 0, it means that the j th data point does not belong to the S_i class.

Equation (1) translates to the following expression:

$$J = \sum_{j=1}^k \sum_{x_i \in S_j} d_{ij}^2, \quad (3)$$

where d_{ij} represents the sum of the squared errors of the data points and the corresponding cluster centers. Therefore,

the objective function is also called the error squared sum criterion function.

In order to obtain the cluster center z_j of each update, the following update iterative formula can be obtained according to the optimization theory of Lagrangian conditional extrema:

$$z_j = \frac{1}{n} \sum_{x_m \in S_j} x_m \quad j = 1, 2, 3, \dots, k. \quad (4)$$

3. Mental Health Assessment Models

3.1. Multi-View Clustering Model. A more effective multi-view clustering technology has been discovered through the investigation of traditional clustering analysis methods. The technology allows multi-view data with multiple features to learn collaboratively in the clustering process, solving the problem of complex data with multiple features. Traditional clustering algorithms may be limited to dealing with a single feature of complex data. Early multi-view clustering techniques took into account each view of the data and treated each view as a separate clustering task. After obtaining the clustering results for each view, the ensemble learning mechanism is used to select an appropriate ensemble learning strategy, and the results of multiple views are combined to produce the final marriage result. It has been discovered that multi-view learning makes use of the connections between various features in the dataset by studying various models of multi-view algorithms [22]. Make full use of the differences and correlations between various points of view so that the final learning outcomes are consistent. Figure 4 depicts the multi-view learning model.

3.2. Multi-View K-Means Algorithm. The evaluation center is the central component of the mental health management model for college students. The evaluation center is primarily used to collect CSMH data and enter it into the trained mental state evaluation model. The model uses the evaluation results to determine each college student's mental state. The most important of these is evaluating the model. CSMH data include many aspects, such as social data, questionnaires, academic data, personal basic information, and so on. As a result, this paper intends to train the evaluation model using a multi-view learning model.

In order to represent the data with multiple features by K-means, a weight vector W is assigned to the clustering of each view, and W satisfies the following conditions:

$$\sum_{v=1}^V w_v = 1, 0 \leq w_v \leq 1, 0 \leq v \leq V. \quad (5)$$

A multi-view dataset X has N samples and V views, $X = \{x_i\}_{i=1}^N$. $x_i = \{x_i^{(v)}\}_{v=1}^V$, $x_i \in R^{d^{(v)}}$ is the view vector of sample x_i . The objective function of MvK-means is as follows:

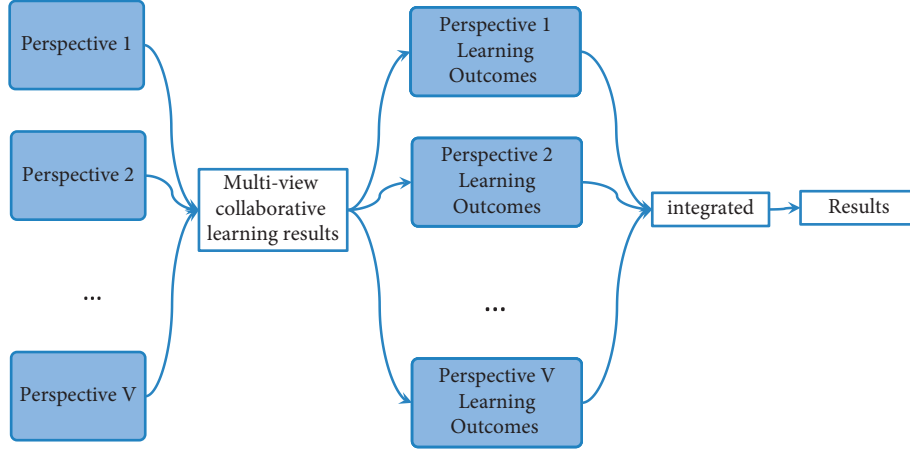


FIGURE 4: Multi-view learning model.

$$\sum_{v=1}^V \sum_{j=1}^C \sum_{i=1}^N w_v^p \delta_{ij} (x_i^{(v)} - z_j^{(v)})^2. \quad (6)$$

$z_j^{(v)} = (z_{j1}^{(v)}, z_{j2}^{(v)}, \dots, z_{jN}^{(v)})$ is the i th cluster center. The number of clusters is C , and the total number of data is N . w_v^p is the weight vector of the v th view, and p is the weight index.

Since the quality of each view is different, the algorithm needs to assign a corresponding weight value to each view to represent the importance of the view. When the data of a certain angle of view is scattered, or when a certain angle of view is greatly interfered by noise, it means that the quality of the angle of view is low, and the weight of the data of this

angle of view is assigned 0. The scalar η is introduced in this paper; when η belongs to the k th cluster, it is equal to 1, otherwise it is equal to 0. The MvK-means algorithm's objective function is as follows:

$$J_H = \sum_{v=1}^V \sum_{j=1}^C \sum_{i=1}^N w_v^p \phi_{ij} (x_j^{(v)} - z_i^{(v)})^2 + \eta \sum_{v=1}^V w_v^p. \quad (7)$$

Using the Lagrange multiplier optimization method to minimize equation (5), the cluster center z_j of the MvK-means algorithm is obtained, and the iterative expression of the weight vector w_v is: Algorithm 1

$$z_j^{(v)} = \frac{\sum_{i=1}^N \phi_{ij} x_i^{(v)}}{\sum_{i=1}^N \phi_{ij}}, \quad (8)$$

$$w_v = \frac{1}{\sum_{v=1}^V \left[\sum_{i=1}^N \phi_{ij} (x_i^{(v)} - z_j^{(v)})^2 + \eta / \sum_{i=1}^N \phi_{ij} (x_i^{(v)} - z_j^{(v)})^2 + \eta \right]}^{1/(p-1)}. \quad (9)$$

4. Experimental Results

4.1. Experimental Data Collection. In order to analyze the performance of the psychometric model used in this paper, the model needs to be trained. Therefore, this study collected the mental health data of a total of 800 students in computer-related majors from freshman to senior year in a university. Mental health data mainly include basic information of students and questionnaire data. The basic information is shown in Table 1. The comparison table for converting each information in Table 1 into numerical value is shown in Table 2. The questionnaire design is shown in Table 3. The two kinds of information shown in Table 1 and Table 3 are taken as two viewing angle data, respectively. A total of 800 samples. Among them, 600

samples are used as training set and 200 samples are used as the test set.

4.2. Experimental Results and Analysis. This section assesses the proposed mental state assessment model's performance on CSMH data. The comparison models used are traditional K-means and fuzzy C-means (FCM). The evaluation indicators use Normalized Mutual Information (NMI) and Rand index (RI). The number of clusters is set to 5, which represent extremely poor, poor, medium, average, and good mental states, respectively. The system generates the initial cluster centers at random. Table 1 shows the attributes of view 1 data, and a total of 8 attributes are used as input variables. Table 3 displays the attributes of view 2 data, and a

Input: Multi-view dataset $D_v = \{D_{v1}, D_{v2}, \dots, D_{vN}\}$. There are a total of v views, and the dataset corresponding to any view is $D_{vk} = \{x_1, x_2, \dots, x_N\}$. The number of clusters $C(2 \leq C \leq N)$, the weight index p . Co-learning parameter η , iteration threshold τ .
Output: partition cluster ϕ , cluster center matrix Z , weight matrix W .
Step1: Randomly select C samples in the dataset as the initial cluster center, the weight value wv , and the initial weight $w_v^p = 1/V$.
Step2: Calculate the cluster center matrix Z by updating equation (8);
Step3: Update the weight value matrix W through equation (9);
Step4: Update the objective function J through equation (7);
Step5: if $\|J^{k+1} - J^k\| < \tau$, then the algorithm ends, jump out of the loop, otherwise return to step2.

ALGORITHM 1: The algorithm flow is as follows.

TABLE 1: Basic information of students.

Gender	Grade	Character	Household income	From rural/urban	Is it an only child?	GPA	Attendance in class
Male	Freshman	Outgoing	Middle	Urban	Yes	Excellent	Good
Female	Junior year	Introverted	Low	Rural	No	Good	Excellent

TABLE 2: Conversion table.

Attributes	Value	Code
Gender	Male	11
	Female	12
Grade	Freshman	21
	Sophomore	22
	Junior year	23
	Senior year	24
Character	Introverted	31
	Outgoing	32
	Inside and outside	33
Household income	Low	41
	Middle	42
	High	43
From rural/urban	Rural	51
	City	52
Is it an only child?	Yes	61
	No	62
GPA	Failed	71
	Pass	72
	Medium	73
	Good	74
	Excellent	75
Attendance in class	Failed	81
	Pass	82
	Medium	83
	Good	84
	Excellent	85

TABLE 3: Questionnaire.

No.	Question	Options
1	Gender	A. male B. female
2	Grade	A. freshman B. sophomore C. junior D. senior
3	Do you feel stressed?	A. very large B. large C. average D. no pressure
4	Source of stress	A. study B. love C. interpersonal communication D. high self-demanding E. do not know F. no pressure
5	How to deal with mental problems	A. talk with friends B. talk with parents C. solve it by yourself D. let it go. E other
6	Whether to attend a mental health education seminar or class?	A. never participated B. listened occasionally C. often participated
7	How to reduce stress?	A. sleeping B. listening to music C. exercising D. talking. E other
8	Can mental health be managed?	A. can B. cannot

TABLE 4: Evaluation results obtained by each model.

Index/model	K-means	FCM	Reference [21]	Proposed
NMI	0.8041	0.8096	0.8351	0.8437
RI	0.7879	0.7923	0.8125	0.8290

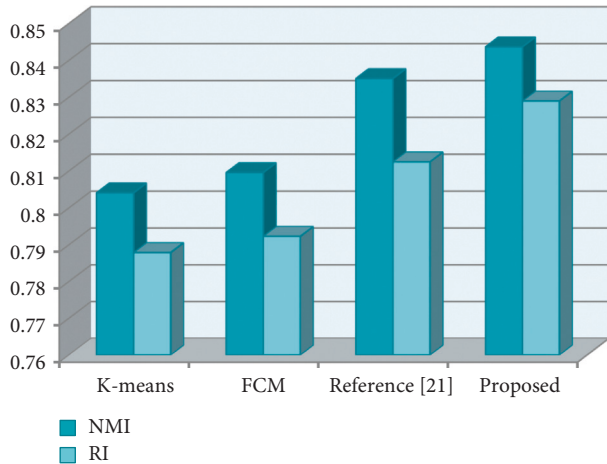


FIGURE 5: Comparison of experimental results.

total of 8 attributes are used as input variables. There is a limit of 100 iterations.

In order to calculate the evaluation index, the real psychological state of each student must be known. As the standard data, this paper uses the student SCL-90 symptom self-rating scale to obtain the psychological state results of 800 college students. The experimental results obtained by each model are shown in Table 4 and Figure 5:

The results show that the model proposed in this paper produces optimal evaluation results. Both reference [21] and the model in this paper belong to the multi-view model, and the experimental results show that the experimental results obtained by the multi-view model are significantly better than the single-view learning models such as K-means and FCM. FCM performs slightly better than K-means due to the fact that FCM introduces membership, which reduces the impact of suffering data on the final result. The proposed model performs better than reference [21] because the proposed model introduces a weighting strategy, and each view data will be given corresponding weights according to its own contribution to the results. The weighting strategy can improve the contribution of high-quality views without losing useful information in low-quality view data.

5. Conclusion

There are many college students, but fewer college teachers and student managers. In this case, the teacher is unable to accurately understand each student's psychological state. College students' psychological problems are becoming more visible as the environment in which they live becomes more complex and pressure increases. The specific manifestation is a daily increase in the number of college students who have poor mental health. The mental health of college

students has a direct impact on the quality of talent training and is linked to the country's future development. As a result, people from all walks of life should pay close attention to how CSMH is managed. Furthermore, relevant mental health management institutions should prioritize CSMH management. Only in this manner will we be able to cultivate high-quality talent capable of adapting to the country's ongoing development in the twenty-first century. As a result, the MvK-means model is proposed in this paper for assessing CSMH. This model is unique in that it employs a multi-view strategy and assigns different weights to different views based on their contribution. The experimental results show that the model has some advantages in assessing college students' mental health. Colleges and universities can implement various management strategies for students with poor mental health based on accurate assessment results. This study's work has the following limitations. First, the data gathered are insufficiently comprehensive. This study primarily collects data from science and engineering students, and the status of CSMH in various major categories will vary. Second, data preprocessing should be improved. To improve data compactness, attributes with high contribution should be selected as much as possible, while attributes with low contribution should be discarded. Third, the model can be improved further to improve the model's evaluation performance. This study will conduct additional research and optimization on the three problems listed above in the future.

Data Availability

The labeled dataset used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Wuxi Vocational College of Science and Technology.

References

- [1] S. Rosenberg, J. Mendoza, H. Tabatabaei-Jafari, and L. S. Carulla, "International experiences of the active period of COVID-19 - mental health care," *Health Policy and Technology*, vol. 9, no. 4, pp. 503–509, 2020.
- [2] A. Gibbons, M. RoseMahoney, and Y. C. Joyce, "The utility of self-perceived health ratings in screening volunteers for mental health research," *Psychiatry Research*, vol. 306, Article ID 114273, 2021.
- [3] C. Zifkin, M. Montreuil, M. È. Beauséjour, S. Picard, C. L. Gendron, and A. Franco, "Carnevale, an exploration of youth and parents' experiences of child mental health service access," *Archives of Psychiatric Nursing*, vol. 35, no. 5, pp. 549–555, 2021.
- [4] E. Ahmed, S. Ahmad, and W. El-Dakhkhni, "Machine learning classification algorithms for inadequate wastewater

- treatment risk mitigation,” *Process Safety and Environmental Protection*, vol. 159, pp. 1224–1235, 2022.
- [5] N. M. Yusof, A. K. Muda, S. F. Pratama, R. Carbo-Dorca, and A. Abraham, “Improved swarm intelligence algorithms with time-varying modified Sigmoid transfer function for Amphetamine-type stimulants drug classification,” *Chemometrics and Intelligent Laboratory Systems*, vol. 226, Article ID 104574, 2022.
- [6] S. M. Ghazali, M. Alizadeh, J. Mazloum, and Y. Baleghi, “Modified binary salp swarm algorithm in EEG signal classification for epilepsy seizure detection,” *Biomedical Signal Processing and Control*, vol. 78, Article ID 103858, 2022.
- [7] J. Waqas and A. Bouchachia, “Iterative ridge regression using the aggregating algorithm,” *Pattern Recognition Letters*, vol. 158, pp. 34–41, 2022.
- [8] U. B. Vyas and V. A. Shah, “Differential evolution based regression algorithm for mathematical representation of electrical parameters in lithium-ion battery model,” *Journal of Energy Storage*, vol. 45, p. 103673, 2022.
- [9] O. Chimere, B. S. Ezuma, M. D. Yining Lu et al., “Camp, A machine learning algorithm outperforms traditional multiple regression to predict risk of unplanned overnight stay following outpatient medial patellofemoral ligament reconstruction,” *Arthroscopy, Sports Medicine, and Rehabilitation*, vol. 4, no. 3, pp. e1103–e1110, 2022.
- [10] G. Caruso, S. A. Gattone, F. Fortuna, and T. Di Battista, “Cluster Analysis for mixed data: an application to credit risk evaluation,” *Socio-Economic Planning Sciences*, vol. 73, Article ID 100850, 2021.
- [11] D. Francisco, “Poor mental health symptoms among Romanian employees. A Two-Step Cluster analysis,” *Procedia - Social and Behavioral Sciences*, vol. 33, pp. 293–297, 2012.
- [12] A. Marinucci, C. Grové, K.-A. Allen, and J. Riebschleger, “Evaluation of a youth mental health literacy and action program: Protocol for a cluster controlled trial,” *Mental Health & Prevention*, vol. 24, Article ID 200216, 2021.
- [13] G. Rubeis, “iHealth The ethics of artificial intelligence and big data in mental healthcare,” *Internet Interventions*, vol. 28, Article ID 100518, 2022.
- [14] P. Mitroshin, Y. Shitova, Y. Shitov, D. Vlasov, and A. Mitroshin, “Big data and data mining technologies application at road transport logistics,” *Transportation Research Procedia*, vol. 61, pp. 462–466, 2022.
- [15] M. Marcia Maja and P. Letaba, “Towards a data-driven technology roadmap for the bank of the future: exploring big data analytics to support technology roadmapping,” *Social Sciences & Humanities Open*, vol. 6, no. 1, Article ID 100270, 2022.
- [16] S. Liu, F. Vahedian, D. Hachen et al., “Heterogeneous network approach to predict individuals’ mental health,” *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 2, pp. 1–26, 2021.
- [17] R. Skaik and D. Inkpen, “Using social media for mental health surveillance: a review,” *ACM Computing Surveys*, vol. 53, no. 6, pp. 1–31, 2020.
- [18] A. Thieme, D. Belgrave, and G. Doherty, “Machine learning in mental health: a systematic review of the HCI literature to support the development of effective and implementable ml systems,” *ACM Transactions on Computer-Human Interaction*, vol. 27, no. 5, pp. 1–53, 2020.
- [19] K. Kirkpatrick, “Artificial intelligence and mental health,” *Communications of the ACM*, vol. 65, no. 5, pp. 32–34, 2022.
- [20] A. Thieme, D. Belgrave, A. Sano, and G. Doherty, “Machine learning applications: reflections on mental health assessment and ethics,” *Interactions*, vol. 27, no. 2, pp. 6–7, 2020.
- [21] A. Gupta, S. Datta, and S. Das, “Fast automatic estimation of the number of clusters from the minimum inter-center distance for k-means clustering,” *Pattern Recognition Letters*, vol. 116, pp. 72–79, 2018.
- [22] R. Xu, “The relationship between psychological quality education and mental health level of college students by educational psychology,” *Frontiers in Psychology*, vol. 13, Article ID 892143, 2022.

Research Article

A Comparative Optimization Model of Japanese Literature Characteristics for Cognitive Retrieval of Cross-Language Information

Tao Feng 

Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China

Correspondence should be addressed to Tao Feng; 2016120064@jou.edu.cn

Received 21 July 2022; Revised 24 August 2022; Accepted 25 August 2022; Published 14 September 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Tao 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.

Cross-language information cognitive retrieval has grown in importance as a study area due to the multilingual character of Internet resources and the diversifying languages spoken by users. This paper analyses and illustrates the key characteristics of Japanese literature from three aspects: ideological structure, structural form, and emotional expression, and makes a straightforward comparison with Chinese literature using a number of well-known Japanese literary masterpieces as examples. An important aspect of this literature is that it is full of lingering feelings, leaving room for the readers to imagine. The lingering charm is endless, obscure, and meaningful, which is the traditional style of Japanese literature. It also focuses on the depoliticization of Japanese literature, that is, the relative separation between literature and politics. The subjectivity of Japanese writing, which goes beyond just conveying feelings but involves less objective description, is related to this. The process of cross-language information retrieval in foreign nations is primarily separated into three stages, according to the changes in research objects. Currently, adding a language conversion mechanism to a monolingual information retrieval system serves as the primary language information retrieval solution. Currently, nearly 40% of the global popularity of literature comes from Japan. In this article, the background and concept of cross-language information retrieval are introduced, and its types, system models, and several key cross-language information retrieval methods are explained, and some solutions to the factors influencing the cross-language information retrieval effect are suggested.

1. Introduction

The traditional information retrieval system is mainly aimed at searching documents in a single language, and when users query, they use the most familiar language to construct retrieval questions. The information age has produced a lot of digital information, among which, text information is the most basic and commonly used form. In order to find their own needs in the vast amount of text information, people urgently need an efficient retrieval tool. How to efficiently store and query unstructured data such as text is a problem worth studying. Among them, full-text retrieval technology and full-text database technology have become the research focus of scholars at home and abroad [1]. Japanese literature has a history of more than 1,300 years since its first work, Ancient Stories. From the oral literature such as ballads that

appeared before The Chronicles of the Ancient, the history is even longer. In the past 1300 years, Japanese literature has experienced both a downturn in which excellent works are scarce and the literary world is declining, and a boom in which excellent works are frequent and the literary world is at its peak. The watershed in the history of Japanese literature is about the 5th century AD. Ecological destruction did not start in the 20th century, but its aggravation started in the 20th century. Since the second half of the 20th century, with the rapid development of social economy, environmental pollution and ecological damage have become more and more serious worldwide. Soil erosion, landslides, land desertification, the increase of saline-alkali land, global warming, seawater pollution, groundwater depletion, waste flooding, ozone hole, etc., have made the environment deteriorating. The unscientific and unlimited economic

development activities of human beings have become the main reason for the continuous deterioration of the ecological environment, and the living environment of human beings has fallen into a dangerous situation. With the globalization of informatization, the information resources provided by the Internet are no longer concentrated in a few languages such as English. At the same time, the proportion of non-English-speaking Internet users is also increasing rapidly. According to the forecast, by 2005, non-English-speaking Internet users will increase to 68% of the total number of users, among which Chinese-speaking users have the fastest growth rate, accounting for about 21% of the total number of users, while other non-English-speaking languages have also increased to varying degrees. For most users who are not proficient in foreign languages, it is difficult to skillfully use foreign language queries, while using native language query conditions to retrieve relevant foreign language information. It is an important feature of national literature [2].

Japanese literary works will reflect the social life, customs, cultural traditions, psychological state, and language features of the Japanese nation, and have the characteristics of Japanese archipelago. The interpretation of Japanese literary works is helpful to understand the living conditions and customs of the Japanese society. The cross-language information retrieval system's objective is to retrieve documents pertinent to the query criteria from a document set other than the language of the user's query criteria. A Lotn's cross-language retrieval study on English-German and English-French-German from the 1970s is where the earliest cross-language information retrieval can be found. The topic of cross-language information retrieval has made significant advancements after decades of relentless investigation by scholars in related domains. In an effort to serve as a resource for other researchers who might be interested, this work conducts a thorough analysis of the cross-language information retrieval technology [3].

The corresponding research is completed in this essay. Models and formulas are developed to explore cross-language information retrieval in research. A related data map is created to assess Japanese literature in accordance with the qualities being studied. The main contributions of this paper are as follows:

- (1) In the article, the argument method is used to further explain it.
- (2) The contribution of this paper is to use the multi-mode teaching method to further understand college English and other related contents.
- (3) In the article, we use the method of multiple evidences to analyze and understand it.

The rest of this paper is arranged: the second part introduces the related work to make corresponding research and analysis. The third part makes corresponding research and analysis on cross-language information exploration. The fourth part studies and analyzes the characteristics of Japanese literature and establishes the corresponding data map.

2. Related Work

With the rapid expansion of the number and scope of Internet users, the languages they have mastered are beginning to show diversification. Because of the diversity of the languages of network resources and the differences of languages mastered by network users, it inevitably brings language barriers to people's information retrieval through the network. For example, more than 90% of the information in the network is in English, while only about 40% of the network users use English, which brings great inconvenience to users in non-English-speaking countries. With the introduction of Chinese characters, Japanese literature began.

Enter the prosperous period of written records. With the large-scale spread of Chinese books in Japan, Chinese Buddhism and Confucianism are also influencing the development of Japanese literature in a subtle way.

In the research, Oard et al. believe that the selected intermediate language should be a language that is easy to be automatically processed by computers, such as English [4]. This method is often used in cross-language information retrieval of more than two languages or cross-language information retrieval without direct corresponding conversion between two languages (for example, German and Italian). The process of this method is to convert the questioning language into the intermediate language, and then convert the intermediate language into the target language, or to convert both the questioning language and the target language into an intermediate language. Nie believes that "natural literature" is biased towards realistic works, and its "nonfiction limitation in genre excludes works that reflect ecological crisis in novels, dramas, poems, and other fields, which cannot cover the current situation of literary creation in the new period. At the same time, there is also the danger of leading this literary trend to realistic reportage, which will lead to the dissolution of the artistic depth of the works" [5]. Ning and Lin think that "natural writing" is too narrow for writing objects. The term is too broad in terms of ideology and subject matter [6]. As long as it is written about nature, it includes nonecological or even antiecological works. Moreover, "natural writing" refers to all writing that takes nature as its object. Broadly speaking, it does not only refer to literary works but also popular science books, reference books, and works on philosophy, natural history, politics, religious studies, cultural criticism, etc., which greatly exceeds the scope of literary studies. Chaware and Rao think that it is impossible for socialist countries to produce public hazards. However, it was only after the comparison that "suddenly I saw the seriousness of China's environmental problems. The pollution degree of China's cities and rivers is no less than that of western countries, but the degree of natural ecological destruction is far above that of western countries" [7]. Zhou et al., lawless believe that literature, as an art, is best not to be close to reality. Only when it is divorced from reality can it have an artistic interest. Moreover, the general tendency of Japanese art is to look for the unfamiliarity, elegance, and symbolic beauty in places divorced from reality [8]. Ma put forward a cross-language information retrieval technology [9]. Sion uses

artificially coded translation knowledge and crosses language barriers through questioning translation strategies. Andreea put forward some unavoidable ecological and environmental problems such as air pollution, pesticides, and fertilizers, pointed out the seriousness of public hazards, and sternly condemned the producers of public hazards. With a unique color of ecological philosophy, the works show a new awakening and pursuit full of ethical judgment and philosophical thinking [10]. Sherif and Ann put forward some unavoidable ecological and environmental problems, such as air pollution, pesticides, and fertilizers, pointed out the seriousness of public hazards, and sternly condemned the producers of public hazards. With a unique color of ecological philosophy, the works show a new awakening and pursuit full of ethical judgment and philosophical thinking [11].

3. Cross-Language Information Retrieval

3.1. Cross-Language Information Retrieval Development and Research. Online information resources are expanding daily, thanks to the growth of the Internet. They are no longer just limited to few languages, such as English and Chinese, and the number of resources available in other languages is also growing significantly. It is more and more common for users to query a multilingual text collection. However, most users cannot skillfully use foreign languages to clearly describe queries and correctly express their needs. If you can use your mother tongue to construct the search questions, retrieve useful information from multilingual information sets, and then browse with the help of translation tools, and the recall rate of users can be greatly improved. Documents in at least two different languages are involved in cross-language information retrieval. It cannot be separated from the “translation” process because it is important to first ascertain the language, form, and coding scheme of the documents before automatically analysing and indexing them and then realising retrieval matching. The major issue with cross-language information retrieval technology is “translation,” and there are four approaches to approach it. In the late 1990s, as the Internet expanded quickly, cross-language information retrieval research truly picked up steam and produced results. The globe saw the publication of numerous relevant studies, and several experimental linguistic information retrieval methods appeared one after another. Cross-language information retrieval research has advanced significantly with the quick development of the Internet and computer technology. In recent years, many related papers have been published and some conferences on cross-language information retrieval technology have been held at home and abroad [12, 13]. Traditional information retrieval systems are evaluated in a standardized laboratory environment to compare the retrieval performance of retrieval systems or retrieval technologies. However, the early test document sets are usually small, and there is a big gap between them and the real retrieval environment. Therefore, the retrieval system based on such test sets cannot achieve good performance in practical application. Cross-language information retrieval

means that the user constructs a retrieval question in a certain language he has mastered, and the computer automatically searches the information in other different languages (including text, voice, and images) according to the user’s retrieval requirements, and the retrieved results can even be translated into the language specified by the user. Multilingualism is one of the characteristics of the Internet world. According to the statistics of ETHNO-LOGUE in 1995, there are as many as 5,703 languages in the world. According to the research of other scholars, there are currently 160 languages of information on the Internet. Since the beginning of the search guide, adding support for multiple languages has been one of the secrets to triumphing in the harsh competition. The discipline of multilanguages has conducted much research on cross-language information retrieval systems over the past ten years. Experts from all around the world have started to focus their attention on the study of the Chinese cross-language information retrieval system in recent years. Figures 1 and 2 illustrate how the research’s associated model diagrams are set up to analyse and investigate them.

MulEnex system, developed in 1997 by the Artificial Intelligence research center of German language and technology research office, is the first cross-language network information retrieval system in the world, which successfully uses cross-language automatic translation technology, so that people can use their own language to effectively obtain information of other languages on the Internet. The system model of cross-language information retrieval mainly includes several functions and modules, such as document preprocessing, constructing retrieval questions, matching, selecting, checking, and transmitting [14, 15]. Literature preprocessing, that is, firstly identifying the language and various forms of literature. According to its research, the corresponding data tables are established to analyze and understand it, such as Tables 1 and 2.

The questioning matching, homologous matching, document translation, and interlanguage conversion technology are the four ways that the cross-language information retrieval system uses to match users’ queries and index information. The selection and check functions of the information retrieval system’s user interface will provide pertinent feedback to the user’s retrieval needs in order to increase recall and precision. The search is finished when the user receives the search results from the system. Cross-language information retrieval technology has received attention and has been applied to international online retrieval in an effort to raise the standard of retrieval on the global web and enable users to access and interpret foreign literature information sources.

3.2. Analysis and Research of Cross-Language Information Retrieval Methods. Documents in at least two different languages are involved in cross-language information retrieval. It cannot be separated from the “translation” process because it is important to first ascertain the language, form, and coding scheme of the documents before automatically analysing and indexing them and then realising retrieval

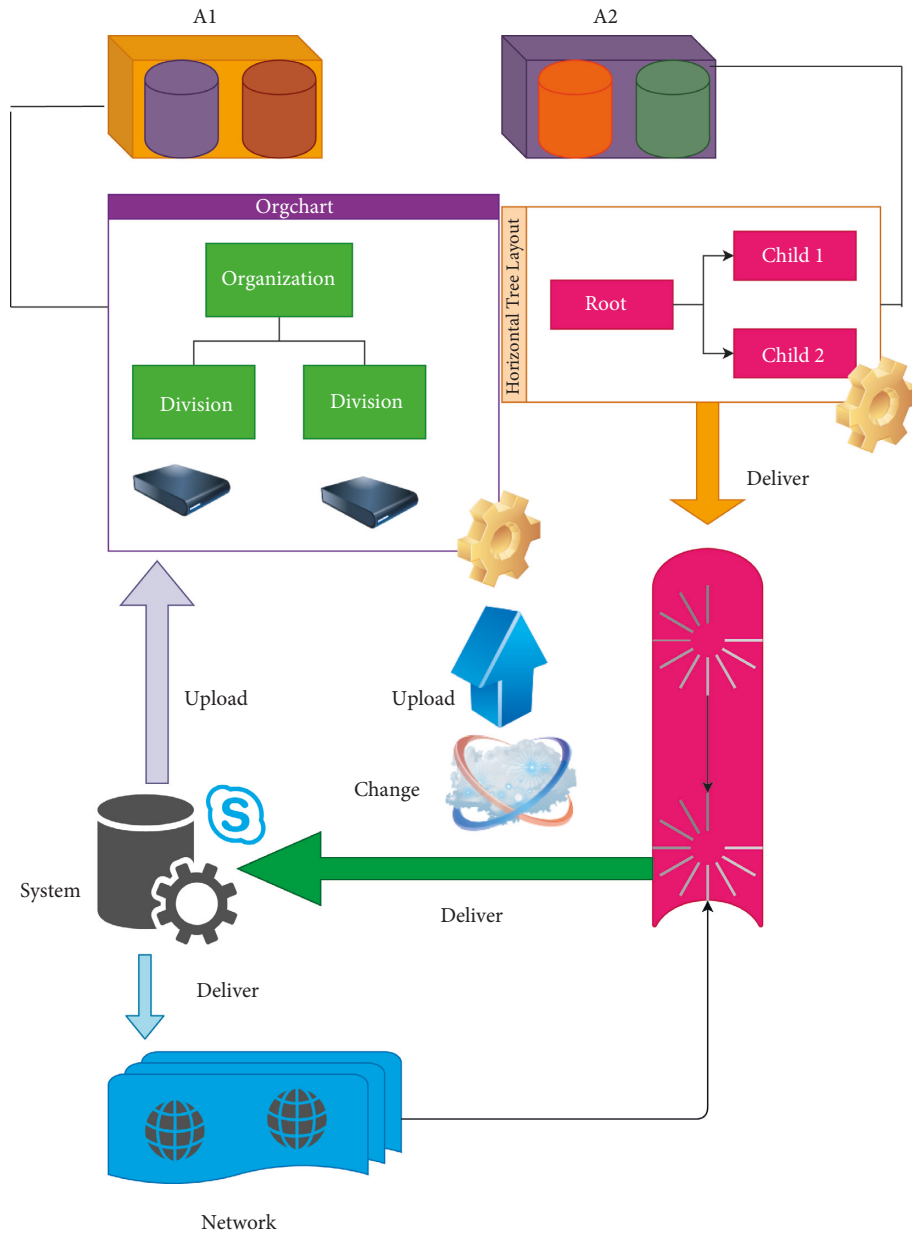


FIGURE 1: Cross-language information model diagram.

matching. The major issue with cross-language information retrieval technology is “translation,” and there are four approaches to approach it. Without any translation, homologous matching evaluates a word’s meaning in one language based on how closely two words are spelled or pronounced [16, 17]. The selection of literature is based on the concept marks being matched in the controlled vocabulary cross-language information retrieval system, which typically uses a multilingual thesaurus to determine the correspondence between the index words of each language supported by the system and a set of universal language-independent concept marks. This method translates the questions input by the user into other languages supported by the system, and then carries out single-language retrieval. Questioning translation strategy is the most commonly used

strategy at present. Most researchers use this strategy to conduct CLIR research and experiments, and put forward a variety of specific methods to translate questioning. Generally speaking, there are three different technical routes to solve the language barrier between query conditions and query document sets, which can translate query conditions into the same language as query document sets or query document sets into the same language as query conditions. With the rapid growth of network information resources, the cross-language information retrieval of controlled vocabulary cannot meet the information needs of students more and more; so, the research of cross-language information retrieval of free text appears. These words are uncontrolled words, which can be directly used by users during retrieval without manual indexing. Because the freedom of

In language information retrieval, translation between two languages is the best way to solve the language barrier. However, all translation methods are inseparable from machine translation, bilingual dictionaries, and corpora. Controlled vocabulary retrieval means that the document collection is searched manually according to the preselected vocabulary, and the user also selects the vocabulary from the same controlled vocabulary to construct the query conditions, and then searches the documents. The information required for automatic analysis, translation, and production of natural language is contained in the machine translation dictionary-based method. The machine translation system can frequently give superior translation results, because this method makes advantage of the syntactic and semantic elements of the context to enhance translation quality, particularly for retrieval inquiries composed of lengthy and complete phrases or paragraphs. The main issue with cross-language information retrieval is translation uncertainty, which has a significant effect on retrieval effectiveness. Foreign academics pay great attention to it as a research hub, and the primary language resources used there are dictionaries, thesaurus, ontologies, corpora, etc. It is not necessary for CLIR to eliminate all translational ambiguities. If the system permits numerous translations for a word, then in conventional information retrieval, the ranking of retrieval results is influenced by the frequency with which the term appears in the inquiry formula. At the beginning of 1990s, a revolutionary thought emerged in the field of knowledge engineering in the method of building knowledge base, that is, the idea of building ontology and ontology engineering. The quality of translation resources has an important influence on the performance of cross-language information retrieval. Therefore, in the research of cross-language information retrieval, foreign scholars have made in-depth studies on the construction of translation resources and their comparison with each other. In ontology-based cross-language information retrieval, the main difference between this semantic level implementation and the traditional CLIR method is that in the process of cross-language conversion of queries, the dictionary or other methods are not blindly used for character level processing, but the keywords of queries are preliminarily distinguished, and the implied semantics of the contents in the ontology library can be identified and retained in the conversion process. Cross-language information retrieval involves two basic concepts: query language and search language. The query language is the language of the user's query request, and the retrieval language is the language of the retrieval target. How to build a bridge between the two is the core and key issue in the research of cross-language information retrieval technology. In the process of information retrieval, instead of using character matching or related optimization strategies to find the target, the retrieval object is semantically processed, and the semantic correlation between the potential target object in the semantic paragraph and the query request is analyzed, so as to decide whether to return it as a result. In the research, the corresponding algorithm formulas are established to analyze it, such as formulas (6)–(10).

$$l_{-T}^{+L^1} \frac{3}{\gamma} \mathcal{N}_{-}^{-1^3} \left(\sum_u 1 = 11 \right)^{-2^1}, \quad (6)$$

$$\mathbb{P} + |L - 41| = +1^2 x_d' \forall r_1 t, \quad (7)$$

$$1 - \bar{l}_1 * (+1) - 1, -\Gamma_{|1-|}^{ch'} \text{fsd}, \quad (8)$$

$$\lambda \cdot 1 - th^1 x \|\pi\|^{+1}, 211, \quad (9)$$

$$\text{Farcct} g \frac{1}{l^1} t \mathcal{V} |11 - r_{-}''|. \quad (10)$$

Currently, the TREC standard document collection is used by many information retrieval systems to evaluate their performance. Generally speaking, a cross-language information retrieval system's efficiency is 40% to 70% lower than a single-language system's efficiency. In general information retrieval, the problem of word span, or how to retrieve publications that do not contain the key terms in the questions but are truly related, is recognised as the difficulty of cross-language information retrieval. Since this primarily involves language interpretation, it is also a matter of meaning or conceptual organisation. Theoretical and technical exploration of this problem is important and meaningful [18, 19]. Knowledge-based cross-language information retrieval methods have many difficulties in establishing large dictionaries and complex multilingual thesauri, thus affecting the recall and precision of information retrieval. Through automatic analysis of large literature collections, the information needed to construct automatic translation technology is extracted. Cross-language information retrieval corpora mainly include parallel corpora, comparable corpora, and misaligned corpora. Parallel corpus includes three alignment methods, which can be aligned at the document level, at the sentence level, or even at the word level. Parallel corpora contain a lot of translation knowledge. There are similarities between language information retrieval and distributed information retrieval in many aspects. The common research of distributed information retrieval includes information representation, information selection, and result combination, but there are few researches on these aspects, especially the first two, which are basically in a blank state, and there is still much work to be done. Thesaurus can also manage domain knowledge. A key feature of each multilingual thesaurus is the cross-language synonym specification. Because the cross-language synonym specification is introduced into the thesaurus used for cross-language information retrieval, the keywords in different languages can be compared with each other. The complicated thesaurus also contains conceptual structure information, antonyms, and related words among words or concepts [20].

4. Japanese Literature Research

4.1. Research and Structure of Japanese Literary Thoughts. Japanese literary works are generally characterized by depoliticization. The earliest Japanese literary works with strong political color are the Japanese Book of Records and Ancient Stories, which involve the description of the origin of gods and the emperor's family. When it comes to Japanese literature, we cannot help but talk about Japanese language. However, in the past, Japanese people often overlooked this point when writing the history of Japanese literature. Japanese is a very distinctive language, and writing with it is bound to have different characteristics from those of European, American and Chinese languages. For example, the difference between prose and verse is distinguished by whether it rhymes or not in Europe, America, and China. It is not only the traditional Japanese literature that is depoliticized but also its modern literature. The Meiji Restoration Movement in Japan has played a great role in promoting the development of modern literature. However, modern literature in Japan is still separated from politics. As rhymes, English, German, and Russian all have a special tune. Only by consciously forming a rhythm according to a certain tune can they be regarded as rhymes. However, Japanese pronunciation has no obvious ups and downs, and it cannot be used to distinguish verse from prose. Japanese is distinguished by the number of syllables, and each line has a certain number of syllables, or seven or five, which is the so-called "seven-five-tone." This is the most typical form of Japanese poetry. Before modern times, Japanese literature was mainly presented in the form of short songs with simple structure, concise narration, and short form. This form has been relatively developed from ancient times to modern times, and it still lasts forever. Japan's diary literature and essay literature, very pursuit of beautiful style. The representative and traditional Japanese literature feature subjectivity. Comparing Japanese literature with European literature and Chinese literature, Homer epics in ancient Greece are mostly objective narratives; reading Chinese novels is subjective, but there are still many objective elements. The most typical Japanese literary work, *The Story of the Heian Dynasty*, contains few events; so, most of them are carried out in the minds of the characters. In the research, corresponding data charts are set up to analyze and study them, as shown in Figures 3–5.

Japanese writers can express the emotional world of the singers and the monks to the fullest with just two or three short songs and haiku. This kind of artistic expression technique that pursues conciseness in a simple structure is really amazing, and this "doll interest" that condenses life and nature is a special expression of lyricism in Japanese literature. It is true, but the vocabulary in this area is poor. In ancient times, when Japanese people wrote objective things, they always loved Chinese. Just as English people used Latin for important things in the Middle Ages, Japanese people used Chinese for important things in the Middle Ages. This form itself is not unique to Japanese literature, but it is a special phenomenon in modern Japanese literature that it has become the mainstream of literature. In other countries,

novels are mostly fictions, not to record the author's own experience, but to write about other things, his friends, created characters, or historical figures. This tendency is also reflected in the works of Ihara Nishiko in the Middle Ages, such as *Lust Generation Men*, *Trouble in the World*, Shiga Naoya's *Dark Night* in modern times, and Kawabata Yasunari's *Snow Country*. In short, these works can come to an end at any time, or they can be finished everywhere. This feature is also very noticeable in Murasaki Shikibu's *Tales of Genji* and other Japanese prose stories.

4.2. Research on the Expression and Characteristics of Japanese Literature. The natural topic is one of the most revered in Japanese literature. People have had a great reverence for nature since the dawn of time, and our country is especially well-known for it. The geographic setting of Japan is inextricably linked to this. Japan is a small island nation in the Pacific Ocean that is well known for its stunning natural landscape, different seasons, and agreeable weather. An important question is how modern Japanese authors understand their own literary heritage. Some authors place a high value on literature from antiquity and the classical period. They painstakingly study these works before using them as the foundation for their own original works. Kawabata Yasunari, who won the Nobel Prize for Literature, is such a writer, and some writers have partially inherited the classical tradition, especially the folklore. For example, Shunji Muxia, a famous playwright, tried to use classical themes and folk legends to create new plays. From the perspective of "flowers" alone, the reason why Japanese people still cannot stop liking cherry blossoms is that cherry blossoms can make them feel that fickle feeling. Contrary to the attitude of westerners who pursue "eternal things" and get beauty from them, most Japanese people feel beauty in moving and changing things, and this tendency is deeply rooted. According to its research, the corresponding data graphs are established to analyze and explain it, as shown in Figures 6 and 7.

All in all, Japan's literary character is very delicate and subtle. Japanese literati mostly express their ordinary life, pursue delicate emotional experience, and show their calm thinking about life and society. Japanese writers have the ability to find the artistic conception of "beauty" in "sorrow." The novel "Dark Painting" is difficult to read. It is always a disadvantage to confuse readers regardless of how it destroys the Japanese tradition. Although it has a good side, as Mr. Noguchi once wrote, in modern Japanese novels, most of them do not have facial descriptions of characters. However, European literary works must have facial descriptions of characters, otherwise people will not know what kind of characters are written. Ye Jian's works describe the characters' faces in detail, and his literature has new ideas. This kind of knowledge is also connected with the principal idea that "the mind is shaken, so literature sprouts." This makes literature show a simple, delicate, and sensitive attitude according to the starting point, or it has no purpose. Sensitive and aimless are the personalities related to the characteristics of Japanese literature.

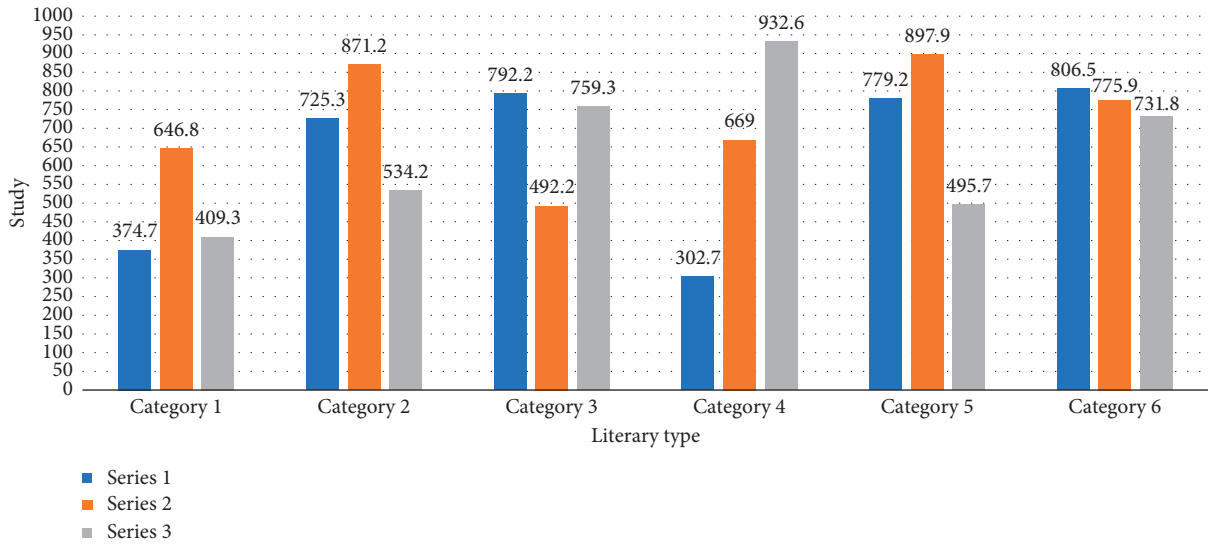


FIGURE 3: Data map of Japanese literature types (1).

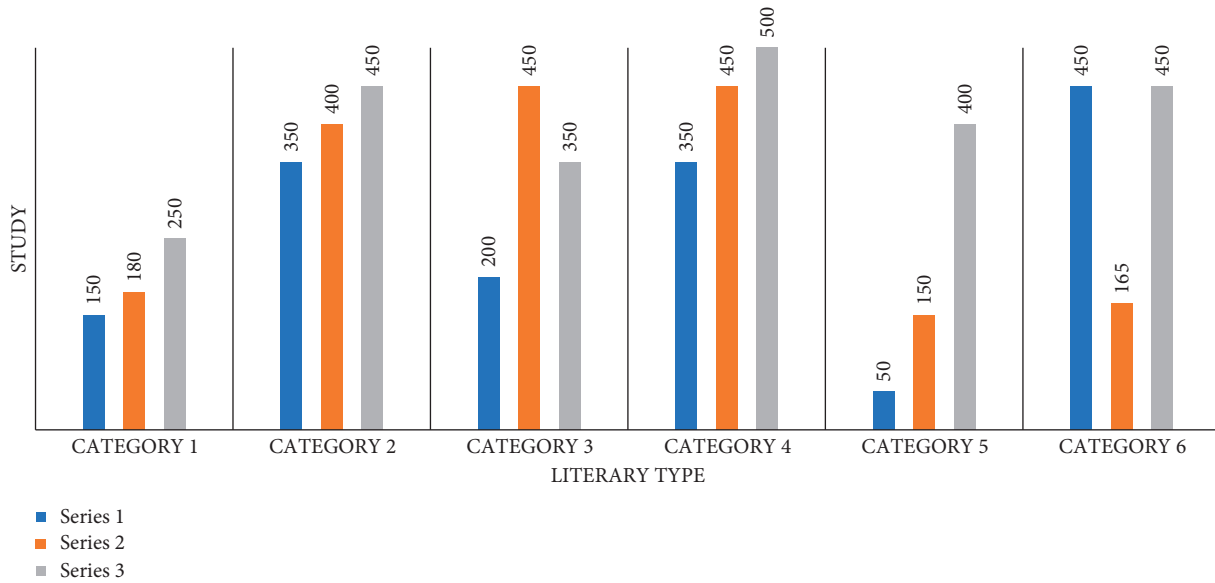


FIGURE 4: Data map of Japanese literature types (2).

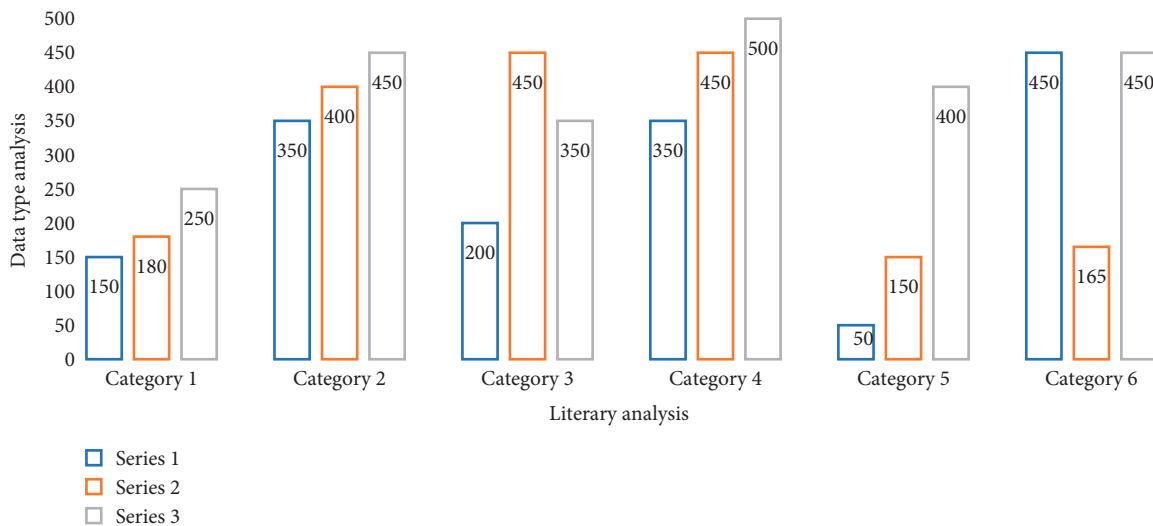


FIGURE 5: Data analysis of literature category gap.

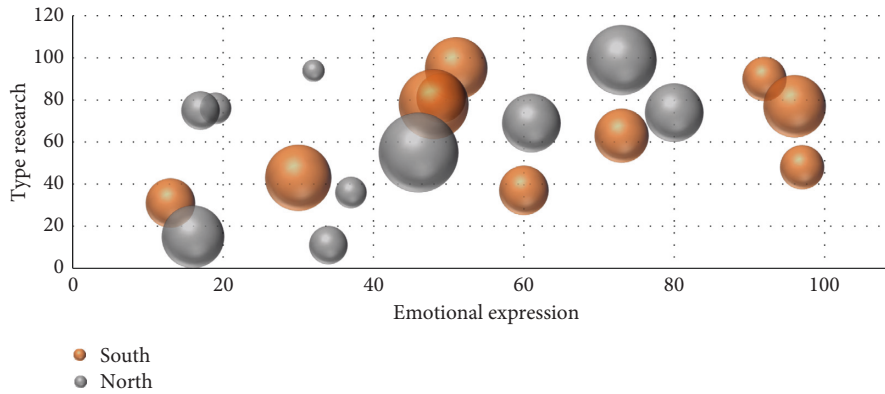


FIGURE 6: Data analysis of literary emotional expression research (1).

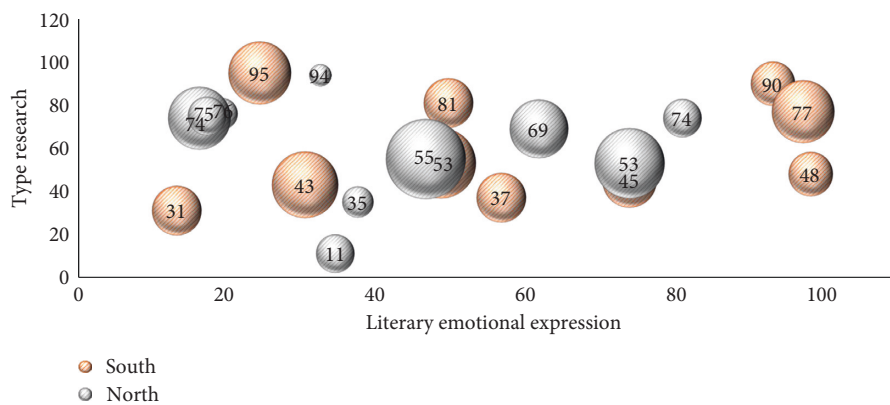


FIGURE 7: Data analysis of literary emotional expression research (2).

5. Conclusion

Cross-language information retrieval has advanced significantly in recent years, both in terms of research techniques and test set development. The result of combining several procedures is superior to that of using only one. Therefore, a clear trend in this area is that more and more academics are starting to think about integrating some of the aforementioned techniques discussed in this work in order to further boost query translation accuracy. Cross-language information retrieval should not be restricted to just document retrieval; it can also include cross-language interactive retrieval, language question-and-answer systems, language new subject discovery, and eye tracking. CLEF has currently conducted fruitful research in connected cities. Technically, the creation of a sizable bilingual or multilingual text corpus, the creation of related tools to investigate the correspondence, and concurrent neutral expression of words in several languages, or the theoretical investigation to create a more general conceptual architecture, such as WordNet, and the combination of these investigations, may be the fundamental solution to the information retrieval problem. The topic of cross-language information retrieval has gained popularity worldwide. Even though the research in this area has advanced significantly, there are still not enough cross-language retrieval systems and related tools available in the

actual corpus environment, and there is still much work to be done before the current technology is truly useful. This continues to be a significant challenge for the field's researchers. Looking ahead, the creation of cross-language information retrieval technology will integrate significant library resources and network information sources. As a result, the way that people interact with information will change significantly, and information will be expressed in a richer and more user-friendly manner. Japanese literature shares these unique distinctive styles and traits with literature from other nations. Since the Yamato era, it has evolved over more than 1000 years, incorporating the best aspects of other civilizations while also adhering to its own features to eventually create this distinctive style. The most thorough and accurate mapping of culture and society can be found in literature, which is a byproduct of the most successful historical stage of culture.

Data Availability

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

Conflicts of Interest

The author does not have any possible conflicts of interest.

Acknowledgments

This research was supported by: Research project of Foreign Language Education Reform in Vocational Colleges of Ministry of Education in 2021 (Foreign Language Teaching Guidance Committee [2021] No. 17) “A Practical Study on The Construction of “Thought and Innovation Integration” Education Mode for Higher Vocational Foreign Language Majors” (WYJZW-2021-2050); Research and Practice project of University-level Education and Teaching Reform of Guangdong Vocational College of Science and Technology in 2021 “Research on reform and Practice of University-Enterprise Cooperation To Construct “Integration of Thought and Innovation” Education Mode in Higher Vocational Colleges” (JG202128).

References

- [1] Z. Ye, J. X. Huang, B. He, and H. Lin, “Mining a multilingual association dictionary from Wikipedia for cross-language information retrieval,” *Journal of the American Society for Information Science and Technology*, vol. 63, no. 12, pp. 2474–2487, 2012.
- [2] T. Talvensaari, J. Laurikkala, K. Jarvelin, M. Juhola, and H. Keskustalo, “Creating and exploiting a comparable corpus in cross-language information retrieval,” *ACM Transactions on Information Systems*, vol. 25, no. 1, p. 4, 2007.
- [3] Y. Wang, T. Wang, J. Wang, X. Zhou, M. Gao, and R. Liu, “Military chain: construction of domain knowledge graph of kill chain based on natural language model,” *Mobile Information Systems*, vol. 2022, pp. 1–11, Article ID 7097385, 2022.
- [4] D. W. Oard, D. He, and J. Wang, “User-assisted query translation for interactive cross-language information retrieval,” *Information Processing & Management*, vol. 44, no. 1, pp. 181–211, 2008.
- [5] J. Y. Nie, “cross-language information retrieval,” *Synthesis Lectures on Human Language Technologies*, vol. 3, no. 1, pp. 1–125, 2010.
- [6] J. Ning and H. Lin, “cross-language information retrieval based on improved latent semantic indexing,” *Journal of Chinese Information Processing*, vol. 6, no. 5, p. 54, 2010.
- [7] S. M. Chaware and S. Rao, “Ontology approach for cross-language information retrieval,” *International Journal of Computer Technology & Applications*, vol. 02, no. 2, pp. 275–298, 2011.
- [8] D. Zhou, S. Lawless, X. Wu, W. Zhao, and J. Liu, “A study of user profile representation for personalized cross-language information retrieval,” *Aslib Journal of Information Management*, vol. 68, no. 4, pp. 448–477, 2016.
- [9] J. Ma, “A methodology for research into the stylistic features of Japanese ancient literature and buddhist scriptures,” *Journal of Japanese Language Study and Research*, vol. 534, no. 64, p. 67, 2019.
- [10] Sion and Andreea, “Some aspects regarding literary translations from Japanese,” *Studies of Science & Culture*, vol. 2009, no. 16, p. 7, 2009.
- [11] A. Sherif, “The demimonde in Japanese literature: sexuality and the literary karyūkai. By cynthia gralla. Amherst, N.Y.: cambria press, 2010. 300 pp. \$114.99 (cloth),” *Journal of Asian Studies*, vol. 72, no. 2, pp. 472–473, 2013.
- [12] M. Galik, “ON the LYRIC(AL)NESS in the SINO-Japanese interliterary process: musings after reading KOKINSHU prefaces,” *Asian & African Studies*, vol. 43, no. 64, p. 4, 2007.
- [13] J. Nangu and A. Minamitani, “Sex and literature as a medium: sexual representations in postwar Japanese literature,” *Research Collection of Social Information Department of Gunma University*, vol. 19, no. 4, pp. 75–94, 2012.
- [14] T. Matsushita, “Exploring the tiers of Japanese vocabulary,” *Academic, literary and beyond*, vol. 3, no. 4, p. 34, 2011.
- [15] Y. Uchida, “Japanese literary studies in taiwan::the present situation and problems,” *Modern Japanese Literary Studies*, vol. 92, no. 23, p. 4, 2015.
- [16] L. You, H. Jiang, J. Hu et al., “GPU-accelerated Faster Mean Shift with euclidean distance metrics,” in *Proceedings of the 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 211–216, IEEE, Los Alamitos, CA, USA, July 2022.
- [17] T. Namigata, “Fusion approaches in Japanese literary studies,” *The Korean Journal of Japanology*, vol. 100, no. 64, pp. 25–33, 2014.
- [18] R. N. Mckinnon, “Japanese literature of the showa period. A Guide to Japanese reference and research materials,” in *The University of Michigan Center for Japanese Studies, Bibliographical Series No. 8*, K. YamagiwaJoseph, Ed., vol. 20, no. 1, pp. 212–113, University of Michigan Press, Ann Arbor, Michigan, United States, 1959.
- [19] C. Yamin, “Research on the precision Japanese literary translation mode from the perspectives of cultural equivalence and metaphor,” *International technical management*, vol. 2016, no. 2, p. 3, 2016.
- [20] H. Osanai and H. Okada, “Construction of the literary response questionnaire for Japanese (LRQ-J),” *Japanese Journal of Psychology*, vol. 82, no. 2, pp. 167–174, 2011.

Research Article

Distributed Scheduling Strategy of Virtual Power Plant Using the Particle Swarm Optimization Neural Network under Blockchain Background

Changchang Lu  and Sijie Chen

College of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Changchang Lu; sjlccls@sjtu.edu.cn

Received 21 July 2022; Revised 8 August 2022; Accepted 9 August 2022; Published 13 September 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Changchang Lu and Sijie 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.

Large-scale and widely dispersed distributed energy resource (DER) can be gathered by a virtual power plant (VPP) in a given area, and its parameters can be combined into a single external operation profile. Each distributed energy source in the VPP has a complete backup of the critical information for the entire network because it is a node of blockchain. The distribution network can be accessed by DER freely and adaptable under the scientific management of the VPP, and it can offer the system high-reliability, high-quality, and high-security power services. An energy blockchain network model based on particle swarm optimization (PSO) to optimise the neural network is proposed in this paper as a solution to the issues with the current VPP models. This will enable distributed dispatching of the VPP and reasonable load distribution among units. According to the simulation results, this algorithm's error is minimal and its accuracy can reach 94.98 percent. This model can more accurately capture demand-side real-time information, which benefits VPP's stable scheduling with a welcoming environment and transparent information. It also enhances the system's data security and storage security. This system can successfully address the issues of subject-to-subject mistrust and high information interaction costs in the VPP.

1. Introduction

The development of all spheres of life must rely on the robust support of the power system given the accelerating advancement of science and technology. Countries all over the world are severely constrained in their ability to develop further by global energy shortages and environmental degradation. The best way to deal with this issue is to actively develop and use DER, which has the properties of low-carbon cleaning, recycling, diversification, and dispersion [1]. Power plants, substations, transmission and distribution networks, and users make up the power system, a sizable system of unified dispatching and operation. Power system automation's primary objective is to guarantee the supply of high-quality power [2]. The operation and control of the entire power system depend heavily on the economic dispatching of the power system. It falls under the power system

planning and the operation dispatching category. It is a typical optimization problem, where the objective is to efficiently use available resources while minimising system operation costs under the constraints of ensuring system load constraints and safe and stable operation [3]. In order to aggregate distributed generation resources and establish virtual power resource transactions, the energy Internet's VPP is a crucial branch. Large-scale and widely dispersed DER can be gathered by the VPP, and its parameters can be integrated into a single external operation profile [4]. DER can freely and flexibly connect to the distribution network while being managed scientifically by the VPP, supplying the system with high-reliability, high-quality, and high-safety power services.

Each component of the VPP is connected to the control centre, and data are transmitted in both directions through the smart grid. Power flow at the machine end, load at the

load end, and the energy storage system are uniformly dispatched to achieve the goals of lowering power generation loss and peak grid load, optimising resource utilisation, lowering greenhouse gas emissions, and enhancing the reliability of the power supply. A consensus mechanism, encryption algorithm, distributed data storage, and point-to-point transmission are all examples of computer technologies that have been integrated and innovated to create blockchain [5]. Information security issues like trusted data transmission in the VPP can now be solved, thanks to blockchain technology's decentralisation, transparency, automation of contract execution, traceability, and other features. The conventional distributed energy system is in an off-design operation state for a considerable amount of time as a result of the dispersed geographic locations of each DER in the VPP and the fluctuation of new energy generation and consumption. The system performance will significantly deteriorate if the current level of technical capability and management tools is used. The smooth power supply and grid-connected operation of the VPP can be achieved more easily with the efficient cooperation of DER [6]. In order to achieve a distributed VPP and achieve a reasonable load distribution among units, this paper proposes to optimise the neural network by combining the PSO algorithm. This is based on the distributed VPP's consistency with blockchain in terms of decentralisation, point-to-point interaction, and decentralised coordination.

The characteristics of transparency and fairness of blockchain provide a new application scheme to solve the problems of aggregation control and opaque transaction of decentralised resources in the VPP [7]. Based on the limited communication mechanism, distributed dispatching realizes the confidentiality of information to a great extent and enables the DER holder to actively participate in the daily trading and operation of the electricity market without disclosing important information. Traditional economic dispatching solutions mostly use centralised methods, but the increasing penetration rate of DER in the power grid poses more challenges to the centralised dispatching. In the traditional centralised dispatching, the dispatching centre or the central coordinator needs to obtain the operation information of each unit through "point-to-point" communication [8]. In this paper, the internal DER coordination problem of the VPP is analyzed under the ubiquitous power of Internet of things background; a decentralised VPP scheduling and control model is proposed, and blockchain technology is used as the cooperative control means among DER. Its main innovations and contributions are as follows:

- (1) In order to increase the parallel processing ability of the network, a certain number of neural networks are trained in parallel by a certain number of subsample sets, respectively, and the training results are optimised by the PSO; finally, an optimal clustering neural network is obtained.
- (2) In order to satisfy the requirements for a dependable power supply and power quality, this paper will optimise the efficiency of the operation of the power system. The algorithm's efficacy is confirmed

through simulation tests, and distributed scheduling is accomplished using blockchain, which offers a workable reference plan for a decentralised VPP operation mode.

2. Related Work

A system called VPP integrates different kinds of power sources to deliver dependable overall power. In comparison to traditional power plants, VPPs have the benefits of flexibility, a high rate of new energy utilisation, and a high rate of self-absorption. For instance, they can quickly produce electricity to meet the peak demand. The VPP offers greater flexibility and efficiency while partially replacing conventional power plants. The system can respond to changes better if it is more flexible, but due to its complexity, this comes with complex optimization, control, and secure communication requirements.

Chen et al. put forward a collaborative dispatching model of wind, water, and wind VPP based on the classic scene set and transformed the random optimization problem into a classic scene [9]. Saad et al. compare several collaborative optimization algorithms of the VPP and simulate them with MATLAB [10]. Zhang et al. emphasized the importance of the VPP in strengthening distributed generation technology, and based on renewable energy, its value is amplified in the electricity market [11]. Al-Saedi et al. use the dynamic weight method to aggregate multi-objective into a series of single-objective optimization problems and then solve them separately [12]. Song et al. introduced the idea of the multiobjective evolutionary algorithm based on decomposition into the PSO. After aggregating multiple objectives into a single objective, the adjacent optimization problems learn from each other, thus reducing the computational cost [13]. Ameli et al. also used PSO to optimise different targets, respectively, and finally integrated them [14]. Wang et al. believe that blockchain technology is one of the most effective ways to communicate between the VPP and microgrid in the future [15]. Petersen et al. analyze the combination of the VPP and blockchain from five dimensions of sci & tech, economy, society, environmental protection, and academics [16]. Xu et al. research shows that the VPP can protect the power system by controlling the process of demand response and participating in the same, and by means of energy storage system, so as to maintain the smooth operation of the power grid [17]. Guo et al. gave the target order, dynamically determined the nearest neighbors on the first-dimension target, and determined the best among the nearest neighbors on the second dimension target, and used the neighborhood best instead of the global best, so that each particle flew to a different best, thus obtaining multiple optimal solutions [18]. Sakamuri et al. introduced the max-min function to compare the target vectors, which not only provided the dominant information between solutions but also contained diverse information [19]. Nappu and Arief improve the visibility of DER by providing interfaces between system components and use the optimal power flow algorithm to describe the VPP [20].

The results of the calculations made using this model will differ significantly from the actual situation, because in actual production, placing too much emphasis on one index will cause other indexes to be ignored or even weakened. This paper suggests a PSO-NN to implement distributed scheduling of the VPP based on the consistency of a decentralised VPP and blockchain in decentralisation, point-to-point interaction, and decentralised coordination.

3. Methodology

3.1. Collaborative Control Technology of the VPP. In VPP collaboration, DER exchanges energy under the guidance of VPP operators or through competitive bidding, and operators generally enhance their ability to restrain DER through restrictions on electricity prices, subsidies, and line power [21]. The issue of optimal power dispatching in the power system has risen to the fore as a result of the ongoing growth in the size of the power system and the increasingly intricate structure of the power grid. Reactive power flow in the network and network impedance parameters both play a role in the loss of active power. The study of the best load-allocation strategy among units is at the technical heart of the VPP dispatching management, and there are many different kinds of DER units. The nonconvex operating conditions of thermal power units, such as the valve point loading effect, forbidden operating area, and multifuel option, should be taken into account when actually dispatching power, in addition to the operating characteristics of renewable energy sources and energy storage systems.

In distributed architecture, agents are independent, completely equal, and have no logical master-slave relationship. According to the predefined agreement, we can determine our respective tasks and coordinate our respective behaviours and activities according to the system's goal and state, as well as our own state, ability, and information. The communication mode between agents is shown in Figure 1.

Because the DERs in the VPP do not trust one another and come from different energy subjects, such as new energy power plants, traditional energy power plants, and energy storage and load, they are unable to verify the validity and legality of the information exchanged. A consensus mechanism, an encryption algorithm, distributed data storage, and point-to-point transmission are all examples of computer technologies that have been integrated and innovated into blockchain. Simply put, blockchain is a technical system that uses cryptography, decentralisation, and de-trust to allow any number of nodes to jointly maintain a trustworthy database. The blockchain's data structure affects the information composition of the successor node and enables the predecessor node to track the information of each block in the chain. The reactive power balance in each area can be roughly maintained by adjusting the ratio of reactive power sources and transformer branches acting as tie lines in each area. By doing this, the flow of reactive power between power grids with various voltage levels in various areas can be decreased, serving the goal of limiting the large-scale flow of reactive power in the power grid.

Assuming that the generator set cost function in DER is quadratic and the cost function is represented by $F_i(P_i)$, the minimum power generation cost to achieve VPP is as follows:

$$\begin{aligned} \min F &= \min \sum_{i=1}^n F_i(P_i), \\ F_i(P_i) &= a_i P_i^2 + b_i P_i + c_i, \end{aligned} \quad (1)$$

where n represents the number of DER units in the VPP, and P_i represents the output power of the unit i . The total cost of the VPP is recorded as F . a_i , b_i , and c_i represent the coefficients of the cost function.

During operation, every DER unit in the VPP satisfies the active power balance of the entire system.

$$\sum_{i=0}^n P_i = P_{LD}. \quad (2)$$

Among them, P_{LD} represents the total load demand of all users. The load is distributed among the units according to the "equal consumption microincrease rate criterion." When the optimal operation is achieved, the microincrement characteristics λ of all the units are consistent, and λ can be calculated by the first-order differential, namely,

$$\lambda = \frac{dF_i}{dP_i} = 2a_i P_i + b_i. \quad (3)$$

Therefore, λ can be used as a consistency variable between nodes in blockchain, and it can be adjusted as the load changes, but the entire network remains consistent.

In the blockchain system, intelligent contracts ensure that both parties' rights and obligations as well as the determination of the contract's execution are upheld. Once the requirements are satisfied, the transaction will be carried out automatically without artificial promotion or oversight by a third party, which greatly increases the efficiency of the transaction execution. Distributed energy and power sources will continue to proliferate, and transactions involving the VPP and related power generation resources will become more frequent. The adoption of adaptive and decentralised energy scheduling will increase, and its isomorphic blockchain will give distributed energy sources a solid and reliable foundation for data interaction. A strong theoretical and practical foundation is provided for the architecture design of a distributed energy system based on blockchain by the decentralised cooperation mode of a power system based on the multiagent consistency theory and current mainstream blockchain technology. The blockchain consensus mechanism enables effective distributed communication between units, and data broadcasting and information interaction are used to ensure the consistency of each unit's operational characteristics. It is therefore possible to implement the blockchain consensus mechanism to achieve the distributed dispatching strategy and the best VPP performance. Through the aforementioned master node selection algorithm, the power supply node and the power consumption node choose the master node. The master node then predicts the power consumption data for the upcoming time period

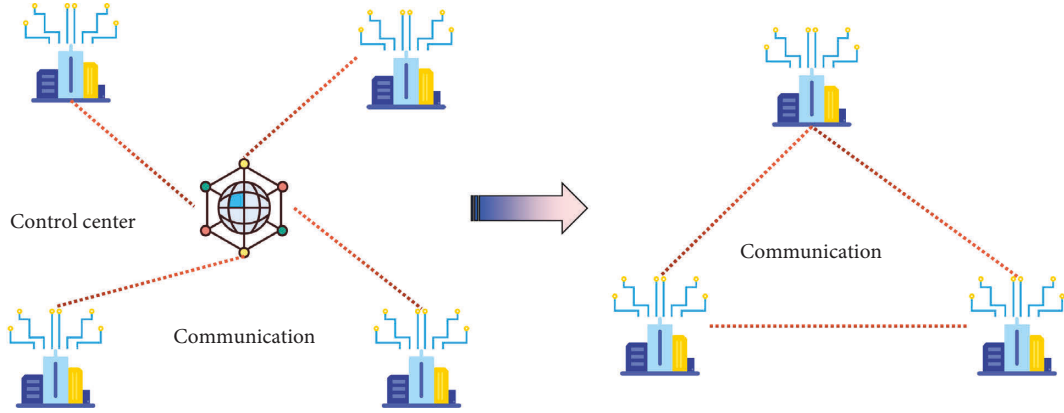


FIGURE 1: communication method between agents.

and sends the predicted data to the blockchain for propagation.

3.2. Distributed Scheduling of the VPP Based on the PSO-NN. Because the complex system contains many variables, the relationship between variables is complex and changeable, and the system scale is relatively large; there are more uncertain factors that have important influence. Multiagent technology has the characteristics of autonomy, distribution, and coordination, which can realize the self-organization, self-learning, and reasoning ability of the system. When multiagent technology is used to solve practical problems, it has robustness, reliability, and high problem-solving efficiency. The decentralised cooperation mode of the power system based on multiagent consistency theory is highly consistent with the concept of current mainstream blockchain technology, which provides a good theoretical and practical foundation for the architecture design of the distributed energy system based on blockchain.

The load forecasting model in the power system makes predictions about the future load demand based on historical data and the current power supply situation. The relationship between historical data is complicated for a number of reasons, and some of the data are even wrong. Additionally, load forecasting using conventional statistical methods must be based on data from a large sample of loads. The outcomes of using traditional methods are frequently very dissimilar from the reality when dealing with the relationship between such a large amount of complex data. Few parameters, a straightforward structure, and straightforward operation are the benefits of the PSO. After improvement, it has a good ability to locate the global optimal solution and can successfully prevent the algorithm's premature convergence. Because the PSO algorithm is better suited to solving single-objective problems, it suffers from some drawbacks when applied to multiobjective problems. These drawbacks include low solving efficiency, a high number of subjective experience factors, a significant amount of calculation, complex algorithm settings, and front-end sensitivity. The structure and operation mode of the VPP system are shown in Figure 2.

The fundamental idea behind artificial neural networks [22] is to mimic how the human brain works in order to transmit data and perform intricate operations between neurons. In essence, there are two processes that make up an artificial neural network's operation. The first is the training process. In this procedure, the neural network's weights and partial weights are acquired through training the neural network. The simulation process is the second. The simulation of the neural network yields information about the prediction output value or accuracy of the network. PSO is a novel kind of swarm intelligence optimization algorithm that excels at parallel search, ease of implementation, simplicity, and computational efficiency. It can find the global optimal solution to the issue with a high probability and is appropriate for complex optimization problems. Particles draw lessons from the groups and their own successful information gathering experiences to inform their next course of action. The flow of the PSO-NN algorithm is shown in Figure 3.

M is the number of neurons in the input layer, N is the number of neurons in the hidden layer, w_{ij} is the network connection weight between neurons in the input layer and the hidden layer, and w_j is the network connection weight between neurons in the hidden layer and the output layer. The expression of the implicit layer function in this paper is as follows:

$$\phi(x) = \cos(1.75x) \exp\left(-\frac{x^2}{2}\right). \quad (4)$$

By scaling and translating the above formula, the wavelet basis function can be obtained as follows:

$$\phi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right), \quad (5)$$

where a and b are the scaling factor and translation factor, respectively. The output layer neuron function expression in this paper is as follows:

$$\delta(x) = \frac{1}{(1 + \exp(-x))}. \quad (6)$$

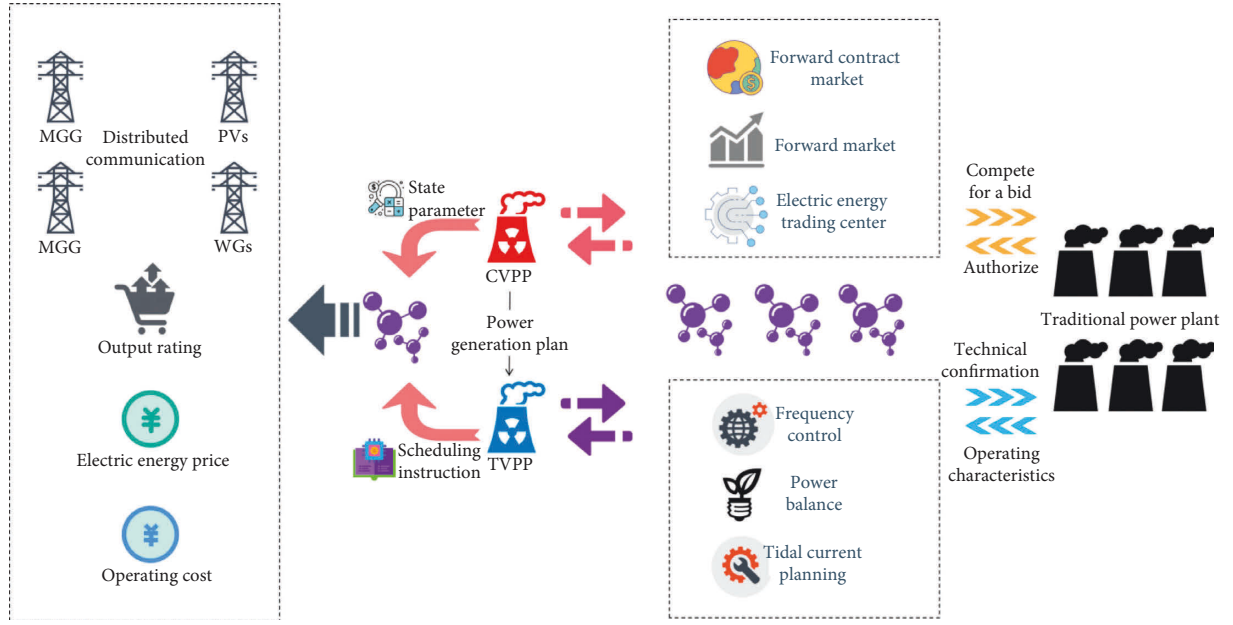


FIGURE 2: The structure and operation mode of the VPP system.

Its output is as follows:

$$y(x) = \delta \left[\sum_{j=0}^N w_j \psi_{a,b} \left(\sum_{m=0}^M w_{ij} x_m \right) \right]. \quad (7)$$

The VPP participates actively in power market transactions as a significant component of the power market. There is currently no two-way information symmetry between distributed energy and the VPP, and as a result, the benefit distribution mechanism in the VPP is not currently accessible to the outside world. This results in higher credit costs and higher transaction costs when buying and selling electricity. The VPP has the characteristics of centralised nodes and requires sufficient authority to coordinate, induce, and control the grid-connected behaviour of each DER in order to ensure the safe and reliable operation of the power system. A decentralised platform is currently required to guarantee fairness. The network's clustering results are unstable because of the sensitivity to data initialization and the difficulty of using the traditional linear connection weight function to show the subtle differences between the input and output linear transformation attribute values. Through two-way communication technology, the VPP implements the scheduling of information and data for each component, such as the power generation side, the demand side, and the electricity trading market. The operation scheduling process of the VPP incorporates a proposed energy blockchain network model, allowing DER to effectively participate in electricity market transactions. Through the cryptographic features of blockchain itself, it guarantees the VPP to obtain a higher level of information security while also increasing the overall operation efficiency of the VPP.

The programmable feature of intelligent contract enables both parties to agree on various transaction terms, and it is

applicable to all kinds of procedural rules and has a very broad application scenario. In the operation of the VPP, the transactions of each node can be automatically and safely executed through intelligent contracts. In the whole process of iterative optimization, the examples of failures experienced by particles can be shown as the positions of poor fitness values of particles themselves or groups of particles. Let us assume that the worst position searched by particles so far is as follows:

$$s_i = (s_{i1}, s_{i2}, s_{i3}, \dots, s_{im}). \quad (8)$$

The worst position searched so far by the entire particle swarm in the iterative optimization process is as follows:

$$s_g = (s_{g1}, s_{g2}, s_{g3}, \dots, s_{gn}). \quad (9)$$

Then, the velocity and position update formula of the i -th particle can be obtained, namely,

$$\begin{aligned} v_{id}^{k+1} &= \omega v_{id}^k + c_1 r_1 (x_{id}^k - s_{id}^k) + c_2 r_2 (x_{id}^k - s_{gd}^k), \\ x_{id}^{k+1} &= x_{id}^k + v_{id}^{k+1}. \end{aligned} \quad (10)$$

The above formula, which states that particles only learn from unsuccessful examples, would obviously lead to an update of particles that is inconsistent with actual experience. The idea of inertia weight is added to the basic PSO in order to enhance the convergence performance and optimise the solution space. The modified version of the original PSO algorithm's velocity and position formula is as follows:

$$\begin{aligned} v_{id}^{k+1} &= \omega v_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k), \\ x_{id}^{k+1} &= x_{id}^k + v_{id}^{k+1}. \end{aligned} \quad (11)$$

In the formula, $v_i = [v_{i1}, v_{i2}, v_{i3}, \dots, v_{im}]$ is the speed of the particle i , which represents the distance between the

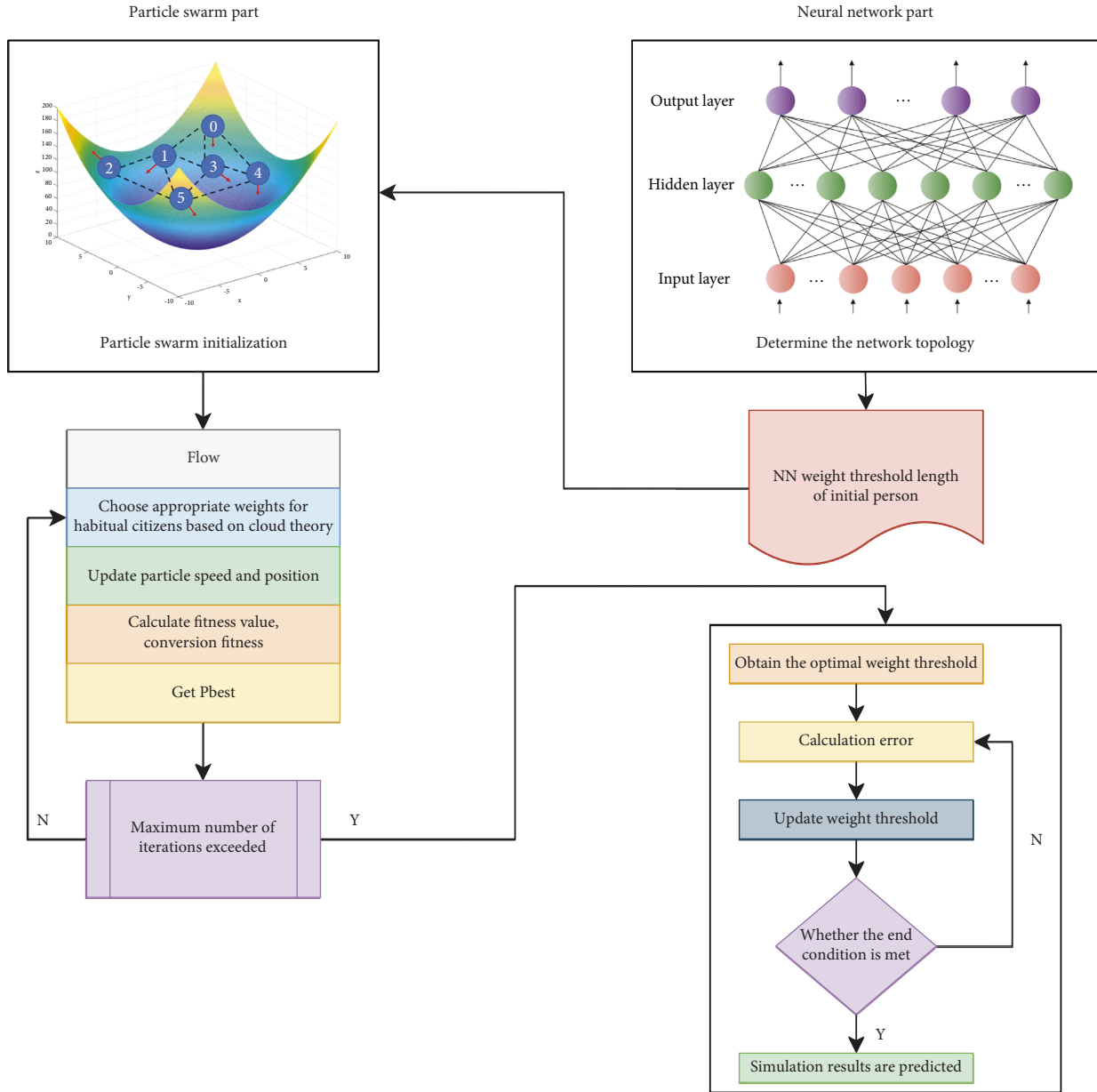


FIGURE 3: PSO-NN algorithm flow.

current position of the particle i and the next target position; $x_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$ is the current position of the particle i ; p_i is the individual optimal solution searched by the particle so far; p_g is the optimal solution searched by the entire particle swarm so far. ω is the inertia weight. Due to the sensible selection of the inertia weight, the particle has a balanced capacity for exploration and development. The inertia weight value is used to characterise the size of the particle's current speed inheritance.

The number of the two subpopulations is constantly changing, so that each particle can get a lot of learning information from its own experience and the group's experience. After every certain iteration, the particles in the population are adjusted according to the proportional coefficient, and the whole population is re-formed into two

new subpopulations [23]. When the optimization is in the late stage, the particles are concentrated near the optimal value. At this time, the number of particles that learn from the failure experience will be far less than the number of particles that get information from the success experience. Then, the particles continue to iterate and update until all the particles in the population adopt the learning strategy of finding the optimal value to iterate. When evaluating the quality of particles, the fitness and concentration of particles should be taken as the standard. If the fitness of particles is better and the concentration of particles is lower, then the quality of particles will be better. Therefore, the improved algorithm will suppress the particles with poor quality, and all particles are equipped with a mutation rate, and the mutation rate will change with the change of particle quality.

4. Result Analysis and Discussion

The pollution discharge can be significantly improved by readjusting and combining the output of different fuel units through dispatching methods, but the cost may go up. The reason is that low-carbon energy units, such as the liquefied natural gas and low-sulfur fuel, will produce more output than other units when the dispatching process is oriented to reduce the pollution discharge, increasing the cost of power generation. The fundamental requirement of the entire power system is to pursue the maximum economic benefit of the system, which is predicated on ensuring the safe and dependable operation of the system. High production, transmission, distribution, and consumption efficiencies are all referred to as aspects of the electric power system’s economy. Even though there are numerous factors that affect the cost of power generation, developing a model is difficult, but it can accurately represent the distribution of output among generators across the entire power system. The simulation is performed through experiments to confirm the efficacy of the algorithm suggested in this paper. For the unit’s operating conditions and starting power, see Table 1.

MGGs are connected through a blockchain network, and each MGG is a node in the network. In the presence of network delay, we test the change of consistency variables.

Figure 4 demonstrates that the system is not functioning at its best because at the initial time $t = 1$, the consistency variables of each unit are different, which does not satisfy the requirement of an equal consumption increment rate. The system reaches its best operational state at time $t = 6.6$, when the consistency variables are consistent. Figure 5 compares the total power consumed by the load and the total power produced by the unit during the consensus-building process.

From Figure 5, we can find the system fluctuation caused by network delay and consensus calculation, but the power balance is finally achieved. Figure 6 shows the active power adjustment of each MGG unit, and the final power value is stable in the optimal operation state.

In order to effectively coordinate the regional electricity demand with the electricity demand of the electricity wholesale market, VPP technology can recognise the possibility that a household or individual load will feed back excess electricity to the power grid. It can also allocate the working hours of periodic distributed generation and distributable distributed generation in a reasonable manner. The blockchain system has strong robustness and reliability in data storage because there is no centralised central control, all nodes can back up the information in blockchain partially or completely, and the data loss of any node will not affect the system’s normal operation. A specific incentive mechanism is used to make sure that every node in the distributed system takes part in the information exchange process in the blockchain system, which is operated and maintained by every node in the network. Enterprises give each objective varying levels of importance when making multiobjective decisions. Businesses can sort different objectives first for better decision-making, prioritise important objectives when choosing schemes, eliminate those that

TABLE 1: Operating parameters of the unit.

Generator	Cost coefficient	Generating power of unit
MGG 1	1.225	35
MGG 2	1.121	40
MGG 3	1.331	65

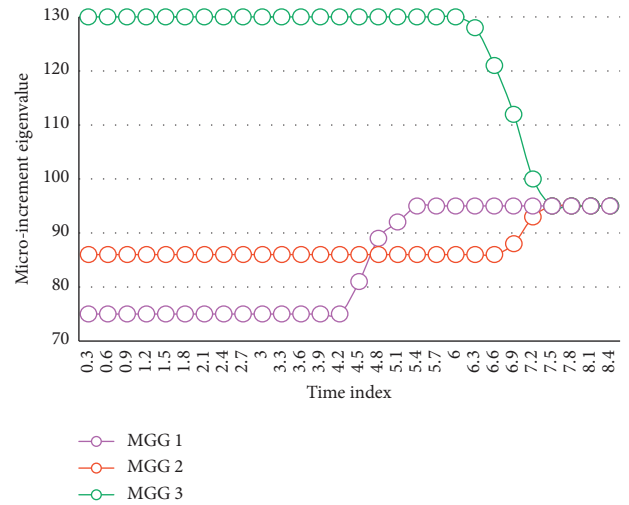


FIGURE 4: Changes of microincrement eigenvalues.

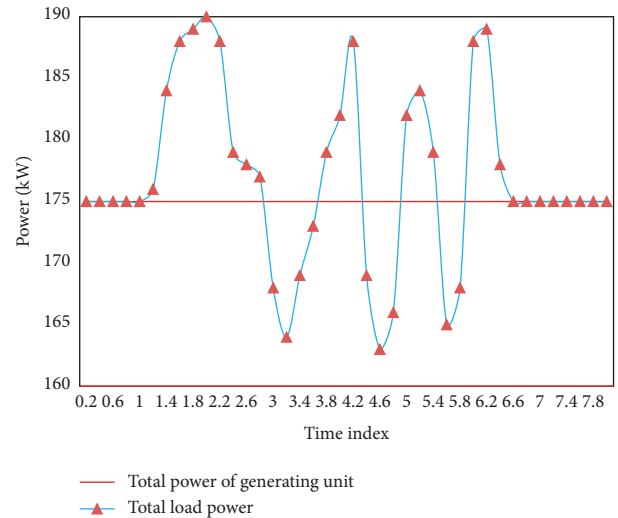


FIGURE 5: Comparison of the total load power and total power generation.

cannot meet important objectives, prioritise secondary objectives when choosing schemes, and choose a few schemes before choosing the best one.

The difference between the actual output and the expected output of power prediction based on the PSO-NN is shown in Figure 7. The corresponding predicted performance indicators are shown in Table 2. The comparison between the short-term power forecast result based on the PSO-NN and the actual value is shown in Figure 8.

The prediction error and prediction result graphs show that when compared to the prediction error and prediction

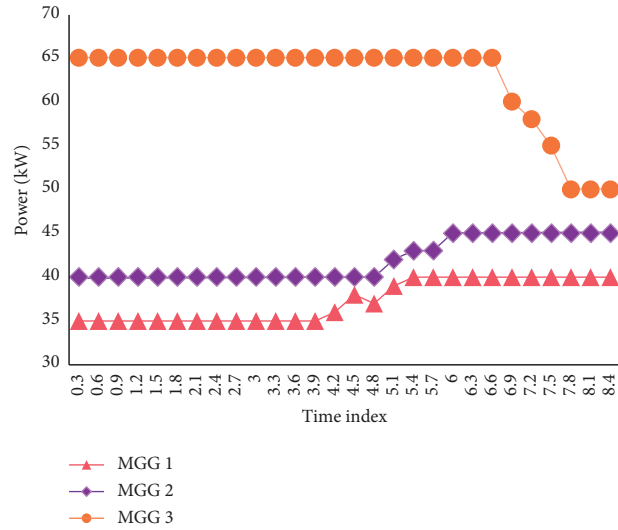


FIGURE 6: Power variation of each unit.

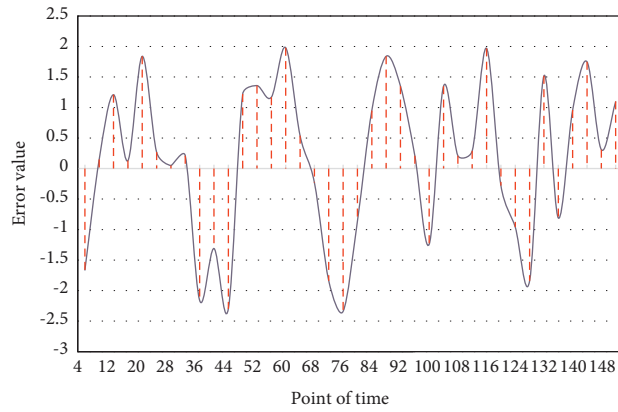


FIGURE 7: Electricity prediction error based on the PSO-NN.

TABLE 2: Comparison of evaluation indicators.

	Average absolute error	Mean square error	Average absolute percent error	Mean square percentage error
PSO-NN	3.510	22.332	0.076	0.018
BPNN	2.991	13.274	0.062	0.008

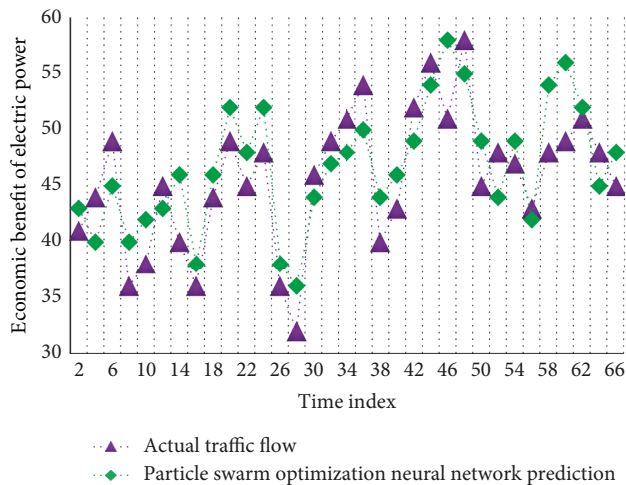


FIGURE 8: Electric power prediction simulation based on the PSO-NN.

graph of the BP neural network, the prediction error of the PSO-NN is smaller, and its prediction effect is better than that of the BP neural network. The prediction accuracy rate of the PSO-NN can reach 94.98 percent. The mean square error of the prediction effect is 13.27, and the average absolute error, mean square error, mean absolute percentage error, and mean square percentage error are all less than those predicted by the BP neural network, according to the evaluation index of prediction performance. Supplying electricity to units with high electric energy utilisation efficiency and low pollution emission as much as possible can effectively improve the electricity utilisation efficiency, protect the environment and protect energy, and achieve the effect of conservation under the conditions of the electricity market, according to the indicators of electric energy utilisation efficiency, pollution emission, and industrial production capacity of each electricity consuming unit.

5. Conclusion

The primary research focus of this paper is on how to make the power system meet all operational stability indicators while also minimising power generation costs, reducing environmental pollution, and minimising waste discharge. The traditional VPP has some drawbacks, including centralised control, mistrust between distributed energy agents, and simple information transmission modification. Blockchain technology's decentralised, anonymous, and transparent transmission can easily fix these issues. The PSO algorithm is used in this paper to optimise the neural network, achieve the distributed VPP dispatching, and achieve a fair load distribution among the units. The PSO algorithm is prone to local convergence and finds the global optimal value with difficulty because once the local optimal value is discovered during the optimization process, other particles will quickly move toward it. According to the simulation results, this algorithm's error is negligible and its accuracy can reach 94.98 percent. In addition to effectively resolving the issues of mutual mistrust and high information exchange costs among the key players in the current VPP, this model can more accurately reflect the real-time information of the demand side. To solve the problem of achieving a more reliable and efficient consensus in the high error rate environment is the focus of the next stage. For the neural network algorithm of PSO, the next step can be to divide the particle swarm into two subpopulations and optimise the subpopulations, respectively.

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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (U1866206).

References

- [1] S. Rahmani-Dabbagh and M. K. Sheikh-El-Eslami, "A profit sharing scheme for distributed energy resources integrated into a virtual power plant," *Applied Energy*, vol. 184, no. DEC.15, pp. 313–328, 2016.
- [2] Z. Chen, A. Luo, H. Kuang, L. Zhou, Y. Chen, and Y. Huang, "Harmonic resonance characteristics of large-scale distributed power plant in wideband frequency domain," *Electric Power Systems Research*, vol. 143, no. Feb, pp. 53–65, 2017.
- [3] M. Delfanti, D. Falabretti, and M. Merlo, "Energy storage for PV power plant dispatching," *Renewable Energy*, vol. 80, no. aug, pp. 61–72, 2015.
- [4] J. Zhang, J. Yan, D. Infield, Y. Liu, and Fs Lien, "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Applied Energy*, vol. 241, no. 20, pp. 229–244, 2019.
- [5] D. Yang, H. Wei, Y. Zhu, P. Li, and J. C. Tan, "Virtual private cloud based power-dispatching automation system-architecture and application," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1756–1766, 2019.
- [6] P. Fu, N. Wang, and X. Li, "CPS-based load dispatching model for the energy conversation and emission reduction of thermal power units," *Proceedings of the CSEE*, vol. 35, no. 14, pp. 3685–3692, 2015.
- [7] J. Yu, H. Ji, Q. Song, and L. Zhou, "Design and implementation of business access control in new generation power grid dispatching and control system," *Procedia Computer Science*, vol. 183, no. 22, pp. 761–767, 2021.
- [8] B. Qza, B. Gwa, and C. Khab, "Framework and technologies of App store in new generation power grid dispatching and control system - ScienceDirect," *Procedia Computer Science*, vol. 183, pp. 754–760, 2021.
- [9] B. Chen, H. Chen, Y. Zhang, J. Zhao, and E. Manla, "Erratum: Chen, B., et al. risk assessment for the power grid dispatching process considering the impact of cyber systems," *Energies*, vol. 12, p. 1084, 2019.
- [10] N. H. Saad, A. A. El-Sattar, and A. E. A. M. Mansour, "Improved particle swarm optimization for photovoltaic system connected to the grid with low voltage ride through capability," *Renewable Energy*, vol. 85, no. JAN, pp. 181–194, 2016.
- [11] H. Zhang, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Peer-to-peer energy trading in dc packetized power microgrids," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 17–30, 2020.
- [12] W. Al-Saedi, S. W. Lachowicz, D. Habibi, and O. Bass, "Power flow control in grid-connected microgrid operation using particle swarm optimization under variable load conditions," *International Journal of Electrical Power & Energy Systems*, vol. 49, no. jul, pp. 76–85, 2013.
- [13] Z. Song, J. Zhang, Z. Zheng, and X. Xiao, "Peak dispatching for wind power with demand-side energy storage based on a particle swarm optimization model," *Utilities Policy*, vol. 56, no. FEB, pp. 136–148, 2019.
- [14] A. Ameli, S. Bahrami, F. Khazaeli, and M. R. Haghifam, "A multiobjective particle swarm optimization for sizing and placement of DGs from DG owner's and distribution

- company's viewpoints," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1831–1840, 2014.
- [15] N. Wang, P. Fu, and D. Chen, "Application of big data analytics in plant-level load dispatching of power plant," *Proceedings of the CSEE*, vol. 35, no. 1, pp. 68–73, 2015.
- [16] M. K. Petersen, L. H. Hansen, J. Bendtsen, K. Edlund, and J. Stoustrup, "Heuristic optimization for the discrete virtual power plant dispatch problem," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2910–2918, 2014.
- [17] J. Xu, Y. Gu, D. Chen, and Q. Li, "Data mining based plant-level load dispatching strategy for the coal-fired power plant coal-saving: a case study," *Applied Thermal Engineering*, vol. 119, no. Complete, pp. 553–559, 2017.
- [18] W. Guo, P. Liu, and X. Shu, "Optimal dispatching of electric-thermal interconnected virtual power plant considering market trading mechanism," *Journal of Cleaner Production*, vol. 279, no. 54, Article ID 123446, 2021.
- [19] J. N. Sakamuri, Z. H. Rather, J. Rimez, M. Altin, O. Goksu, and N. A. Cutululis, "Coordinated voltage control in offshore HVDC connected cluster of wind power plants," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, pp. 1592–1601, 2016.
- [20] M. B. Nappu and A. Arief, "Network losses-based economic redispatch for optimal energy pricing in a congested power system," *Energy Procedia*, vol. 100, no. 5, pp. 311–314, 2016.
- [21] J. Branke, T. Hildebrandt, and B. Scholz-Reiter, "Hyperheuristic evolution of dispatching rules: a comparison of rule representations," *Evolutionary Computation*, vol. 23, no. 2, pp. 249–277, 2015.
- [22] Y. Ding, Z. Zhang, X. Zhao, and Y. Cai, "Self-supervised locality preserving low-pass graph convolutional embedding for large-scale hyperspectral image clustering," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.
- [23] J. Jin, P. Zhou, M. Zhang, X. Yu, and H. Din, "Balancing low-carbon power dispatching strategy for wind power integrated system," *Energy*, vol. 149, no. APR.15, pp. 914–924, 2018.

Research Article

Comparison and Suggestions of Logistics Performance Index of Main Countries of Belt and Road Strategy Based on Bootstrap DEA Model

Wen-Tsao Pan ¹, Bingqian Jiang ², Yuting Wang ², Yueyuan Cai ², and Xiaoxia Ji ²

¹School of Management, Guangzhou Huashang College, Guangzhou, China

²School of Business, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence should be addressed to Yuting Wang; 20190401763@gdufs.edu.cn

Received 2 May 2022; Revised 14 June 2022; Accepted 21 July 2022; Published 13 September 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Wen-Tsao Pan 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 an important economic sector, logistics is becoming more important, if not crucial, in economic growth. In our nation, the logistics industry is booming, and it's just getting better. However, in addition to focusing on the positive aspects of our country's logistics industry's development, we should also analyze and address the negative aspects of our country's logistics industry's development. The overall logistics pattern has not yet been formed, and there is an urgent need for systematic construction. The regional development is extremely unbalanced. By comparing the logistics performance indices of various Belt and Road countries, this research aims to examine the major elements influencing overall logistics performance. Second, we introduce the Moran index to explore the geographical association of the subdivision indicators of the logistics performance index using the spatial econometric model. The bootstrap DEA analysis method examines and ranks the countries' logistics performance indexes, determines our country's advantages and disadvantages in comparison to other Belt and Road countries, and executes specific improvement strategies that will enhance logistics and boost the overall growth of our country's logistics sector.

1. Introduction

As an important sector in the economic field, the logistics business plays an increasingly critical and even decisive role in economic development. In our nation, the logistics industry is booming, and it's just getting better. The logistics business in our nation has gradually developed since the founding of the People's Republic of China, and the amount of both supply and demand has expanded dramatically. However, in addition to focusing on the positive aspects of our country's logistics industry's development, we should also analyze and address the negative aspects of our country's logistics industry's development. The overall logistical plan has not yet been established, and systematic building is urgently required. There is a lack of specialization in the logistics industry, as well as few connected needs. Good international integration has been established; nevertheless, the level of logistics informatization is poor, and

efficiency must be addressed immediately [1]. Logistics performance is frequently reflected by the Logistics Development Index as an important indicator of the level of logistics development (LPI). Low-performance logistics will increase trade costs, stifle the flow of goods, and weaken market competitiveness significantly. Customs clearance efficiency, logistics infrastructure quality, international transportation convenience, logistics service quality, ability to track products, and goods transit timeliness are all part of the logistics performance index system. The lower a country's trade expenses and the stronger its position in the global value chain, the higher its LPI score.

Tang Xiaoming et al., based on Tibet's geographical advantages and actual logistics development, analyzed the existing logistics development in Tibet Conditions, logistics nodes, and channels, and proposed the "five-in-one" Tibet logistics development strategy system framework [2]. Tao Zhang and Qiao Sen not only pointed out that national

logistics performance has an impact on the “One Belt and One Road,” but also that national logistics performance has an impact on the “One Belt and One Road.” The route has a substantial positive impact on the amount of trade in the countries and regions that it passes through [3]. Wang Chao and others pointed out that, compared to foreign nations, notably industrialized countries in Europe and America, logistics performance evaluation technology is in its infancy. However, logistics performance analysis is beneficial to increasing a company’s competitiveness and supporting its growth. As a result, future domestic study on logistics performance is crucial [4]. “Influence of Logistics Performance of the twenty-first Century Maritime Silk Road on China’s Export of Mechanical and Electrical Products,” by Liu Zuankuo and colleagues, established an expanded trade gravity model and empirically analyzed the logistics performance of the twenty-first Century Maritime Silk Road. The essay makes specific policy recommendations based on the empirical findings [5].

The bulk of previous research has concentrated on a basic examination and analysis of the logistics efficiency index. The complete index of logistics efficiency index is rated and compared with a single analysis to assess its influence on the growth of our country’s logistics industry and to define the development patterns and laws of the logistics industry. Existing research approaches include fuzzy comprehensive evaluation, multiple regression, entropy weight method, gray correlation analysis method, and others. No one has attempted a second in-depth analysis of the logistics development index using the bootstrap DEA technique.

This study intends to analyze the primary variables impacting overall logistics performance by comparing the logistics performance indices of different Belt and Road nations. Second, using the spatial econometric model, investigate the geographical correlation of the subdivision indicators of the logistics performance index by introducing the Moran index. The bootstrap DEA analysis method examines and ranks the logistics performance index of the countries involved, identifies our country’s advantages and disadvantages in comparison to other Belt and Road countries, and implements targeted improvement measures that will promote the overall development of our country’s logistics industry and improve logistics competitiveness.

2. Research Method

This study will adopt the following three research methods.

2.1. Visual Data Analysis. With the help of visual graphics software, we will show the lengthy data tables in the form of charts, such as line charts and bar charts, to make the data expression more visual and help to convey the expressed problems to the readers. Converting the original statistical tables into intuitive and visual charts is conducive to further comparison and analysis, as well as in-depth discovery. In terms of sample selection, we selected 22 countries as important representatives according to the world bank’s

logistics performance index and the distribution of countries along the “the Belt and Road”. The relevant logistics performance index and other data of 22 countries selected in this study are complete. The Central Asian economic belt includes Russia, Afghanistan, India, Pakistan, Iran, Turkey, Saudi Arabia, Iraq, Syria, and Jordan (the Asia Europe economic belt. Includes Germany, France, Britain, Italy, Ukraine, and Egypt). See Table 1 for the coverage and division of sections of the Silk Road Economic Belt.

2.2. Spatial Econometric Model. At the annual conference of the Netherlands Statistical Association, J. Paelinck first proposed spatial econometrics. As scientists continue to develop spatial autoregressive models, the spatial autoregressive model (SAR), the spatial error model (SEM), and the spatial Durbin model (SDM) are three classic spatial econometric models that are often used in research. The spatial autocorrelation model is one of them, and it seeks to determine whether a variable is connected in a geographical space area, as well as the degree of correlation. A typical statistic is the spatial autocorrelation coefficient. There are two types of autocorrelation: global space autocorrelation and local space autocorrelation. This approach is used to investigate the global and local spatial autocorrelation of China’s territory in this study.

This study will realize the construction and analysis of the spatial econometrics model based on GeoDa software. As indicated by Moran’s I, the global spatial autocorrelation mainly explores the degree of spatial reliance of each attribute variable over the whole area. Following data processing, the Moran scatter plot generated by the GeoDa program will be separated into four quadrants based on the attribute level of the area and the surrounding area, as well as the different types of the area space, namely high-high, high-low, low-low, and low-high. Scholars might use the divided four quadrants to see if a region has substantial spatial agglomeration characteristics.

Local spatial autocorrelation starts from the specific areas divided within the overall range and analyzes whether the space between attribute variables obeys the trend, that is, whether they are similar. First, we will use the GeoDa software to calculate the Moran index and use the results of the Moran index to get the change of attribute variables with the position. When the value is greater than 0, it is a positive correlation, and objects with similar attributes gather together, otherwise, objects with different attributes gather together. In addition, there is another possibility that when the Moran index approaches 0, it means that there is no spatial autocorrelation in the sample. Secondly, according to the results calculated by the single variable Moran index, the Lisa clustering map will be generated, and the local spatial autocorrelation will be visualized, which can further intuitively analyze the influencing factors.

2.3. Bootstrap DEA Model. Due to the restricted observation sample, it is difficult to avoid the issue of sample sensitivity and extreme value effect on the computed efficiency value since the DEA model has some of the benefits of different

TABLE 1: The coverage and division of sections of the Silk Road Economic Belt.

Arrangement	Region	Major economies	Country abbreviation
Core area	Central Asian economic Belt (five Central Asian countries)	China	CHN
		Kyrgyzstan	KGZ
		Kazakhstan	KAZ
		Turkmenistan	TKM
		Tajikistan	TJK
Important area	Central Asia Economic Belt (Central Asia, West Asia, Russia, India, Pakistan)	Uzbekistan	UZB
		The Russian Federation	RUS
		Afghanistan	AFG
		India	IND
		Pakistan	PAK
		Islamic Republic of Iran	IRI
		Turkey	TUR
		Saudi Arabia	SAU
		Syrian Arab Republic	SYR
		Iraq	IRQ
Jordan	JOR		
Expansion area	Asia Europe Economic Belt (Central Asia, Europe and North Africa)	Germany	GER
		France	FRA
		The U.K.	GBR
		Italy	ITA
		Arab Republic of Egypt	ARE
		Ukraine	UKR

parameter estimation techniques (DMU). The efficiency value achieved with the DEA model is actually “relative efficiency.” In terms of absolute efficiency values, this estimate is biased and inconsistent. To overcome this flaw, Simar [6] et al. presented the Bootstrap-DEA method¹⁹, which can estimate the confidence interval, correct the bias of the DEA estimates, and establish the significance level¹⁸. The primary idea behind this method is to use the Bootstrap²⁰ concept to sample the original samples repeatedly, create multiple Bootstrap sample data, collect many Bootstrap efficiency values, and construct confidence intervals using the empirical distribution of Bootstrap efficiency values. To increase the consistency of standard DEA estimators, use statistical inference. Overall, the DEA estimated efficiency value predicts the real efficiency value of the original sample, while the Bootstrap efficiency value is calculated by estimating and bias-correcting the DEA estimated efficiency value using a large number of simulated Bootstrap samples.

The method steps are as follows:

- (1) For each uses the traditional DEA method to calculate the efficiency value of the sample data
- (2) For the efficiency obtained in the first step, use the Bootstrap method to randomly sample n efficiency values where b represents the use of the b -the iteration of the Bootstrap method
- (3) Calculate the Bootstrap method to simulate the sample
- (4) Use the traditional DEA method to simulate each Bootstrap method sample and calculate the efficiency value again

- (5) Repeat steps 2–4 for a total of B times to generate the efficiency value

Three variables of Input (X), Output (Y), and DMUH-been selected tedd for bootstrap DEA model analysis, and the DUM ranking and deviation correction under the combined action of two variables of x and y are discussed. Among them, DUM represents the name of the ranking unit being evaluated.

3. Empirical Analysis

3.1. Data Sources. The data used in this study are obtained from World Bank publications [7–11].

3.2. Variable Selection and Description. Table 2 shows the independent variable (INPUT) and dependent variable (OUTPUT) settings. Since the composite index is the upper-level indicator of other subindices according to the World Bank’s official indicators, the composite index is used as the output index, while the other six subindices are used as the input index.

3.3. Research Methods and Data Computing

3.3.1. Visual Chart Analysis. In 2007, 2010, 2012, 2014, 2016, and 2018, the World Bank published four logistics performance index reports. This post will use visual charts to compare the horizontal and vertical situations of the Belt and Road’s logistics development, as well as the specific situation of our country’s logistics development in the Belt and Road countries.

TABLE 2: INPUT and OUTPUT introduction.

OUTPUT (Y)	INPUT (X)
Y ₁ : composite index	X ₁ : how often do the goods arrive at the consignee within the scheduled or expected time
	X ₂ : quality of trade and transport-related infrastructure
	X ₃ : efficiency of customs clearance procedures
	X ₄ : ability to track and inquire about goods
	X ₅ : ease of arranging competitively priced shipments
	X ₆ : ability and quality of logistics services

Due to a large amount of raw data, only the comprehensive logistics performance score is selected for visual analysis here.

From the fluctuation trend of Figure 1, it can be seen that the comprehensive scores of the logistics performance of the major countries along the “Belt and Road” are quite different. The highest comprehensive score is Germany in the expansion area, which is as high as 4.20, while Afghanistan in the important area, whose comprehensive logistics performance score is only 1.73, is in a relatively inferior state, and the gap with other countries is obvious.

Figures 2 and 3 show the Logistics Performance Index for Core Area and Important Area from 2007 to 2018. On the whole, there is a large difference in the comprehensive scores of countries in important regions. Turkey has had a high comprehensive score in the past six years, followed by India and Saudi Arabia. However, compared with China, there is still a certain gap, which has not broken through 3.5. Afghanistan and Iraq have poor comprehensive scores of logistics performance, both of which are no more than 2.5. From the data, these two countries have low levels in the capacity and quality of logistics services and the quality of trade transportation-related infrastructure. They should strengthen exchanges with other countries and constantly improve own level in logistics services and trade and transportation.

Of the countries in the expansion zone, Germany has the highest score, as shown in Figure 4. And its comprehensive score has stabilized at more than 4.0 from 2007 to 2018, followed by Britain, France, and Italy, with relatively high comprehensive scores. The Arab Republic of Egypt and Ukraine have the lowest comprehensive scores, and their logistics performance needs to be strengthened. It is not difficult to see that Germany has high scores in the ability and quality of logistics services, the ability to track and query goods, and the quality of trade and transportation-related structure. Although China is in a relatively stable state in all aspects, compared with Germany, China’s logistics development level is still relatively backward. We should give full play to the comparative advantages of China’s logistics industry and realize industrial upgrading according to the principles of complementary advantages, mutual benefit, and win-win results.

In general, our country’s logistics development is at a medium level, and the logistics performance indexes of several established developed countries with a small proportion of expansion areas are all high. Among them, Germany ranks among the best in the past few years from 2007 to 2018. Followed. On the whole, the logistics

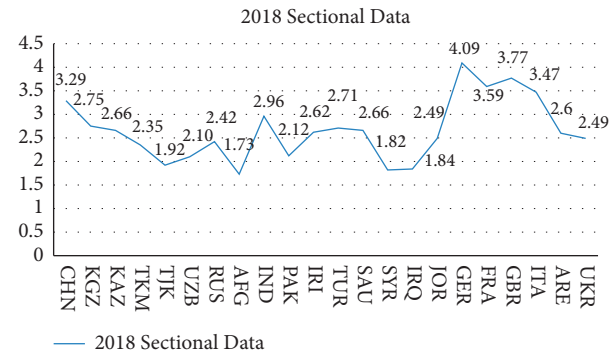


FIGURE 1: 2018 logistics performance index of major belt and road countries: composite score (1 = very low to 5 = very high).

development of the key countries of the “Belt and Road” has the characteristics of “the central part is poor, and the east and west are better”.

3.3.2. Spatial Econometric Model. In this section, we will discuss the relationship between countries in the Belt and Road by using spatial econometrics and calculating the univariate Moran index by GeoDa software. Since GeoDa can only handle cross-sectional data, we only choose data for 2018. Table 2 shows the calculation results of each variable from X_1 – X_6 with Y . It is evident from the results that the Moran’s I of each variable is positive indicating that there is spatial autocorrelation between the independent variables and the dependent variable, and a positive correlation, indicating that objects with similar attributes are clustered together. Among them, the strongest correlation with the composite score is the ease of arranging competitively priced shipments.

Based on the calculation results in Table 3, LISA clustering maps were generated, as shown in Figure 5. Since the graphical results of the clustering maps generated by X_1 – X_6 are almost the same, we have selected one of the six maps for display.

The LISA clustering map visualizes the local spatial autocorrelation and represents the impact of the “Belt and Road” on the countries along the “Belt and Road”. In the clustering map, the red area indicates High-High clustering; the blue area indicates Low-High clustering; the white area indicates insignificant spatial correlation; and the light gray area represents that the location has no neighbors among the 22 countries selected in this paper. The rest of the dark gray parts of the world map, which belong to the regions outside the research object of this paper, are not imported data.

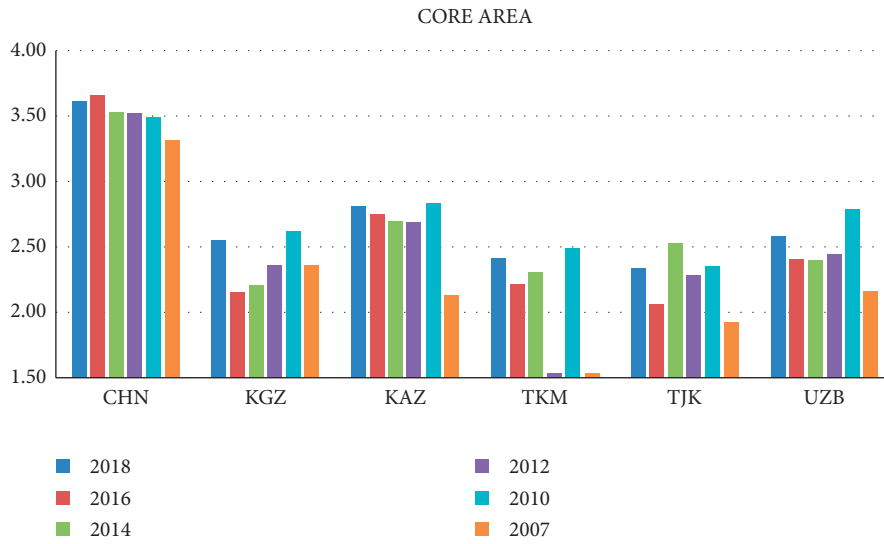


FIGURE 2: 2007–2018 logistics performance index of core area: composite score (1 = very low to 5 = very high).

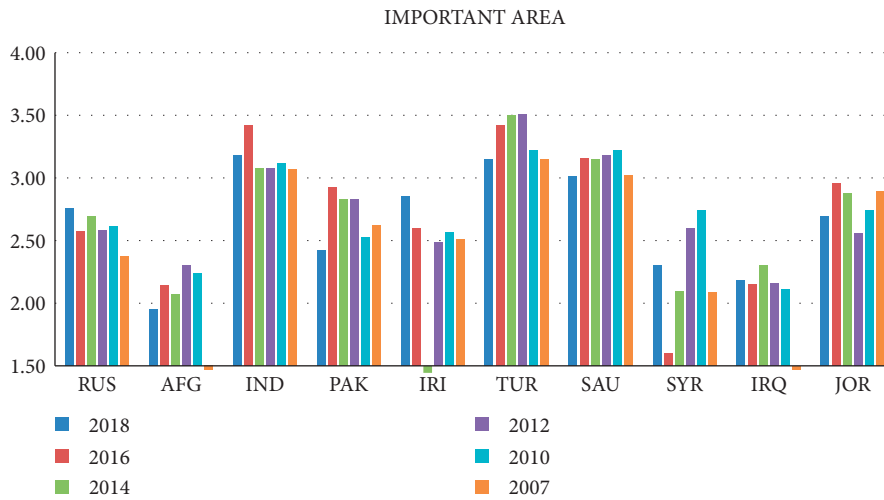


FIGURE 3: 2007–2018 logistics performance index of important area: composite score (1 = very low to 5 = very high).

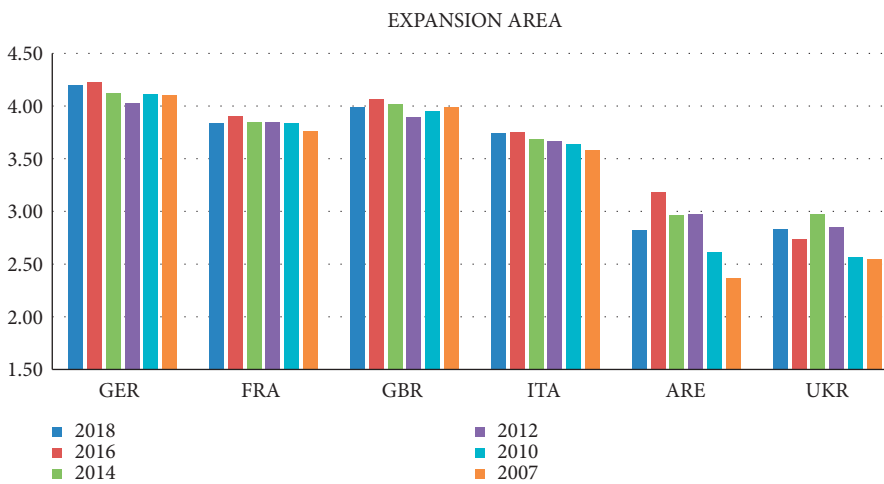


FIGURE 4: 2007–2018 logistics performance index of expansion area: composite score (1 = very low to 5 = very high).

TABLE 3: Moran's I calculation results.

	X1	X2	X3	X4	X5	X6
Moran's I	0.486	0.495	0.510	0.530	0.536	0.487

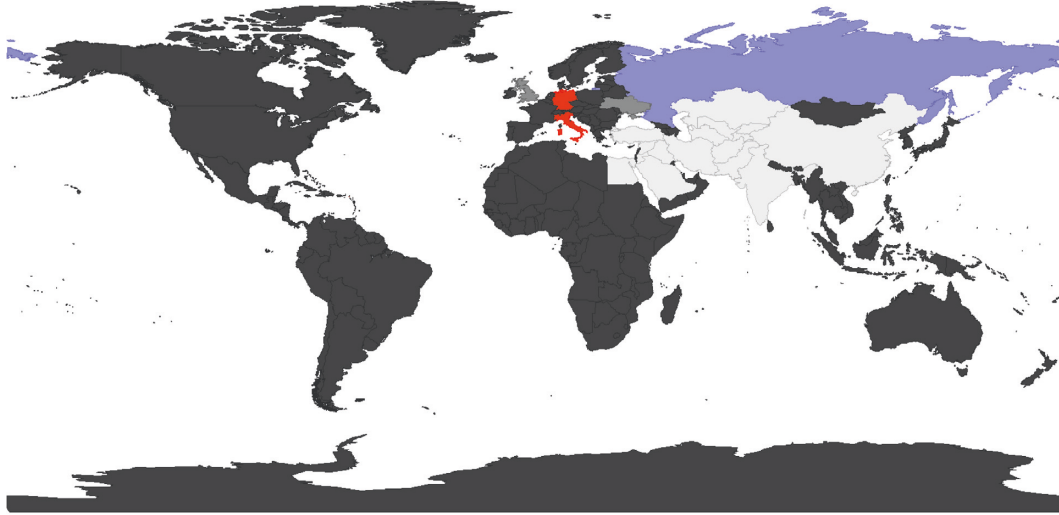


FIGURE 5: Moran cluster map between X1–X6 and Y.

In the figure, the regions of Germany and Italy, show High–High aggregation. This indicates that when these two regions have high LPI, the surrounding regions have correspondingly high scores. Russia, on the other hand, shows a Low–High aggregation, indicating that when Russia has a low logistics performance score, the surrounding regions have a high score instead. It can be seen that only individual countries show a more obvious spatial correlation in LPI changes, and the geographical connectivity among most countries along the Belt and Road has yet to be enhanced.

Then further spatial regression analysis was conducted in this paper by Spatial Error Model (SEM) and Spatial Lag Model (SLM), and the results are shown in Table 4. It is obvious that the numerical results of the two model runs are very similar, and there is an extremely strong explanatory power for both dependent variables ($1 > 0.999978 > 0.999977$). Also, the combination of LogL value ($93.34 > 92.90$), AIC value ($0 > -170.683 > -171.803$), SC value ($-162.327 > -164.492$), and p value of LP test ($0.5837 > 0.2773$) shows that the Spatial Lag Model works better if the optimal spatial regression model is to be built. Since the focus of this paper is to explore the efficiency of logistics through MaxDEA, we will not expand on it here.

3.3.3. Bootstrap DEA Model Analysis. The comprehensive score index (y_1) and six subdivisions ($X_i, i = 1, \dots, 6$) of the logistics performance of major nations along the Belt and Road in 2018 were extracted independently and entered into Max DEA for analysis.

In this article, the MAX DEA software is used to operate bootstrap DEA. First enter the original data, X includes 6 inputs $x_1, x_2, x_3, x_4, x_5, x_6$, and Y includes y_1 . Let DUM

denote major economies. Select the envelope model, keep other default options, and select bootstrap for data analysis. The results are shown in Table 5.

The corrected values are re-ranked, and the results are as follows (Table 6).

From the analysis of the bootstrap DEA model, it can be seen that the first selected model, that is, the 2018 LPT report, comprehensively and completely describes and compares the logistics performance of major countries along the Belt and Road. The comprehensive ranking after the correction is more accurate and detailed than the original model, and the difference in comparison is more obvious, which is convenient for our further analysis. It can be seen that China's logistics performance is at a medium to a high level. China has performed well in terms of logistics capacity, international freight, timeliness, and cargo tracking, especially in infrastructure construction, but not outstanding in logistics performance rankings. This study speculates that the possible reasons are that, on the one hand, in terms of customs clearance, there are still problems such as cumbersome turnover procedures and low import and export efficiency; on the other hand, some logistics costs are low, such as low labor costs and high population density.

Compared with other countries, Germany has won the championship, followed by the United Kingdom, but the customs and infrastructure of the five Central Asian countries are relatively backward, and Uzbekistan's cargo tracking ability is medium, which needs to be further improved. Among them, Kazakhstan and other countries have a better delivery capacity, and Tajikistan performs better in international freight, ranking very high among the five Central Asian countries. At the same time, we can also see that the performance of several countries in the important

TABLE 4: Results of spatial regression model analysis.

	R-squared	Log-likelihood	Akaike info criterion	Schwarz criterion	Likelihood ratio test prob
SEM	0.999977	92.901735	-171.803	-164.492	0.58370
SLM	0.999978	93.3417	-170.683	-162.327	0.27730

TABLE 5: Results raw data.

No.	DMU	Score (original)	Bias	Mean	Median	Sd	CI_LowerBound	CI_UpperBound
1	Afghanistan	1	0.0018889	1.0018889	1.0009291	0.0026905	1.0001232	1.0114327
2	Egypt	0.9928662	0.0008557	0.9937219	0.9935008	0.0007915	0.9929146	0.996882
3	Pakistan	1	0.0019365	1.0019365	1.0009292	0.0027918	1.0000848	1.0107582
4	Germany	1	0.0021274	1.0021274	1.0010236	0.002828	1.000059	1.0115292
5	Russia	1	0.0013026	1.0013026	1.0009146	0.0014179	1.0000634	1.0074079
6	France	1	0.001225	1.001225	1.0008382	0.0011593	1.0000513	1.0052738
7	Kazakhstan	1	0.0013736	1.0013736	1.0009893	0.0013802	1.0000678	1.0068713
8	Kyrgyzstan	1	0.0020506	1.0020506	1.000973	0.00278	1.0000731	1.0115743
9	Saudi	1	0.0018772	1.0018772	1.0009005	0.0026984	1.0000672	1.0111135
10	Tajikistan	1	0.0016653	1.0016653	1.0008854	0.0024274	1.0000731	1.0110763
11	Turkey	1	0.0013324	1.0013324	1.0009115	0.0012414	1.0000479	1.0057658
12	Turkmenistan	1	0.0020379	1.0020379	1.0009489	0.0028231	1.0000851	1.0110627
13	Ukraine	1	0.0021894	1.0021894	1.0009496	0.0029749	1.000069	1.0120212
14	Uzbekistan	1	0.00194	1.00194	1.0009283	0.0025734	1.0000439	1.0100235
15	Syria	1	0.0022273	1.0022273	1.0009279	0.0029162	1.000093	1.0103861
16	Iraq	1	0.0019241	1.0019241	1.0009189	0.0026837	1.0000542	1.0106874
17	Iran	0.9894613	0.0008688	0.9903301	0.9901487	0.0006542	0.9895422	0.9930757
18	Italy	0.9990447	0.0008299	0.9998746	0.9997653	0.0005505	0.999129	1.0025317
19	India	0.9914746	0.0009293	0.9924039	0.9921703	0.0007987	0.9915093	0.9961749
20	The U.K.	1	0.0013702	1.0013702	1.0009941	0.0012389	1.0001018	1.0058977
21	Jordan	1	0.0020418	1.0020418	1.0009684	0.0028026	1.0000804	1.0118942
22	China	1	0.0016956	1.0016956	1.0007966	0.0022515	1.0000791	1.0083734

TABLE 6: Ranking results.

Rank	DMU	Median
1	Germany	1.001023584
2	U.K.	1.000994064
3	France	1.000989313
4	Italy	1.000973002
5	China	1.000968445
6	India	1.00094961
7	Turkey	1.00094893
8	Russia	1.000929174
9	Kazakhstan	1.00092911
10	Kyrgyzstan	1.000928327
11	Jordan	1.000927938
12	Ukraine	1.000918908
13	Turkmenistan	1.000914579
14	Pakistan	1.000911534
15	Afghanistan	1.000900459
16	Uzbekistan	1.000885444
17	Syria	1.000838196
18	Iraq	1.000796621
19	Saudi	0.999765322
20	Tajikistan	0.993500798
21	Egypt	0.992170308
22	Iran	0.990148732

regions of the Central Asian Economic Belt is quite different. In the context of “One Belt, One Road”, in line with the principle of complementary advantages, mutual benefit, and win-win, countries should conduct in-depth exchanges to jointly promote the improvement of logistics levels.

4. Conclusions

By comparing the logistics performance indexes of various countries in the Belt and Road, this paper explores the main factors affecting the comprehensive performance of logistics, introduces the Moran index, and uses the spatial econometric model to explore the geographical relationship of the subdivision indicators of the logistics performance index. The country’s logistics performance index is compared.

From the visualization analysis and bootstrap DEA model analysis, it can be seen that China’s logistics is at a medium level, and its scores are mainly in the efficiency of customs clearance procedures and the quality of trade and transportation-related infrastructure. So far, there are still large logistics vacancies along the Belt and Road. On the Silk Road Economic Belt, European countries have relatively high incomes and logistics performance indices, whereas other countries and regions, such as the five Central Asian countries, have lower national logistics performance indices, with their highest levels of logistics performance at the bottom. The data shows that in 2014, compared with Afghanistan, the best-performing country, Afghanistan was 2.05 points worse than Germany, and Afghanistan ranked 158 in that year. In 2012, the logistics performance of various countries was 141, and the corresponding score was 141.1.87. China, Turkey, and India in the economic belt are middle- and upper-income countries, and their logistics performance is correspondingly high. This “logistics vacancy” is not conducive to the overall development of the economic belt.

However, the relationship between regional income level and logistics performance is less well understood than the relationship between physical location and logistical performance. Overall, countries with higher incomes also have higher logistics performance, but looking at the relative performance of incomes aside from absolute incomes, there is no evident link between earnings and performance improvement. Taking our country as an example, in 2014, our country's per capita income ranked 94th in the world, at US\$7,476, but our country's global ranking on the logistics performance index was 28, which is an example of lower-income and higher logistics performance ranking. Moreover, for countries with extremely uneven regional distribution, such as our neighbor Russia, there are large intra-country differences, and the LPI score cannot well reflect the country's logistics development level.

Here, this study makes the following recommendations:

- (1) It is necessary to improve the customs clearance mechanism and customs clearance coordination mechanism of the countries along the route, establish the "Belt and Road" free trade zone, improve customs efficiency, and reduce the difference in customs clearance efficiency. Accelerate the improvement of the customs clearance coordination mechanism between China and other countries along the Belt and Road, solve the problems of low customs clearance efficiency and low timeliness of logistics services, make a trade in services and goods more convenient, and promote the facilitation and efficiency of economic and trade exchanges between countries. At the same time, it is necessary to strengthen the infrastructure construction cooperation with the countries along the "Belt and Road" to communicate with each other, speed up the improvement of logistics infrastructure, maintain and upgrade existing equipment, and improve the operating efficiency of existing facilities, increase the net throughput of ports, and reduce construction development of new technical equipment.
- (2) Strengthen the development of regional linkage of logistics system among Belt and Road countries. The above spatial econometric analysis results show that the main countries of the Belt and Road have weak spatial correlations, while the Belt and Road aim to carry out wider, higher-level, and deeper regional cooperation, and advocates the interconnection of countries. Territorial connectivity is essential. As a logistics industry with prominent geographical elements, it plays an indispensable role in enhancing spatial correlation. In the future, China should strengthen trade links with countries along the route with high levels of economic development, a high degree of openness, and strong radiation capabilities. It is conducive to driving the liquidity of the entire "Belt and Road" trade, improving the level of logistics, and enhancing the regional connectivity between countries.

- (3) Improve the efficiency of infrastructure development and customs passage. The five Central Asian countries are the only way for the Silk Road in terms of geographical conditions, but it is precise because of their disadvantageous geographical conditions that their geographical conditions are poor and their resource distribution is extremely uneven, resulting in its Economic development lagging behind other coastal countries and regions with rich products and superior natural conditions. In 2014, only Kazakhstan's logistics performance ranked among the top 100 in the world, ranking 88th. Only in 2010 did the five Central Asian countries have their best year in the logistics performance index, and they all joined the world's top 100 for the first and only time. The five Central Asian countries are relatively backward in all aspects of logistics. Infrastructure construction and customs clearance efficiency must both be enhanced, in the end [6].
- (4) Develop green ecological logistics. "Lucid lakes and beautiful mountains are priceless assets," says the logistics business, which also requires green growth to have a low environmental impact. Ecological issues are not just national or regional issue; they are also a global issue. Many nations will respond positively to the request to develop a green logistics ecological chain, as worldwide attention to environmental issues has increased in recent years. Active participation is required. To limit the negative effect on the environment, we must implement ecological civilization in all parts of logistics, from the procurement of raw materials to delivery to clients, and aim to reduce resource consumption, carbon emissions, and recycling as much as feasible [12]. The Green Silk Road's economic logistics are collaboratively constructed and shared by all countries.

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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was financially supported by Guangzhou-Huashang College 2021 Key Discipline Project, School-Level Key Discipline-International Business (project number: 2021HSXK05), Guangzhou Huashang College Key Discipline Project - Applied Statistics (Project Number: 2021HSXK03), Guangdong Planning office of philosophy and Social Sciences Project (Youth): Research on Cross - border social responsibility of private foreign trade

enterprises in Guangdong and the reconstruction of legitimacy -- Perspective of the Organization to piece together. (Project Number: GD20YGL09), School-Level Key Discipline-Business Management (project number: 2021HSXK10), and Special Innovation Project of Guangdong University of Foreign Studies (project number: 299-GK20GS43).

References

- [1] X. Jiang, "Development and suggestions of my country's logistics industry," *Hebei Enterprise*, vol. 12, pp. 88-89, 2020.
- [2] X. Tang, "A study on the logistics development strategy of Tibet under the construction of the south asian great corridor and the integration of the silk road economic belt," *Tibet Science and Technology*, vol. 10, pp. 17-20, 2015.
- [3] T. Zhang and S. Qiao, "Research on the influencing factors of the "belt and road" international trade-an empirical test based on trade agreements and logistics performance," *Journal of Social Sciences*, vol. 1, pp. 63-71, 2020.
- [4] C. Wang, Y. Gao, and C. Liu, "The current situation of research on logistic performance evaluation and the trend of that," *China Business and Market*, vol. 31, no. 3, pp. 16-24, 2017.
- [5] Z. Liu, Li Xin, and F. Cao, "The influences of logistics performance of 21st century maritime silk road on the export of Chinese mechanical and electrical products," *East China Economic Management*, vol. 32, no. 11, pp. 52-59, 2018.
- [6] L. Simar and P. W. Wilson, "A general methodology for bootstrapping in non-parametric Frontier models," *Journal of Applied Statistics*, vol. 27, no. 6, pp. 779-802, 2000.
- [7] J.-F. Arvis, *World Bank Logistics Performance Index Report 2010 Linking to Competition Trade Logistics in the Global Economy*, China Fortune Press, China, 2010.
- [8] J.-F. Arvis, *World Bank Logistics Performance Index Report 2012 Linking to Competition Trade Logistics in the Global Economy*, China Fortune Press, China, 2012.
- [9] J.-F. Arvis, *World Bank Logistics Performance Index Report 2014 Linking to Competition Trade Logistics in the Global Economy*, China Fortune Press, China, 2014.
- [10] J.-F. Arvis, *World Bank Logistics Performance Index Report 2016 Linking to Competition Trade Logistics in the Global Economy*, China Fortune Press, China, 2016.
- [11] J.-F. Arvis, *World Bank Logistics Performance Index Report 2018 Linking to Competitive Trade Logistics in the Global Economy*, China Fortune Press, China, 2018.
- [12] M. R. Chernick, *Bootstrap Methods: A Guide for Practitioners and Researchers*, Wiley-Intercedence, Hoboken, 2nd edition, 2008.

Research Article

Font Design Optimization Model for New Media Short Video Based on Virtual Reality Digital Processing Technology

Hongwei Hu ¹ and Wenyao Zhu ²

¹Graduate School of Design, Chosun University, Gwangju 61452, Republic of Korea

²College of Engineering, Lishui University, Lishui, Zhejiang 323000, China

Correspondence should be addressed to Wenyao Zhu; zwy@lsu.edu.cn

Received 19 July 2022; Revised 8 August 2022; Accepted 10 August 2022; Published 5 September 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Hongwei Hu and Wenyao 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.

Time is not a static idea, but neither is the evolution of the media and design industries. Our creative tools and the media have seen a significant transformation in the last 20 years. Digital technology will shape the media and design industries in the future. Until the next major technological revolution, digitalization will have a lasting effect on the media and design industries. The development and liberation of many designers' ideas and perspectives thanks to modern digital processing technology for virtual reality has sparked an unprecedented "boom" in design. People's senses of sight, sound, and touch will be completely satisfied thanks to the incorporation of such technologies in the design process. A vast history and rich cultural heritage can be found in the field of font design. It has continuously played a crucial role in the advancement of science and technology. The creation of a new media short video typeface based on digital processing technology for virtual reality is suggested in this study. After mastering the font style, the new media short video font is extracted using virtual reality digital processing technology, and the identification system is built utilising virtual three-dimensional technology. The simulation test and analysis are done last. The proposed approach has an accuracy that is 9.34% greater than the conventional technique, according to simulation findings. This outcome demonstrates in detail how font design becomes more humanized when virtual reality digital processing technology is used. It demonstrates how people and information interact and genuinely stress the importance of human participation and dominance. Ethics and aesthetics are combined in font design. The fashion and aesthetic ideas of the new century are reflected in it like a mirror. As a result, font design is now being pursued in a new way, and its new application concept unquestionably has a significant impact on the design sector today.

1. Introduction

The term "new media short video" refers to a new type of video whose length is measured in seconds, which primarily uses mobile intelligent terminals to achieve quick shooting and beautification editing and which can be shared in real time and seamlessly connected on social mobile media platforms. It combines text, images, voice, and video to more intuitively and stereoscopically meet users' needs for expression and communication and for displaying and sharing among individuals. It has distinctive characteristics and value. Font design is the foundation of the whole design industry. In the digital era, characters are no longer limited to simple deformation writing or freeze frame in the

conventional traditional mode [1]. No matter how perfect some models are, they have become somewhat tacky and "old-fashioned," which has become a stumbling block in the information age and does not adapt to the development of society. Interactive font design is an art of communication. This kind of communication is the communication between creators and viewers, producers and browsers, and browsers and other browsers [2]. It is this feature of two-way communication that makes the original font design works be reinterpreted, reconfigured, and "translated," so this kind of work may become richer, more detailed, and more interesting. The injection of new ideas is conducive to the establishment of a multi-perspective, diversified, and new font design mode. Font design is an essential design element in

visual communication design. It is not only the carrier of language but also the graphic carrier of image information. It is the most important, personalized, dynamic, and unique design element among many design elements in modern visual communication design. However, it is a part of modern design as a whole, which is closely linked and complementary to other design elements [3]. Font design must obey the form of design platform, conform to the content of design creativity, and meet the needs of layout.

Short videos have a lot of value and function and are used in all spheres of life, especially as the new media development trend picks up steam. It can spread knowledge and culture, record the details of daily life, and aid in product promotion and drainage. It can even improve brand communication, help with brand crisis public relations, support agriculture and efforts to reduce poverty, and promote the brand. In the age of new media art, 3D virtual imaging technology [4, 5] has replaced the traditional physical display, creating many unique and nonexistent objects [6]. To provide people with a range of sensory experiences in a single arrow, it can simultaneously change different spatial forms in a constrained space, which significantly conserves space, materials, transportation, and other material resources [7]. One of the most promising technologies in many fields is virtual reality digital processing technology [8]. A computer simulation system called virtual reality digital processing technology allows users to both create and experience virtual worlds. It creates a simulation environment using computers. Users can become fully immersed in the environment thanks to this interactive, three-dimensional dynamic scene and entity behavior simulation system with multisource information fusion. Virtual reality digital processing technology can interactively provide comprehensive information of multiple perceptions, create relatively realistic virtual scenes, and give participants an immersive and real feeling. This is in contrast to the sensory information transmitted by traditional print media. This study adopts the three-dimensional visualisation technology in virtual reality digital processing technology to reduce the execution cost of the algorithm in light of the benefits of this technology. The practice has shown that this combination can increase the effectiveness and efficiency of new media short video font design while also cutting down on calculation time.

In the age of new media, short videos are well-liked by online users. Its emergence serves as a helpful complement to social media's current primary content providers and signifies a new use of postmodern aesthetics [9]. The blurred line between art and non-art is one way that short videos differ from the familiar movies, TV dramas, and short videos. Short videos have become more common as a result of the growth of mobile Internet and their ability to be shared at any time and from any location. It also satisfies the demands of mass consumption while having a distinct aesthetic meaning of its own. People are increasingly inclined to watch videos through mobile clients thanks to the development of the 5G network and the widespread use of mobile devices. The core of the relationship between words and media is the typesetting design of words. Although

information is usually used to actually "express" some meanings, it can also be modified and strengthened by the appearance of fonts, and it is these words that form the discourse to be "expressed." Even if there is no explicit meaning, words can also be used to be constructed into a pure aesthetic demonstration, because they will form an intuitive and beautiful effect visually. This study establishes a visual feature reconstruction model of new media short video font design image, explores the key variables of font design, and extracts the fuzzy feature [10] of font design optimization design image. Its innovation lies in the following:

- (1) For the purpose of lowering the cost of the algorithm's execution, this study uses a three-dimensional visualisation technique in virtual reality digital processing technology.
- (2) This study constructs the key feature quantity of the new media short video font optimization design image and uses the establishment of a database to realize the optimization design and optimization recognition of the new media short video variables.

This study studies the optimization design of new media short video fonts.

2. Related Work

The dissemination of written information is more personalized, fashionable, and thorough in the age of new media. People also pay attention to participation and immersion and anticipate turning passive acceptance into active participation [11]. The research on dynamic font design has only recently become available, and font design has not received the proper attention in design research that incorporates digital media. Few monographs exist; the majority are papers. New design technologies and specifications have been made possible by the development of digital media. The digital media platform has undergone a significant change as a result of font design [12].

Parsons and others believe that the history of human civilization must be a history of the development and creation of media, and the creation and promotion of a media often breed a new culture or civilization. As he pointed out, we shape tools, and the remarkable sign of tools shaping our postmodern society is the dominant civilization of electronic media and digital media. Therefore, the research on font design on digital media platform has certain value and prospects [13]. Leng et al. pointed out that the new font design rekindled the desire to read in the digital age. Their design gave new life to the text, and their interaction with the reader's vision and emotion was reflected between the lines [14]. The American company founded by Jiang et al. specializes in the production of film titles, including a large number of excellent font design works. The font design in the title of *Shrek* is an excellent case. In the combination of font and modern technology, Western font design is more representative, and the development state is also tending to be mature, while Chinese font design does not have its own complete system in the combination with modern

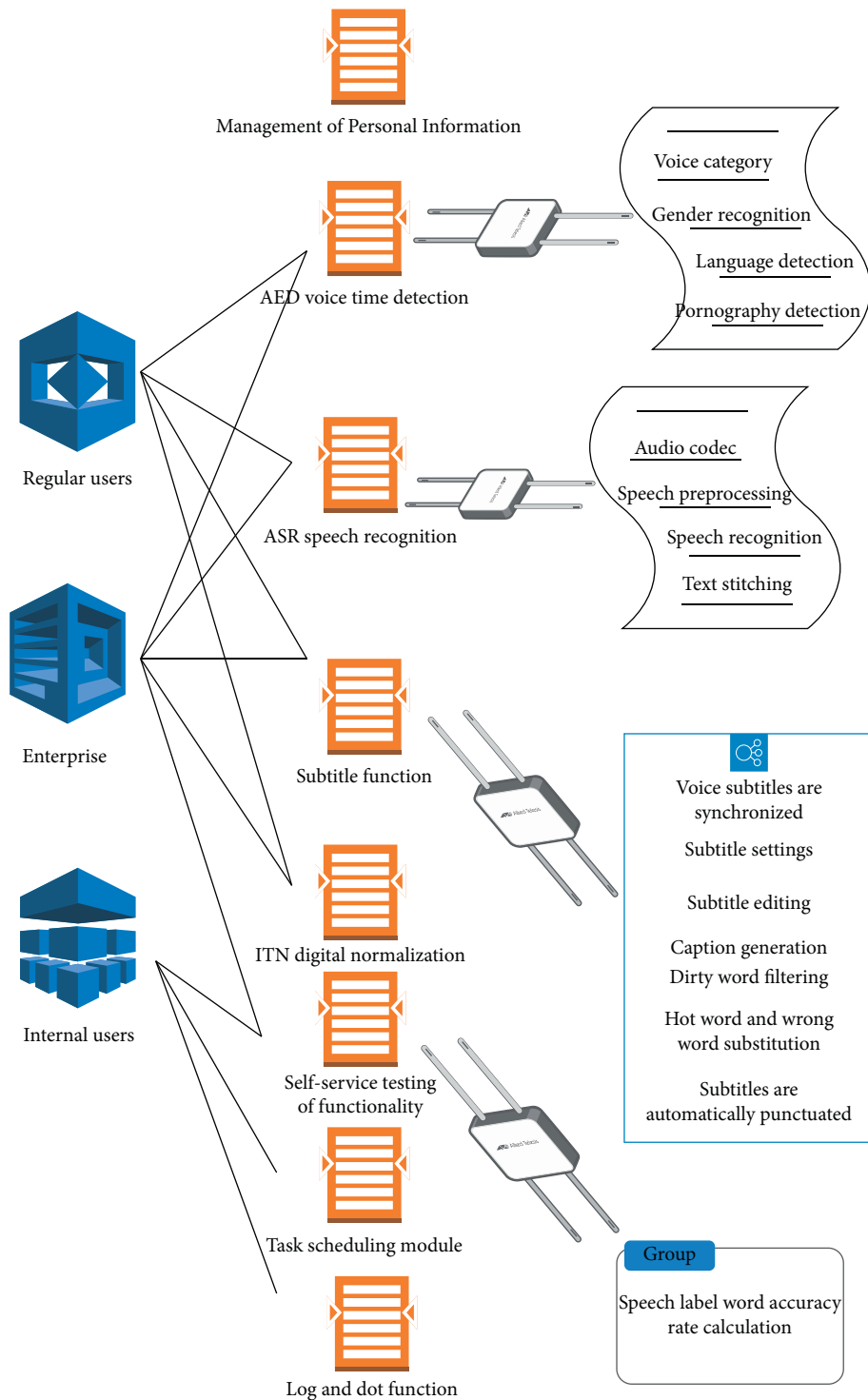


FIGURE 1: Use case diagram of automatic caption generation system.

technology, so it needs more research and practice to explore the new development of font design [15]. Parsons and others mainly started with the analysis of visual elements, discussed the characteristics of various elements of font design in the new era, and discussed the development trend of visual form of font design from two aspects of semiotics and new sense of order [13]. Liu and others mainly analyzed the basic personality and communication characteristics of font design in

the post-digital era, made a detailed explanation on the pen shape, structure, image, and other aspects of font design, and discussed the interactive performance from the perspective of public art as “the blurring of the boundary between entertainment and exhibition; the integration of education and leisure; the integration of information acquisition and aesthetic experience, entertainment experience, and so on” [16]. Yang et al. did in-depth research and elaboration on

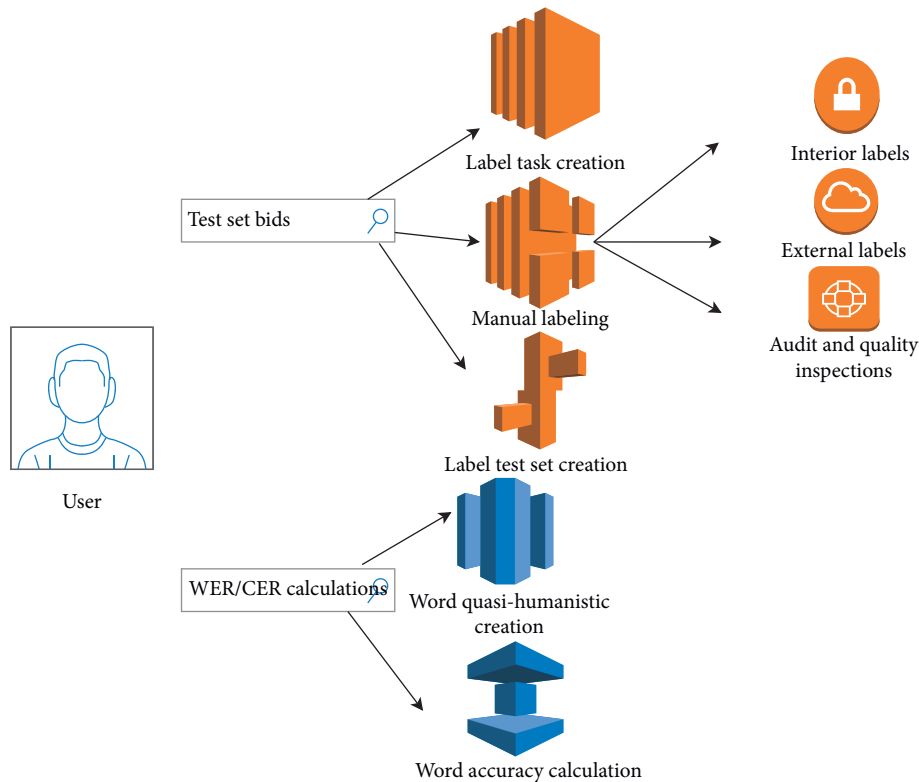


FIGURE 2: Voice self-test function use case diagram.

interaction, but they did not talk about how font design should change under the interactive characteristics of new media and only simply study the characteristics of new media [17]. Dang and Peng studied the information presentation mode and image characteristics with the screen as the media and proposed the creative thinking path of dynamic word effect [18]. Tutschek and others mainly combed the changes in Chinese characters in detail and summarized the development law of font shape design based on the character formation characteristics of Chinese characters. Zhang Jing mentioned that font design should rely on interaction and make an in-depth study of interaction, mainly discussing the development of dynamic character effect, which has a certain breakthrough in font interaction design compared with other papers [19]. Mouzaki and others believe that the direction of new media art will be the combination of “dry” silicon crystal computer technology and “wet” biology related to life systems. This new media art is called “wet media.” Microscopic biological phenomena are becoming a new channel for transmitting information and a new direction for us to expand our communication space [20]. For visually conveyed works that are disseminated in any way, type design is both a basic factor and a reasonable starting point. Type design, like the message itself, is a living entity. Based on the digital processing technology of virtual reality, this study establishes a new media short video font design system, finds the common factor of font design, explores the key variables of font design, and extracts the amount of fuzzy features of font design optimization design image.

3. Methodology

3.1. Automatic Caption Generation System for Font Recognition. The automatic caption generation system mainly includes AED voice event function, ASR voice recognition function, log and dot function, caption function, timed task function, voice self-test function, API gateway function (a single point of entry between the client and the API, which acts as a reverse proxy to route client requests to the subsequent set of APIs), ITN digital normalization function, and task scheduling function [21]. Figure 1 is the overall functional use case diagram 1 of the caption generation system.

The voice self-test function is mainly to solve the problem of difficult voice testing. A part of the short video is extracted from the random algorithm as the test. First, the small flow test is carried out, and then, the full flow test is carried out. If the small flow test results are not aligned with the expected results, the cause is checked, and the full flow is carried out after alignment. Its use case diagram is shown in Figure 2.

Voice self-test functions mainly include the following:

- (1) **Test Set Bid Submission:** the main purpose is to extract a part of the test set from some short video business parties using a random algorithm and then publish the test set to the annotation platform, so that the annotation personnel can annotate it. After the annotation is completed, the annotation test set is saved, and then, the annotation test set is uploaded to the cloud.

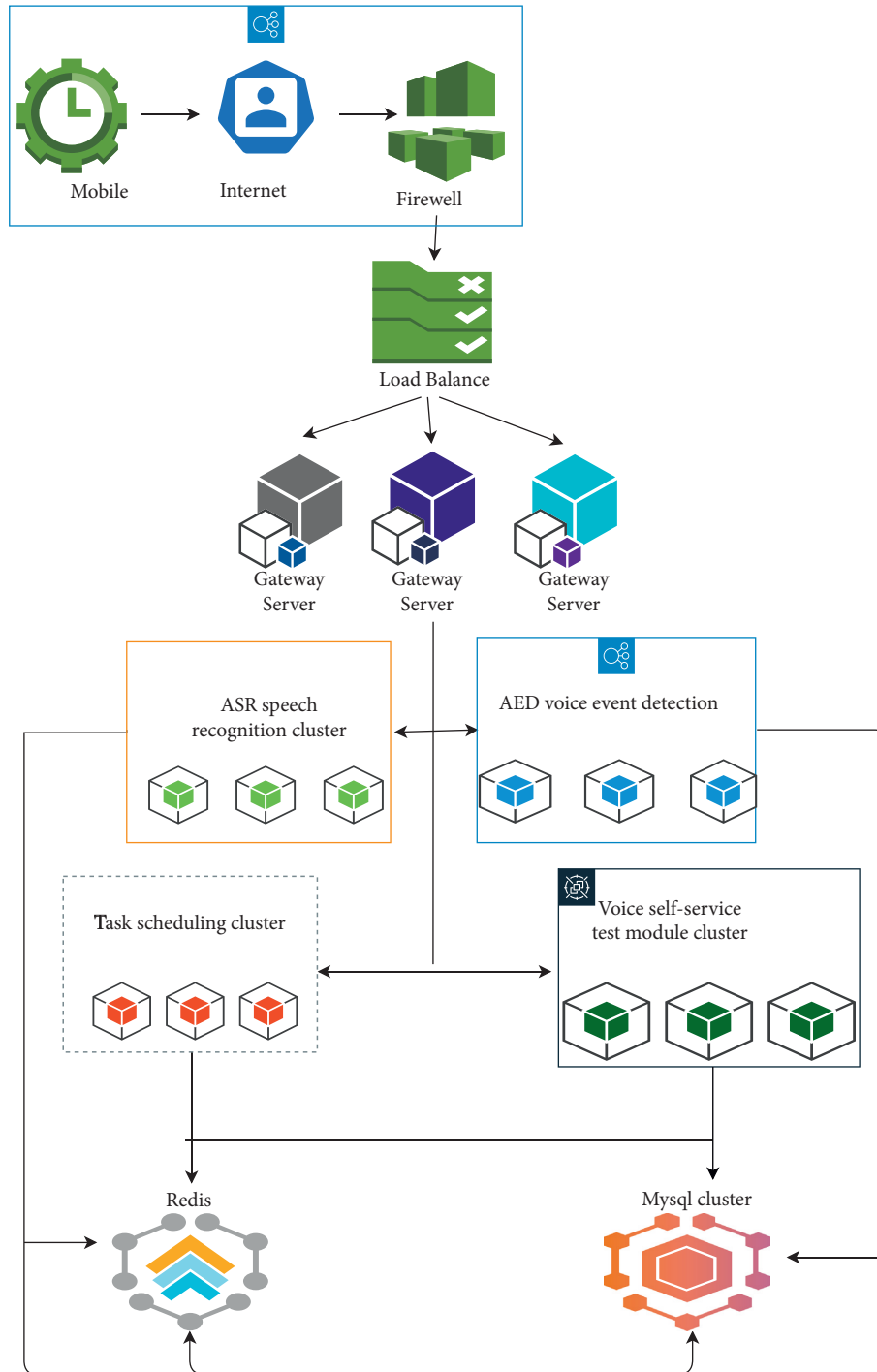


FIGURE 3: General system architecture.

(2) WER/CER Calculation: the word error rate and sentence error rate are calculated. The main purpose is to evaluate the quality of the algorithm and whether it can be aligned with the online and offline environment model.

According to the subtitle request process from the Web end or mobile end, the overall architecture of the system can be obtained by analyzing the system request process. The overall frame composition of the system is shown in Figure 3.

According to the general architecture of the system, it can be found that users send requests through mobile phones or Web terminals, and then, these requests resolve the requested IP through DNS (it is a server that translates IP addresses into their corresponding domain names) domain names through the Internet. After passing through the firewall, the firewall intercepts some illegal requests. Then, the corresponding API gateway through the load balancer is found, and then, the corresponding cluster through API

gateway and Consul (is a service management software. Distributed high availability, service discovery, and configuration sharing in multiple data centers are supported. Raft algorithm is adopted to ensure high availability of services) is found, the corresponding interface is requested, and the result is returned. The clusters here mainly include ASR speech recognition cluster, AED speech event detection cluster, task scheduling cluster, and speech self-service test cluster. Each cluster uses its own Redis cache (to put it another way, the remote dictionary service, an open source log type, key value database written in ANSI C language, supports the network, can be based on memory and can be persistent, and offers APIs in various languages), the corresponding MySQL data storage, and the corresponding log management and dotting processing.

3.2. Font Processing and Reorganization Design Based on 3D Visualisation Technology. Some basic concepts, such as volume data and voxels, are used in 3D visualisation (it is a means of describing and understanding models and a representation of data bodies, not simulation technology) in the field of new media. The following is a description of these basic concepts: volume data can be defined as a discrete bounded sampling function in three-dimensional space. If the sampling is regular and structured, this kind of volume data is called regular and structured. In most cases, the volume data are regular and structured; that is, the sampling is uniform in three directions of space. This kind of data is also called three-dimensional discrete image, which can be expressed as follows:

$$f(x, y, z), \text{ among } \begin{cases} x = x_1, x_2, \dots, x_i; (x_i - x_{i-1} = \Delta x), \\ y = y_1, y_2, \dots, y_m; (y_i - y_{i-1} = \Delta y), \\ z = z_1, z_2, \dots, z_n; (z_i - z_{i-1} = \Delta z). \end{cases} \quad (1)$$

Sampling point and sampling value sampling point are the spatial location of sampling. The sampling value is the quantitative value of a physical attribute of a substance at the sampling point. A voxel is defined as a small box area. Its length, width, and height are the sampling spacing in three directions, respectively. Voxels can be regarded as containing only the same substance, and their physical properties are the same as the sampling value. It can also be regarded as filled with nonuniform material, and the physical property change in the internal material is determined by the trilinear interpolation of the sampling values at its eight corners. The latter model is more accurate and reasonable than the former one in describing the distribution of physical characteristics in the body domain. The sampling space, which is made up of many voxels, is represented by volume space (also known as volume domain). If the sampling distribution in volume space is structured, then the data are said to be structured on a regular basis. This refers to the relationship between spatial data that complies

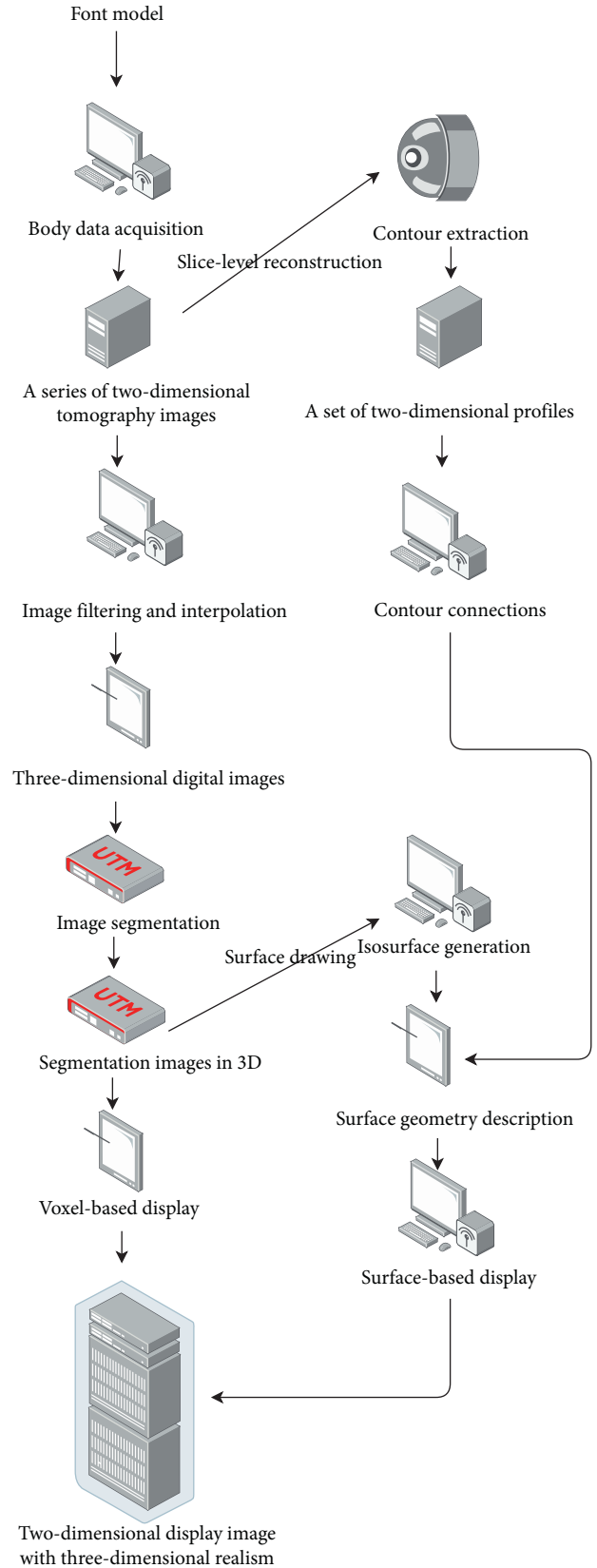


FIGURE 4: Three-dimensional visualisation flow chart.

with the rules of topological geometry. There is a distinct topological adjacency between sampling points, i.e., the adjacency, association, inclusion, and connectivity between entities represented by nodes, arcs, and polygons. The volume data are referred to as regular if the distribution of samples in one direction of volume space is equidistant. These fundamental ideas are necessary for the two 3D visualisation techniques that will be discussed later. The short video font image data must also be segmented and labelled before the necessary rendering work is carried out using the 3D visualisation method. Let us now discuss font image segmentation and annotation preprocessing. The flow chart of 3D visualisation is shown in Figure 4.

An indicator used to gauge how similar different fonts are is structural similarity. To analyze the differences between fonts more thoroughly, it makes the font structure independent of brightness and contrast and combines the three. Following is a definition of how structurally similar the two fonts are:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (2)$$

Among them, x and y are two different images, respectively, μ_x and μ_y are the mean of x and y , respectively, σ_x^2 and σ_y^2 are the variance of x and y , respectively, and σ_{xy}^2 is the covariance of x and y . $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$ are constants used to maintain stability. L is the dynamic range of pixel values. $k_1 = 0.01$, and $k_2 = 0.03$. The range of structural similarity is 0 to 1.

Font distortion is frequently assessed using the peak signal-to-noise ratio, which is an engineering term for the ratio of a signal's maximum possible power to the destructive noise power that affects the accuracy of its representation. Although its evaluation results are frequently at odds with the subjective perception of human eyes as a result of its mathematical analysis based on pixel grey values, this does not diminish its referentiality as an objective evaluation index. Mean square error, or MSE, is a straightforward way to define PSNR, and its solution formula is as follows:

$$MSE = \frac{1}{mn} \sum_{i=0}^m \sum_{j=0}^n \|I(i, j) - K(i, j)\|^2, \quad (3)$$

$$PSNR = 10 \cdot \log\left(\frac{MAX_I^2}{MSE}\right) = 20 \cdot \log\left(\frac{MAX_I}{\sqrt{MSE}}\right).$$

Among them, $I(i, j)$ represents the pixel value of the predicted font at point (i, j) , $K(i, j)$ represents the pixel value of the original font at that point, m and n represent the size of the font, and MAX_I represents the possible maximum pixel value of the picture. Obviously, the smaller the MSE value, the closer the predicted font is to the original font. That is, the higher the PSNR value, the lower the font distortion. Further, the lower the noise level contained in the predicted font, the better the reconstruction effect.

The network does not calculate the similarity of the whole image, but calculates in blocks. Because the sizes of I^{LR} and I^{Ref} are different, I^{LR} pairs are upsampled by bicubic

interpolation first, so that the sizes of $I^{LR\uparrow}$ and I^{Ref} are the same. At the same time, considering that the fuzzy degree of reference image I^{Ref} is different from that of low-resolution image $I^{LR\uparrow}$ after upsampling, I^{Ref} pairs are downsampled by bicubic interpolation and then upsampled, so that the fuzzy degree of $I^{Ref\downarrow\uparrow}$ is close to that of $I^{LR\uparrow}$. Because texture features are more robust to changes in color and illumination, and even if there is no similarity between I^{Ref} and I^{LR} in RGB or other image spaces, similar blocks may exist in high-level or low-level feature spaces, and the network does not directly calculate the similarity of pixels in image blocks, but calculates the similarity on higher-level feature maps. The inner product is used to measure the similarity between neural texture features:

$$S_{i,j} = \left\langle P_i(\varphi(I^{LR\uparrow})), \frac{P_j(\varphi(I^{Ref\downarrow\uparrow}))}{\|P_j(\varphi(I^{Ref\downarrow\uparrow}))\|} \right\rangle. \quad (4)$$

Among them, $P_i(\cdot)$ represents the i image block sampled from the neural feature map, and $S_{i,j}$ represents the similarity between the i LR image block and the j reference image block. The reference image feature block characteristics are normalized to select the best match on all j . Each reference image feature block is used as a convolution to check all the LR feature blocks for convolution in order to efficiently implement similarity calculation on all the LR feature blocks as a set of convolution (or cross-correlation) operations. Each LR pixel block will eventually have a reference image block with the highest similarity to its corresponding one by comparing the results obtained from multiple cores:

$$S_j = \varphi(I^{LR\uparrow}) * \frac{P(\varphi(I^{Ref\downarrow\uparrow}))}{\|P_j(\varphi(I^{Ref\downarrow\uparrow}))\|}, \quad (5)$$

where S_j is the similarity diagram of the j reference image block, and $*$ represents the correlation operation. $S_j(x, y)$ indicates the similarity between the LR image block centered on position (x, y) and the j reference image block. LR blocks and reference image blocks are both densely sampled from their images. Based on the similarity score, the network constructs an exchange feature map m to represent the texture-enhanced LR image. Each image block in M centered on (x, y) is defined as follows:

$$P_{w(x,y)}(M) = P_j^*(\varphi(I^{Ref})), j^* = \arg \max_j S_j(x, y). \quad (6)$$

4. Result Analysis and Discussion

The development of modern technology has expanded the dimension of font design and paid attention to diversified expression. Digital virtualization is to liberate human thinking activities from the brain and convert them into computer software operations. Human thinking has become a tangible and visible operating system. At this time, thinking is behavior. In virtual thinking and its practical activities, thinking has been behaviorized.

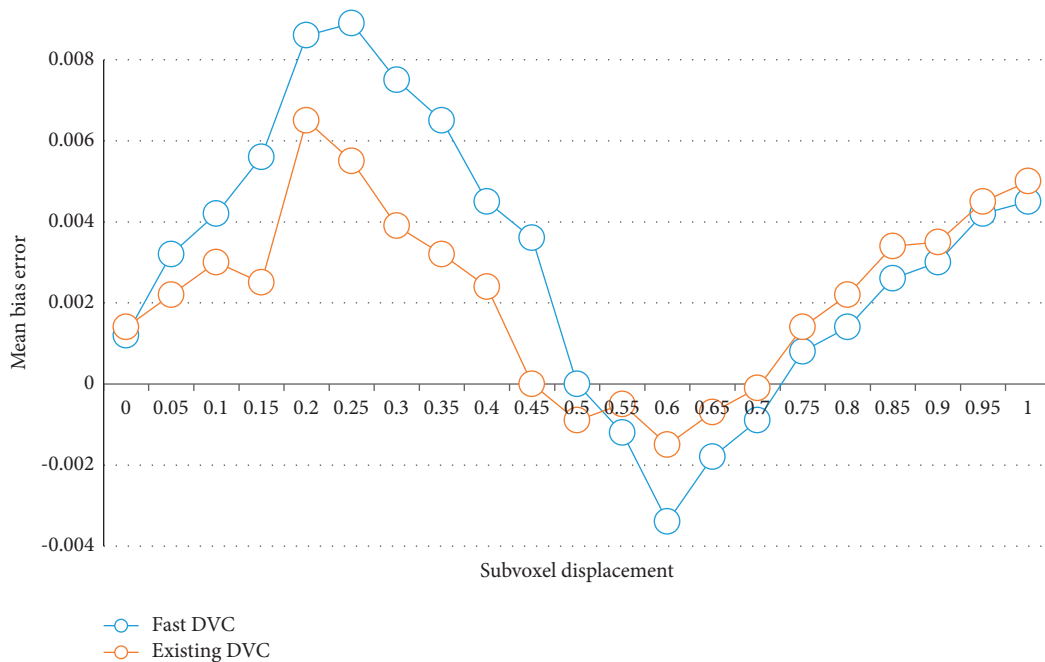


FIGURE 5: Average deviation and standard deviation of displacement calculated by fast DVC method.

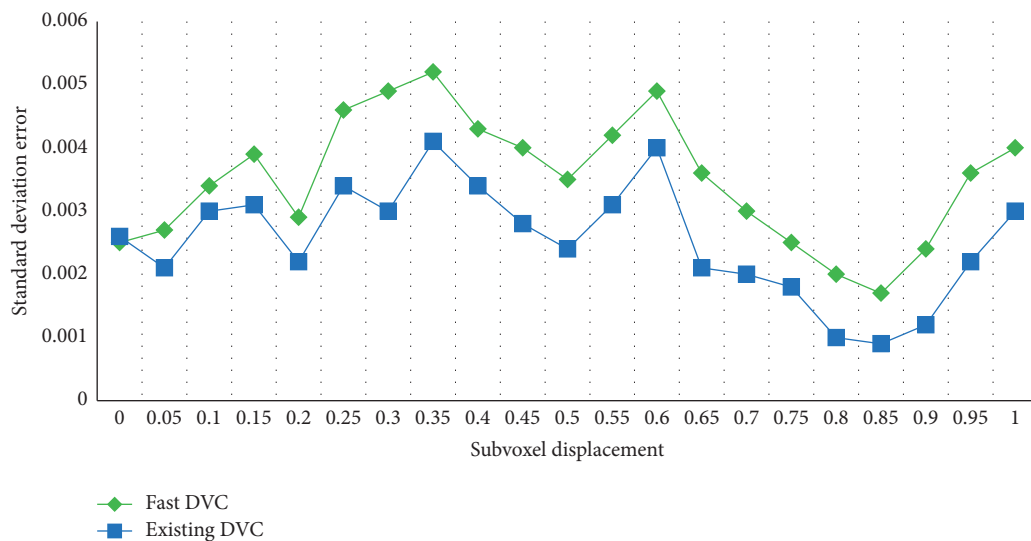


FIGURE 6: Average deviation and standard deviation of displacement calculated by existing DVC method.

In this study, a reliability-oriented displacement tracking strategy combined with an initial value estimation transfer method is applied to IC-GN algorithm to avoid global prime displacement search. This displacement tracking strategy starts with a seed point, guided by the correlation coefficient of each calculation point, and automatically extends to its adjacent calculation points. At the same time, the initial value estimation transfer strategy can automatically transfer accurate and complete deformation initial value estimation to the current calculation point. This calculation strategy reduces the number of iterations of the algorithm and ensures its rapid convergence.

A reference volume image and ten deformed volume images containing Gaussian noise are generated by

computer simulation, and the latter applies 0-1 voxel displacement, respectively. The ten volume images are calculated by the fast DVC method and the existing DVC method, as shown in Figures 5 and 6.

It can be seen that the two methods have similar standard deviations, and the maximum average deviation of the fast DVC method is about half of the existing DVC method. In general, experiments have effectively proved the computational accuracy of the fast DVC method compared with the existing DVC methods. By comparing the calculation efficiency in the above calculation process, the results show that compared with the existing DVC method, the fast DVC method can reduce the number of iterations by 65% and improve the calculation speed by nearly 30–45 times.

TABLE 1: Comparison of recognition results in mixed font dataset.

Algorithm model	Number of test sets	Correctly identify quantity	Number of errors identified	Recognition rate
Tradition DenseNet-201	63131	59346	3875	94.00
ResNet-50	63131	57364	5716	90.87
AlexNet	63131	52164	10942	82.63
GoogLeNetV4	63131	58301	4925	92.35
DPN-92	63131	51423	11364	81.45
Method in this study	63131	60125	2469	95.24

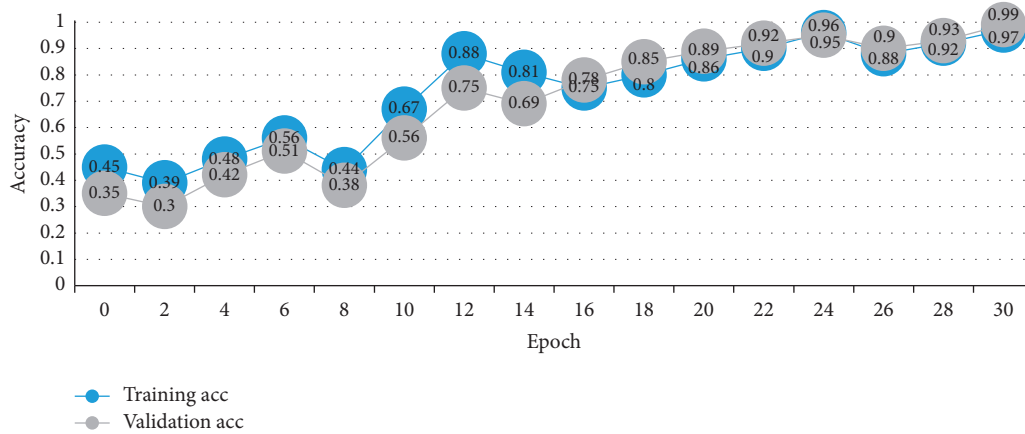


FIGURE 7: Change trend of recognition accuracy.

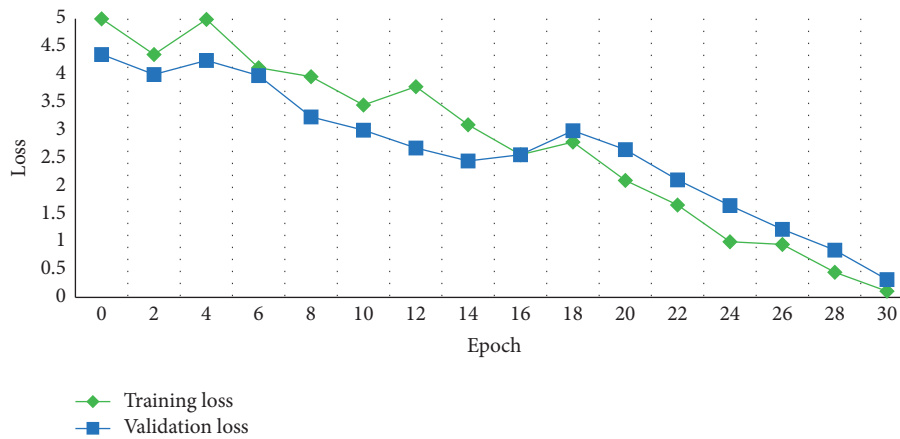


FIGURE 8: Change trend of loss function.

This study trains the regular script, seal script, official script, and running script recognition models to assess the efficiency and dependability of this algorithm. The improved DenseNet-201 algorithm is used to train them by dividing the data set into training set and test set in a 7:3 ratio. The algorithm was put to the test on a test set, and it was compared to five well-known ones: DenseNet-201, ResNet-50, AlexNet, GoogLeNetV4, and DPN-92. The recognition results are shown in Table 1.

From the experimental results in Table 1, it can be seen that compared with the other five algorithms, this method has the highest average recognition rate of 95.24%, which is 1.24%

higher than the traditional DenseNet-201 algorithm. In addition, compared with the experimental results in the classified font dataset, the traditional DenseNet-201, ResNet-50, GoogLeNetV4, and the method in this study have achieved a higher average recognition rate, indicating that with the increase in the dataset, the model can learn more features.

AlexNet’s network structure is relatively simple, so it cannot extract the features of calligraphy images well, and the average recognition rate is only 82.63%. Traditional DenseNet-201, ResNet-50, and GoogLeNetV4 perform well due to their deep network structure, with an average recognition accuracy of more than 90%, as shown in Figures 7

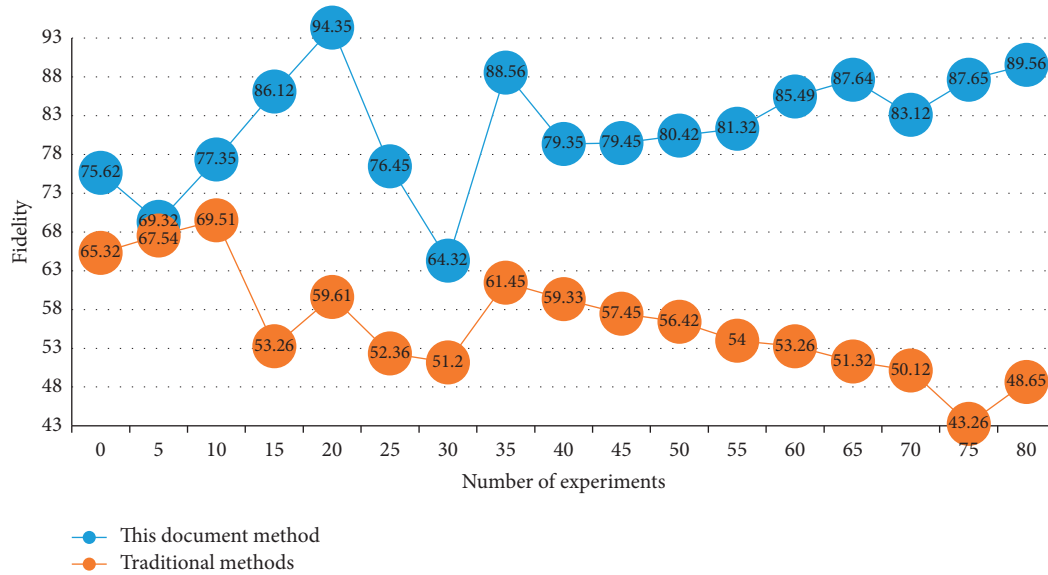


FIGURE 9: Comparison of experimental results.

and 8. The average recognition rate of the improved DenseNet algorithm in this study is significantly higher than that of the other five algorithm models, demonstrating that the improved DenseNet-201 algorithm can achieve higher performance in the task of calligraphy font recognition, even though the accuracy of the clerical script recognition model trained by the traditional DenseNet-201 algorithm is higher than that trained by the improved DenseNet algorithm in this study.

To see the practical application effect of the virtual reality digital processing technology proposed in this study more clearly and concretely, we compare it with the traditional font design method to compare the suitability of its font changes in different scenes. To ensure the accuracy of the experiment, the two font design methods in the same experimental environment are put to test the adaptability of font design in different scenarios. The comparison of experimental results is shown in Figure 9.

It can be seen from Figure 9 that the font adaptation of the digital processing technology method based on virtual reality designed in this study is high, and the initial fidelity is as high as 75%. With the increase in the number of experiments, the fitness increases steadily. When the number of experiments is 80, the fidelity is close to 90%. The initial fidelity of the traditional method is 65%, which is far lower than the design method in this study. With the increase in the number of experiments, the adaptability of font design fluctuates greatly, even lower than 43%. Therefore, the method designed in this study is much higher than the traditional method in font design adaptability and stability and has obvious advantages.

5. Conclusions

In this study, a new media short video font design scheme based on virtual reality digital processing technology is proposed. The new media short video font is extracted using

virtual reality digital processing technology to master the font style, and then, the recognition system is constructed using virtual three-dimensional technology. Finally, the simulation test and analysis are carried out. Simulation results show that the proposed algorithm has certain accuracy, which is 9.34% higher than the traditional algorithm. This result fully shows that the use of virtual reality digital processing technology makes font design more humanized. It provides the interaction between people and information and truly emphasizes people's participation and dominance. Its most distinctive advantage is that it can save time, space, human, financial, and material resources to a great extent, making font design more possible. It makes many wonderful ideas from impossible to possible, which was beyond the reach of relying on physical media display in the past. For font design, virtual reality digital processing technology is still a relatively new technology. In the actual application process, there are still problems such as high cost, great technical difficulty, and the need for a large number of related technical support, especially software development. However, from its development prospects, virtual reality digital processing technology is bound to be an important direction for the development of font design to the high-tech field 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 do not have any possible conflicts of interest.

Acknowledgments

This research was supported by the Science and Technology Public Welfare Project of Zhejiang Province and Research

and Development of BIM-3DGIS Photo Reality Integrated 3D Online Platform and Key Technology Based on WebGL, no. LGG20F020020.

References

- [1] K. Zhong, X. Zhou, and J. Huo, "Digital signal processing for short-reach optical communications: a review of current technologies and future trends[J]," *Journal of Lightwave Technology*, vol. 2018, no. 9, p. 1, 2018.
- [2] H. Shi and D. Niu, "Application research of virtual reality technology in ocean environmental art design," *Journal of Coastal Research*, vol. 104, no. sp1, p. 35, 2020.
- [3] C. M. Lo, J. H. Wang, and H. W. Wang, "Virtual reality human-robot interaction technology acceptance model for learning direct current and alternating current[J]," *The Journal of Supercomputing*, vol. 2022, no. 7, pp. 1-24, 2022.
- [4] S. Qi, X. Ning, G. Yang et al., "Review of multi-view 3D object recognition methods based on deep learning," *Displays*, vol. 69, Article ID 102053, 2021.
- [5] W. Cai, D. Liu, X. Ning, C. Wang, and G. Xie, "Voxel-based three-view hybrid parallel network for 3D object classification," *Displays*, vol. 69, Article ID 102076, 2021.
- [6] L. Tabbaa, C. S. Ang, P. Siriaraya, W. J. She, and H. G. Prigerson, "A reflection on virtual reality design for psychological, cognitive and behavioral interventions: design needs, opportunities and challenges," *International Journal of Human-Computer Interaction*, vol. 37, no. 9, pp. 851-866, 2020.
- [7] E. Segura Ortí and A. García Testal, "Intradialytic virtual reality exercise: increasing physical activity through technology[J]," *Seminars in Dialysis*, vol. 2019, no. 6, p. 67, 2019.
- [8] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7982-7991, Long Beach, CA, USA, June 2019.
- [9] K. Zhong and X. Zhou, "Digital signal processing for short-reach optical communications: A review of current technologies and future trends[J]," *Journal of Lightwave Technology*, vol. 2018, no. 7, p. 65, 2018.
- [10] J. Li, Z. Chen, L. Cheng, and X. Liu, "Energy data generation with Wasserstein deep convolutional generative adversarial networks," *Energy*, vol. 257, Article ID 124694, 2022.
- [11] R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr, and P. Debevec, "A system for acquiring, processing, and rendering panoramic light field stills for virtual reality," *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1-15, 2018.
- [12] A. Yethiraj and P. Jungwirth, "More than virtual reality: important new physical insights in simulations of biomolecules and synthetic polymers," *The Journal of Physical Chemistry B*, vol. 121, no. 26, p. 6294, 2017.
- [13] T. D. Parsons, G. Riva, S. Parsons et al., "Virtual reality in pediatric psychology," *Pediatrics*, vol. 140, no. 2, pp. S86-S91, 2017.
- [14] Y. Leng, C. C. Chen, and Q. Sun, "Energy-efficient video processing for virtual reality[J]," *IEEE Micro*, vol. 2020, no. 99, p. 1, 2020.
- [15] H. Jiang, Z. Lin, Y. Li et al., "Projection optical engine design based on tri-color LEDs and digital light processing technology," *Applied Optics*, vol. 60, no. 23, p. 6971, 2021.
- [16] Y. Liu, F. Castronovo, J. I. Messner, and R. Leicht, "Evaluating the impact of virtual reality on design review meetings," *Journal of Computing in Civil Engineering*, vol. 34, no. 1, 2020.
- [17] Z. Yang, C. Chen, Y. Lin, D. Wang, H. Li, and W. Xu, "Effect of spatial enhancement technology on input through the keyboard in virtual reality environment," *Applied Ergonomics*, vol. 78, no. 8, pp. 164-175, 2019.
- [18] T. Dang and M. Peng, "Joint radio communication, caching and computing design for mobile virtual reality delivery in fog radio access networks[J]," *IEEE Journal on Selected Areas in Communications*, vol. 88, no. 99, p. 1, 2019.
- [19] B. Tutschek and J. Ebert, "OC05.02: virtual reality objects improve learning efficiency and sustained abilities in fetal ultrasound," *Ultrasound in Obstetrics and Gynecology*, vol. 52, no. 8, p. 10, 2018.
- [20] M. Mouzaki and A. T. Trout, "Virtual reality: new insights regarding the prevalence of nonalcoholic fatty liver disease in children and adolescents with obesity using magnetic resonance imaging," *The Journal of Pediatrics*, vol. 207, no. 4, pp. 8-10, 2019.
- [21] F. Cheng, Y. Huang, B. Tanpure, P. Sawalani, L. Cheng, and C. Liu, "Cost-aware job scheduling for cloud instances using deep reinforcement learning," *Cluster Computing*, vol. 25, no. 1, pp. 619-631, 2022.

Research Article

Comparison and Analysis of Several Clustering Algorithms for Pavement Crack Segmentation Guided by Computational Intelligence

Dan Wang ^{1,2,3}, Zaijun Zhang ^{1,2}, Jincheng Zhou,^{2,3,4} Benfei Zhang,^{2,3} and Mingjiang Li^{2,4}

¹School of Mathematics and Statistics, Qiannan Normal University for Nationalities, Duyun 558000, China

²Key Laboratory of Complex Systems and Intelligent Optimization of Guizhou, Duyun 558000, China

³Key Laboratory of Complex Systems and Intelligent Optimization of Qiannan, Duyun 558000, China

⁴School of Computer and Information, Qiannan Normal University for Nationalities, Duyun 558000, China

Correspondence should be addressed to Zaijun Zhang; zzj@sgmtu.edu.cn

Received 6 August 2022; Revised 22 August 2022; Accepted 23 August 2022; Published 3 September 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Dan 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.

Cracks are one of the most common types of imperfections that can be found in concrete pavement, and they have a significant influence on the structural strength. The purpose of this study is to investigate the performance differences of various spatial clustering algorithms for pavement crack segmentation and to provide some reference for the work that is being done to maintain pavement currently. This is done by comparing and analyzing the performance of complex crack photos in different settings. For the purpose of evaluating how well the comparison method works, the indices of evaluation of NMI and RI have been selected. The experiment also includes a detailed analysis and comparison of the noisy photographs. According to the results of the experiments, the segmentation effect of these cluster algorithms is significantly worse after adding Gaussian noise; based on the NMI value, the mean-shift clustering algorithm has the best de-noise effect, whereas the performance of some clustering algorithms significantly decreases after adding noise.

1. Introduction

Pavement crack detection methods that are efficient and reasonable must be used to evaluate the degree of pavement damage, and appropriate maintenance measures must be taken in order to achieve the goal of objectively detecting pavement damage and ensuring the safety, efficiency, and convenience of transportation on pavement [1]. In order to achieve this goal, pavement damage must first be detected. In recent years, there has also been some progress made in the research of pavement crack detection; however, in the actual operation process, manual visual inspection is still used for the purpose of crack detection. Manual detection has been unable to meet the rapidly developing needs of highway maintenance, and the detection speed of cracks has not been able to keep up with the requirements of pavement development [2]. Only relying on manual detection of pavement cracks is not an accurate enough method because it will lead

to issues such as missed detection and false detection, both of which will disrupt traffic and are unable to guarantee personal safety. Even if a large number of people are sent out to inspect and evaluate the pavement of the highway, it will still take a significant amount of time to compile statistics based on the data that has been collected. As a result, it is of the utmost importance to conduct cracks in the pavement in a prompt and efficient manner [3–5].

The most common types of pavement cracks are transverse cracks that run in the same direction as the lane line, longitudinal cracks that run perpendicular to the lane line, block cracks and network cracks generated by transverse and longitudinal staggered, and so on [6]. The low contrast between the crack area and the background, the interference caused by the background, and the uneven damage degree of the cracks are the three factors that make it the most difficult to detect and identify cracks. Inspections were performed manually and visually for the most part in

the traditional method. There will be a great deal of difficulty, as evidenced by a single method of detection, a low detection efficiency, a lengthy cycle, and a high risk. The proliferation of computer technology has led to the development of an increasing number of automatic detection methods. Automatic detection technology has significantly improved in terms of both its efficiency and accuracy in comparison with manual detection, which eliminates the need for human intervention wherever possible. Image processing is typically what is utilized when looking for cracks [7–9].

The advantages of high precision, high efficiency, and strong robustness are all possessed by modern image processing technology, which is analogous to the visual system of the human brain. Information about crack features can be extracted after various features, including the texture and geometry of cracks, which have been analyzed [10]. Image enhancement, feature extraction, and object segmentation are all components of the overarching process that comprises detection methods in image processing [11, 12]. The detection of crack defects using image processing techniques has been the subject of a significant amount of research activity in recent years, both domestically and internationally. Cracks were found by Oliveira et al. by analyzing the light and dark changes of box-borders and the distribution of dark pixels, but this method is susceptible to noise because of its reliance on light and dark changes. In order to achieve crack segmentation, they searched for the valley that was located near the highest peak in the gray histogram as a threshold. In order to improve the limitations of the conditional texture anisotropy (CTA) directional detection, Schmugge et al. proposed a free-form anisotropy (FFA) algorithm to detect pavement cracks. Using this algorithm, the authors sought out the path with the least amount of weight in each direction using an iterative search algorithm. However, due to the high amount of computational work required, this algorithm is not appropriate for use in real-time detection. In order to remove noise from crack images, Yan et al. utilized a variety of fusion and recombination filtering methods. Using this method, it is possible to remove various types of noise; however, it is very easy to lose information regarding cracks. Yang et al. first performed curve fitting on the pixels in the segmented image and then used the SVM model to achieve classification. For SVM training, the authors extracted the grayscale histograms of the local crack area and the background area. This method selected a smaller number of feature parameters, which resulted in a poor recognition accuracy. An automatic algorithm for the classification of cracks was suggested by Lettsome et al. After the crack image has been vectorized, the geometric features are separated, and the linear cracks are classified according to the slope and width. This allows the type and degree of crack damage to be determined, allowing for the accurate and automatic extraction of crack information. Lettsome et al. developed an improved pavement image segmentation algorithm, which extracts the statistical data, structure, and shape of pavement cracks in image sub-patches to construct crack-feature vectors and identifies the categories of sub-patches through sparse representation, so as to improve the accuracy of pavement crack recognition. This was done in order to improve the quality of the pavement crack

recognition process. A method for detecting cracks in asphalt pavement based on its spatial aggregation feature has been proposed by one researcher. This method takes into consideration the spatial distribution characteristics, grayscale characteristics, and geometric characteristics of cracks. An adaptive threshold method was proposed by Liang et al., which is capable of realizing the extraction of small cracks. J. Knig et al. were successful in achieving their goal of detecting pavement cracks by combining the ICS-LBP method. This method is superior to the method of directly using gray value features because it can achieve real-time crack detection. Even though the methods of image processing have made some headway in the detection of pavement cracks, they are all based on the texture and geometric features of the surface. They are able to produce satisfactory results when dealing with a limited number of crack images. In spite of this, when confronted with a large number of complex crack images, the parameters of the texture geometry need to be adjusted in order to obtain satisfactory results, which cannot adapt to a greater variety of crack images [13–15].

With the advancement of technology, the use of image processing technology to detect pavement cracks has attracted an increasing amount of attention from scientific research institutions. As a result, many pavement detection systems have been developed, such as road crack, which is a pavement detection system that was developed in Australia. In accordance with the complexity of block cracks and network cracks, these two types of cracks are intertwined with horizontal and vertical cracks, and they develop into pavement cracks, which are not pavement cracks that appear early on in the process. As a result, the adaptability of the standard algorithm for crack segmentation is insufficient. In this paper, we will implement these mature clustering algorithms that are applied in engineering, and through the performance comparison of various algorithms, we will provide new ideas for pavement crack detection. In addition, the results of the tests can provide a decision-making basis for the work of maintaining the pavement, thereby lowering the costs of maintenance, improving economic benefits, ensuring driving safety, and developing a positive environment for infrastructure.

2. Related Works

2.1. Noise Model of Crack Image. Noise, as we are all aware, is a random phenomenon that can only be characterized statistically. There are various types of noise, such as additive noise and multiplicative noise, among others. The statistical distributions of various noises, as well as the effects that those noises have on images, are each unique [16, 17]. Complex noise types can be seen in pavement images, with additive noise being the primary type. The noise model is presented in the following format for ease of analysis:

$$f(x) = x + n, \quad (1)$$

where $f(x)$ is the observed value; and x and n are the true value and the noise matrix, respectively. Gaussian noise is the most common additive noise. Figure 1 shows the noisy crack image with different noises. For example, Figure 1(d)

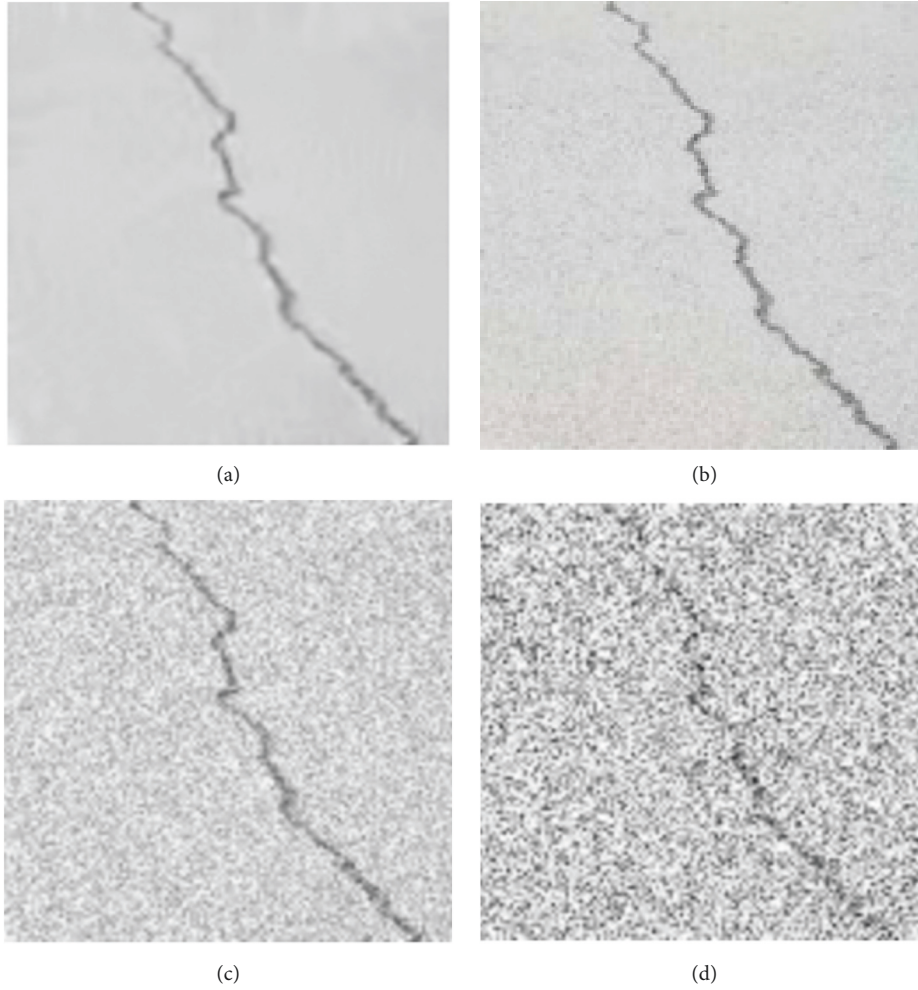


FIGURE 1: Noisy crack image. (a) RAW image; (b) noisy image with zero mean and 5 variance; (c) noisy image with zero mean and 10 variance; (d) noisy image with zero mean and 30 variance.

shows the image with zero mean and 30 variance. It can be seen that the noise obviously interferes with the crack and some details have been lost.

In general, the topological structure of a pavement fracture looks like a tree, while in specific areas it has a linear appearance. There are a number of very fine cracks, and the contrast between them varies quite a bit. In general, there are two types of crack images, which are known as crack backgrounds and false-crack backgrounds. One Rayleigh distribution and two normal distributions were chosen to simulate the statistical distribution of the crack background; the high gray area represents the fracture target, and a normal distribution was selected because it fits the statistical distribution of the crack image. The statistical results show that the probability of medium and low gray areas is designated as the crack background. One way to explain the statistical model is as follows:

$$f(x_i) = \omega_4 f_4(x_i) + \sum_{l=1}^3 \omega_l f_l(x_i), \quad (2)$$

where $\omega_l (1 \leq l \leq 4)$ is weight coefficient; the constraints are $\sum_{l=1}^4 \omega_l = 1$; $f_4(x_i)$ is the density function of Rayleigh

distribution with parameter γ ; and $f_l(x_i) (1 \leq l \leq 3)$ is a normal distribution density function with mean and variance $\{u_l, \sigma_l^2\}$. Therefore, the parameter set of crack image statistical distribution model is $\{\omega_l, \gamma, u_l, \sigma_l^2\}$, $l = 1, 2, 3, 4$. In the practical application of engineering crack detection, the shape and structure of cracks are complex and random. The contrast of cracks in different positions is large, and there are a large number of peripheral cracks and fine cracks. These characteristics increase the difficulty of crack segmentation and seriously affect the accuracy of crack segmentation. Therefore, the traditional single grayscale feature of crack cannot effectively segment the crack, so the influence of crack characteristics and noise on the crack segmentation algorithm must be eliminated by introducing additional information. Therefore, image enhancement is used to eliminate the influence of noise and interference in this paper.

2.2. Crack Image Enhancement. The intensity of the light and the angle at which the photograph was taken caused the surface of the fallen concrete to produce shadows that

resembled crack structures. This occurred during the process of acquiring the photograph. The false detection rate of cracks is increased as a result of these noncracks. It is challenging to differentiate between cracks and noncracks using only two-dimensional information because there is insufficient information regarding the depth of the three dimensions. The image of the crack that was collected in the real scene can be seen in Figure 2. The effect of the lighting on the image results in a lack of contrast, and the uneven pavement surface results in an uneven background area, both of which contribute to the lack of contrast between the cracks in the pavement. In addition, the red areas that can be seen in the image are not cracks but rather road scratches. However, the gradient characteristics of cracks are more readily apparent than those of false cracks, and appropriate filtering methods can be utilized to reduce the interference caused by false cracks.

Mean filtering, median image filtering, Gaussian image filtering, and nonlocal mean image filtering have emerged in recent years as the primary image filtering methods. The mean filter is the simplest, but it significantly reduces the prominence of edge features; the salt-and-pepper noise can be effectively removed by the median filter, but the crack image, which contains the vast majority of speckle noise, is not an appropriate candidate for this filter; the Gaussian filter has a beneficial effect on smoothing, and its overall performance is superior to that of the mean filter and the median filter; nonlocal mean filtering has a noise reduction effect that is comparable to that of Gaussian filtering. It also has the ability to preserve edge characteristics, despite the fact that the calculation is more complicated and the processing time is longer. In order to get rid of image noise and smooth out the gradient characteristics of false cracks, a two-dimensional Gauss low-pass filter has been chosen. This decision was made after taking into account the processing speed and effect in their entirety. The expression of filter can be shown like this:

$$G(u, v) = \frac{1}{2\pi\sigma} e^{-u^2+v^2/2\sigma^2}, \quad (3)$$

where $\sigma = 1$, and the sum of squares of u and v represents the distance from the pixel (u, v) to the center.

The crack edge characteristics will be softened to some amount after image de-noising, and some crack details will be lost. As a result, it is required to improve the crack structure as well as the contrast of the crack. The Frangi filter, which is based on the Hessian matrix, is used to extract the linear structure in the image and suppress background clutter, which has a noticeable enhancing effect on the cracks. To calculate the gradient change rate of pixels, the Hessian matrix is built of second-order partial derivatives. The relationship between the eigenvalues λ_1 and λ_2 reflects the structure of the image. The Frangi filter detects the linear structure in the image by using the relationship between the eigenvalues λ_1 and λ_2 of the Hessian matrix and defines the eigenvalue linear parameters R_b and S to construct an enhancement filter function, which can distinguish the background and the target. According to the literature [8], these linear parameters can be written as follows:

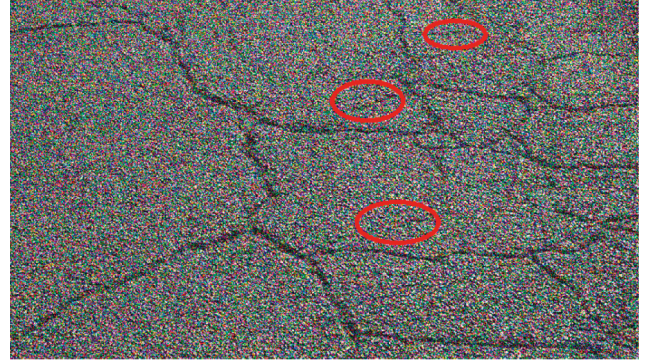


FIGURE 2: Comparison between cracks and noncracks.

$$V_0(S) = \begin{cases} 0, & \lambda_2 > 0, \\ \exp\left(-\frac{R_b}{2\beta^2}\right) \left[1 - \exp\left(-\frac{S^2}{2C^2}\right)\right], & \text{others,} \end{cases} \quad (4)$$

$$\begin{cases} R_b = \frac{\lambda_1}{\lambda_2}, \\ S = \|H\|_F = \sqrt{\sum_{j \leq 2} \lambda_j^2}, \end{cases} \quad (5)$$

β and c are used to adjust the sensitivity of R_b and S , which is generally set to 0.5 and 15.

3. Classic Computational Intelligence Algorithms for Crack Image

3.1. K-Means Clustering. The fundamental premise of the K-means clustering algorithm is to group data samples into k -classes, calculate the distance from other points to the center of the cluster, assign the samples that are geographically closest to the center of the cluster to a single class, calculate the new center of the cluster, and update the attribution of each sample iteration after iteration until the center of the cluster does not change significantly and satisfactory clustering results are obtained. The K-means clustering technique is a part of the hard clustering division. Within this division, the Euclidean distance is utilized as the similarity measure, and the sum of squares of errors is utilized as the criterion function to calculate the cost value of repetitive operations. The total number of deviations that exist between the sample and the cluster center is referred to as the sum of squared errors. When determining the accuracy of a cluster, the greater the cost value, the more significant the mistake. The optimal partition can be obtained only in the case where the cost function is the smallest possible value.

Assuming that the data samples are a data set $tX = \{x_1, x_2, \dots, x_n\}$ with n samples, the n samples will eventually be divided into k clusters, and the k clusters meet the following conditions: (1) each classification cannot be empty; (2) each data sample exists in only one classification. First, k samples are randomly selected as the initial cluster centers

of K clusters, where Euclidean distance is used as the similarity measurement and each sample is classified into the category of the nearest cluster center according to the size of the distance. The average value of all data samples in each cluster is calculated with Equation (6) as the new cluster center:

$$m_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}, i = 1, 2, \dots, k, \quad (6)$$

where N_i is the number of samples in the i th cluster.

The similarity between each sample and the new cluster center is recalculated, and the samples are reclassified before calculating the new cluster center. If the cluster center moves little or no samples are reassigned before and after clustering, the sample's clustering iteration terminates and the clustering cost function reaches a minimum. The following is the equation for the cost function of K-means clustering:

$$J = \sum_{i=1}^k \sum_{j=1}^{N_i} \|x_{ij} - m_i\|_2^2. \quad (7)$$

In the process of each iteration, judging whether each sample is assigned to the correct class is the key of K-means. If the classification is incorrect, it is necessary to make changes. After all samples are classified and corrected, it is judged that all samples are classified into the correct class, and then the cluster center is calculated and modified, and the next iteration is started. The steps of K-means clustering algorithm are described as follows.

Step 1. Divide the samples into k groups.

Step 2. Randomly select k points as the cluster center (m_1, m_2, \dots, m_k) of k groups.

Step 3. Calculate the distance from each sample x_i to k cluster centers (m_1, m_2, \dots, m_k).

Step 4. For each sample x_i , assign it to the cluster with the closest center.

Step 5. Calculate the new cluster center according to Equation (5).

Step 6. Calculate the cost function according to Equation (6).

Step 7. If the value of the cost function J converges, output the cluster centers (m_1, m_2, \dots, m_k) and the algorithm terminates. Otherwise, return to Step 2.

3.2. Fuzzy C-Means. In FCM algorithm, it is given that the data set $X = \{x_1, x_2, \dots, x_n\}$ with n samples has divided the samples into c fuzzy groups, where $2 \leq c \leq n$ is the clustering centers of c fuzzy groups. The Euclidean distance function is used to calculate the similarity between the samples and the clustering centers, assign membership values to them, obtain

the membership matrix U of all samples, and then obtain a new clustering center until the objective function representing the dissimilarity index finally reaches the minimum value. Since the membership matrix U has a normalization regulation, the total membership degree of a sample for each category is equal to 1.

$$\sum_{i=1}^c U_{ij} = 1, \forall j = 1, 2, \dots, n. \quad (8)$$

Therefore, the objective function of FCM can be described as follows:

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2. \quad (9)$$

When the objective function $J(U, c_1, c_2, \dots, c_c)$ takes the minimum value, the best clustering effect can be obtained.

$$\begin{aligned} \hat{J}(U, c_1, c_2, \dots, c_c, \lambda_1, \dots, \lambda_n) \\ = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right), \end{aligned} \quad (10)$$

where λ_j ($j = 1, \dots, n$) is the Lagrangian multiplier of n constraint expressions in (9). The obtained value when the derivative is 0 is the value that minimizes the (10), so the necessary conditions are as follows:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \quad (11)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij}/d_{kj})^{(2/m-1)}}.$$

The above two essential requirements yield the new cluster center and membership matrix. The fuzzy C-means clustering algorithm is a straightforward iterative procedure.

3.3. Maximum Entropy Clustering. The maximum entropy clustering algorithm is a deterministic annealing clustering algorithm based on the entropy. By taking the entropy as a constraint, it tries to maximize the entropy while minimizing the fuzzy distortion. The algorithm regards the entropy function as a component of the objective function to achieve a better division of the data set.

The maximal entropy clustering algorithm can be described by the following mathematical model. Given n vector $X = \{x_1, x_2, \dots, x_n\} \in R^s$ with s -dimensional space, it is assumed that there are c fuzzy class according to some similarity measures. The center of the i th fuzzy class is expressed as y_i , $i = 1, 2, \dots, c$, and the posterior probability of each vector belonging to the fuzzy class is represented by a set of real numbers between 0 and 1, which can be denoted as $\{P = (p_{y_i | x_i}) \in R^{cn} (i = 1, \dots, c; k = 1, \dots, n)$. This set of real numbers satisfies the following constraints: $\sum_{i=1}^c p(y_i | x_k) = 1$; $p(y_i | x_k) \in [0, 1]$; and $\sum_{i=1}^c p(y_i | x_k) > 0$. Therefore, the objective function of the maximum entropy clustering algorithm is defined as

$$J = \sum_{i=1}^c \sum_{k=1}^n p(y_i | x_k) (x_k - y_i)^2 + T \sum_{i=1}^c \sum_{k=1}^n p(y_i | x_k) \log p(y_i | x_k), \quad (12)$$

where $Y = (y_1, y_2, \dots, y_c)^T \in R^{CS}$; $d(x_k, y_i) = \|x_k - y_i\|^2$. The maximum entropy clustering algorithm is to find a probability distribution and a set of cluster centers to make the fuzzy distortion degree L the smallest and the entropy H to be the largest. Generally, the Lagrange multiplier method can be used to obtain the minimum value (p^*, y^*) of J .

It can be seen from the description of the maximum entropy clustering algorithm that when T is large, $\sum_{i=1}^c \sum_{k=1}^n p(y_i | x_k) \log p(y_i | x_k)$ plays a major role in the objective function. In addition, the probability of each sample belonging to c cluster centers is equal, which is approximately equal to $1/c$. As T gradually decreases, $\sum_{i=1}^c \sum_{k=1}^n p(y_i | x_k) (x_k - y_i)^2$ plays a more and more important role in the objective function. In addition, the probability of each sample belonging to the cluster center closest to it is increasing. When $T=0$, each sample belongs to the nearest cluster center with probability 1.

3.4. Gaussian Mixture Model. Gaussian mixture models are about quantifying things precisely with multiple Gaussian probability density functions. Each Gaussian mixture model is composed of N Gaussian distributions, each of which is called a cluster. These Gaussian distributions are combined to form the probability density function of the Gaussian mixture model.

$$p(x) = \sum_{n=1}^N p(n) p(x | n), \quad (13)$$

$$p(x) = \sum_{n=1}^N \pi_n N(x | \mu_n, \Sigma_n),$$

where N is the number of models; π_n represents the weight coefficient, which means the probability that each cluster class is selected, and $\sum_{n=1}^N \pi_n = 1$; and $N(x | \mu_n, \Sigma_n)$ is the Gaussian distribution density, $\Sigma_n = \delta_n^2$. δ_n^2 represents the standard deviation of the n th class, so the n th Gaussian model is denoted as follows:

$$N(x | \mu_n, \delta_n^2) = \frac{1}{\sqrt{2\pi}\delta_n} \exp\left(-\frac{(x - \mu_n)^2}{2\delta_n^2}\right). \quad (14)$$

Assuming that there are K collected samples, these samples can be considered to obey a Gaussian distribution, and then the likelihood function of GMM can be written as

$$\prod_{i=1}^K p(x) = \sum_{i=1}^K \log \left(\sum_{n=1}^N \pi_n N(x | \mu_n, \delta_n^2) \right). \quad (15)$$

Since the maximum value cannot be obtained directly in (14), so the EM algorithm must be adopted. The steps of EM algorithm for parameter estimation of Gaussian mixture model are described as follows:

Step 8. Determine initial value π, μ, δ .

Step 9. Calculate the probability that the sample is Gaussian distribution.

$$\gamma(i, n) = \frac{\pi_n N(x_i | \mu_i, \delta_i^2)}{\sum_{j=1}^N \pi_j N(x_i | \mu_j, \delta_j^2)}. \quad (16)$$

Step 10. Calculate maximum likelihood function of μ_n and δ_n^2 .

$$u_n = \frac{1}{K_n} \sum_{i=1}^K \gamma(i, n) x_i, \quad (17)$$

$$\delta_n^2 = \frac{1}{K_n} \sum_{i=1}^K \gamma(i, n) (x_i - \mu_n)(x_i - \mu_n)^T.$$

Step 11. Calculate maximum likelihood function of π_n , namely, $\pi_n = K_n/K$.

Step 12. Repeat iteration steps 2–4 until the parameter likelihood function value converges.

3.5. Mean-Shift Clustering. Mean-shift is a prediction method of gradient search through nonparametric probability density estimation, which classifies the sample data and counts the pattern categories in the feature space of the data samples. The mean-shift-based image segmentation is to first estimate the probability density, then find the convergence point through gradient search, use the convergence point to filter the image, and finally achieve image segmentation.

Nonparametric estimation refers to the process of estimating the density function using sample data sets. Kernel density estimation is the most commonly used nonparametric estimation, which is the process of processing the sample according to the kernel function $K(x)$ to obtain the density function. For a Euclidean space with d -dimensional, x is one of the eigenvectors, and R is a real number field. If a function $K: R^d \rightarrow R$ follows $K: [0, +\infty] \rightarrow R$, we can obtain $K(x) = C_k k(\|x\|^2)$, where the modulus of x is denoted as $\|x\|^2 = x^T x$; C_k is a standardized constant; and k obeys non-negative and piece-wise continuous and $\int k(r) dr < \infty$.

Given the known kernel function $K(x)$ and bandwidth matrix $H_i(x)$ of n sampling points $\{x_i, 1 \leq i \leq n\}$ in the space R^d , the kernel density estimation formula of the density function can be denoted as

$$\hat{f}(x) = \sum_{i=1}^n c_k \omega_i |H_i|^{-1/2} K\left(\|x - x_i\|^2 |H_i|\right), \quad (18)$$

$\omega(x_i) \geq 0$ is the weight of the sampling point, which satisfies $\sum \omega(x_i) = 1$ and abbreviates as ω_i . The kernel function $K(x)$ is the similarity measurement between the sample x_i and the center x . The bandwidth matrix H_i represents the range of

the kernel function estimation. The density function estimate $\hat{f}(x)$ is obtained by the weighted sum of the kernel function at each sample.

Generally, the density at x is always smaller than that at $m_{H_i}(x)$, and the mean-shift vector should move in the direction of high density. The convergence point of the mean-shift algorithm is the local density maximum point. In order to describe the mean-shift algorithm more simply, it is given that the weights between sampling points are equal, namely, $\omega(x_i) = 1/n$. Since the bandwidth matrix is proportional to the identity matrix $H_i = h^2 I$, then the mean-shift iteration formula can be written as follows:

$$m_h(x) = \frac{\sum_{i=1}^n g\left(\|(x - x_i)h^{-1}\|^2\right)x_i}{\sum_{i=1}^n g\left(\|(x - x_i)h^{-1}\|^2\right)}. \quad (19)$$

The mean-shift algorithm firstly and randomly selects a search area circle in the sample and calculates the average value of all sample points in the search area through the iterative formula. The density of the newly obtained mean point must be greater than the density at the initial center point. Repeat the above steps until the density change is less than a certain value, and then converge to the maximum density point.

3.6. Hierarchical Clustering. In the method known as hierarchical clustering, a structure that looks like a tree is built to correlate with the samples that are going to be clustered. This is done in order to facilitate the clustering process. The process of hierarchical clustering can be further subdivided into a variety of distinct clustering algorithms depending on the direction in which the hierarchy is constructed, either from the bottom up or from the top down. These distinct algorithms can then be used to cluster data in a variety of different ways. Because of the flexibility of the hierarchical structure, it is possible to make this distinction in either direction during its construction. When applying the bottom-up methodology, the first thing that needs to be done is to treat every single sample as if it were its own cluster. After that, it combines these atomic clusters to produce clusters that are progressively larger until either all of the samples are included within a single cluster or a termination condition is satisfied, whichever occurs first. After that, it moves on to the next step. The concept of similarity between clusters can be understood quite differently depending on the approach that is taken, despite the fact that the vast majority of hierarchical clustering algorithms fall into this category. Hierarchical clustering takes a different approach, one that works from the bottom up, as opposed to the strategy that operates in the opposite direction, from the top down. It starts by grouping all of the samples together into a cluster, and then it gradually divides that larger cluster into smaller and smaller clusters until either every object spontaneously forms its own cluster or it satisfies a condition that causes it to stop.

The relative connectivity between two clusters c_i and c_j is defined as the absolute connectivity between two clusters c_i and c_j divided by two relative connectivities in c_i and c_j .

$$RI(c_i, c_j) = \frac{|\text{EC}(c_i, c_j)|}{\left(|\text{EC}(c_i)| + |\text{EC}(c_j)|\right)/2}, \quad (20)$$

where $|\text{EC}(c_i, c_j)|$ is the edge-out containing c_i and c_j clusters, so that the cluster can be decomposed into $|\text{EC}(c_i)|$ and $|\text{EC}(c_j)|$. The relative approach degree between two clusters c_i and c_j is the absolute approach degree between two clusters c_i and c_j divided by the approach degree within two clusters c_i and c_j :

$$RC(c_i, c_j) = \frac{S_{\text{EC}(c_i, c_j)}}{\left(\alpha_i S_{\text{EC}(c_i)} + \alpha_j S_{\text{EC}(c_j)}\right)}. \quad (21)$$

4. Experiment

4.1. Evaluation Indicators. In order to provide a reasonable evaluation of the clustering performance of each clustering algorithm, this study uses the NMI and the RI evaluation indicators to analyze the performance of each comparison algorithm. These two indicators have a value range of [0, 1], and a good rule of thumb to follow is that a higher number indicates a higher level of clustering performance. Calculating the degree of error that occurred during the process of judging the results of the fracture segmentation can be done with the help of an index known as the NMI index. When this value is increased, the crack segmentation will become more accurate as a result. The magnitude of the value is inversely proportional to the extent of this improvement. The RI index examines and quantifies the degree to which the segmentation results of the crack segmentation algorithm and the results of the ground-truth segmentation are consistent along their edges. This is accomplished by comparing the segmentation results of the two sets of data. This is accomplished by contrasting the outcomes of the two studies with one another. When the value is increased, the algorithm will produce segmentation results that are of a higher quality than they were before.

These cluster methods are first tested on a Windows computer equipped with a 2.50 GHz Intel Core i5-1135G7 processor and 4 gigabytes of RAM, and then they are implemented in MATLAB using the MATLAB R2012b software, which is run on a Windows 10 64-bit operating system. Finally, the cluster methods are validated on a Windows computer.

4.2. Data Sources. According to the types and characteristics of pavement cracks, this paper selects some images of different locations, different environments, and different time periods. These samples cover typical cracks such as bifurcation, uneven thickness, uneven illumination, occlusion, road markings, and their shadows, which can meet the testing requirements of crack image clustering algorithm. These crack images can be found and downloaded from the Internet, such as <https://img.xianjichina.com/editer/20210113/image/e80456e50e92c5eb0b8b03b.jpg>. For testing



FIGURE 3: The samples of typical crack images.

requirements, the crack data set with a size of $128 * 128$ or $64 * 64$ pixels is formed by cutting out the crack images. Many of them suffer from the problems of occlusion and shadow interference cracks and fine cracks and carrying a large number of peripheral cracks, which is difficult to segment. Figure 3 shows the typical crack images selected in this paper, the first row is single-crack images, and the second row is cracks with complex background interference. We manually annotate the ground-truth crack on test images for objective performance evaluation.

4.3. Experimental Analysis. In this article, several pavement crack images are processed with the help of six different clustering algorithms. The degree of complexity of the crack image has a significant bearing on the segmentation results achieved by the clustering algorithm. Additionally, because the acquisition equipment is susceptible to the influence of light intensity and noise, the brightness of various cracks in the crack image is typically uneven. This will have a negative impact on the effectiveness of the crack image segmentation. In order to properly segment cracks in an image, it is necessary to preprocess the image first.

In this study, an image enhancement algorithm was applied during the image preprocessing stage in order to get rid of any noise and interference that might have been present. Image enhancement will, simultaneously, improve the response value of crack image similarity measure, which will be helpful in improving the performance of crack segmentation. Even though the main body of the crack has been preserved thanks to image enhancement, the fine cracks and peripheral cracks have been lost in a significant way, which has resulted in the crack having poor continuity. In addition, the phenomenon of under-segmentation is more serious, and the segmentation itself presents a fracture phenomenon. The noise in the crack segmentation result has been greatly reduced, but the segmentation result makes it easy to lose details; the smaller crack area is easy to connect,

but it is easy to distinguish the subtle and peripheral cracks as noncrack area, and the cracks are disconnected, which results in serious over-segmentation of the cracks. Therefore, the preprocessing method that was used in this paper helps retain the details, and the results of this method do not affect how well the six different clustering algorithms perform when compared to one another.

K-means clustering (K-means), fuzzy c-means clustering (FCM), maximum entropy clustering (MEC), Gaussian mixture mode (GMM), mean-shift, and hierarchical clustering (HC) are used as comparison clustering algorithms in order to verify the comprehensive segmentation performance of these clustering algorithms. Test samples are randomly selected from the pavement crack database. Because of the constraints imposed by the available space, we can only select a select few examples of typical image segmentation results for analysis, as can be seen in Figures 4–6.

It can be seen from the segmentation results that different algorithms can basically achieve crack segmentation, but the segmentation results show obvious differences. The K-means algorithm is highly sensitive to outlier noise, and the selection of cluster centers directly affects the segmentation results. The segmentation results only preserve the main structure of the crack. Noise and the interference of the complex background cause the segmentation algorithm to lose a lot of details, especially the edge information.

The segmentation results show a fracture phenomenon, there is a lot of noise, and there is an over-segmentation phenomenon such as judging the complex background as a crack area. The neighborhood information that introduces crack pixels into the feature is improved compared with the segmentation result of the K-means algorithm, but the noise sensitivity is still high, which easily leads to the islanding effect. It is easy to judge the background as a crack at the tip and subtle parts of the image, and there is still a lot of noise in the noncrack area.

Based on the gray distribution characteristics of cracks, the GMM algorithm uses Gaussian distribution to fit the

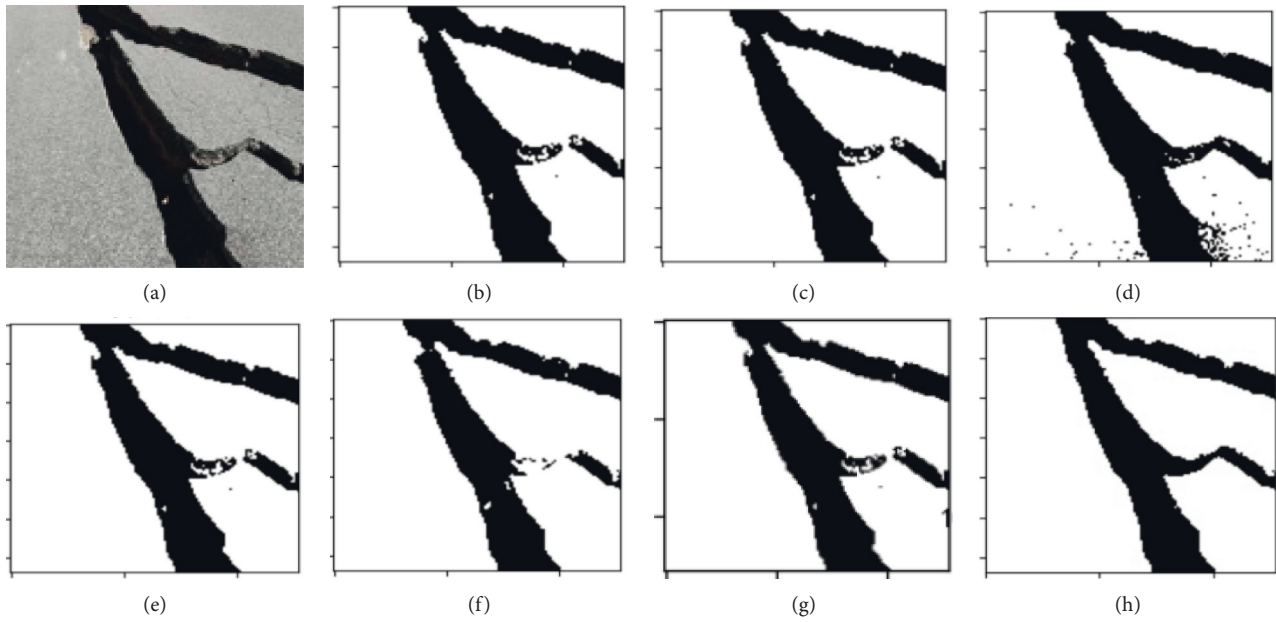


FIGURE 4: Performance comparison for different cluster algorithms. (a) RAW images; (b) K-means; (c) FCM; (d) MEC; (e) GMM; (f) mean-shift; (g) HC; (h) ground-truth.

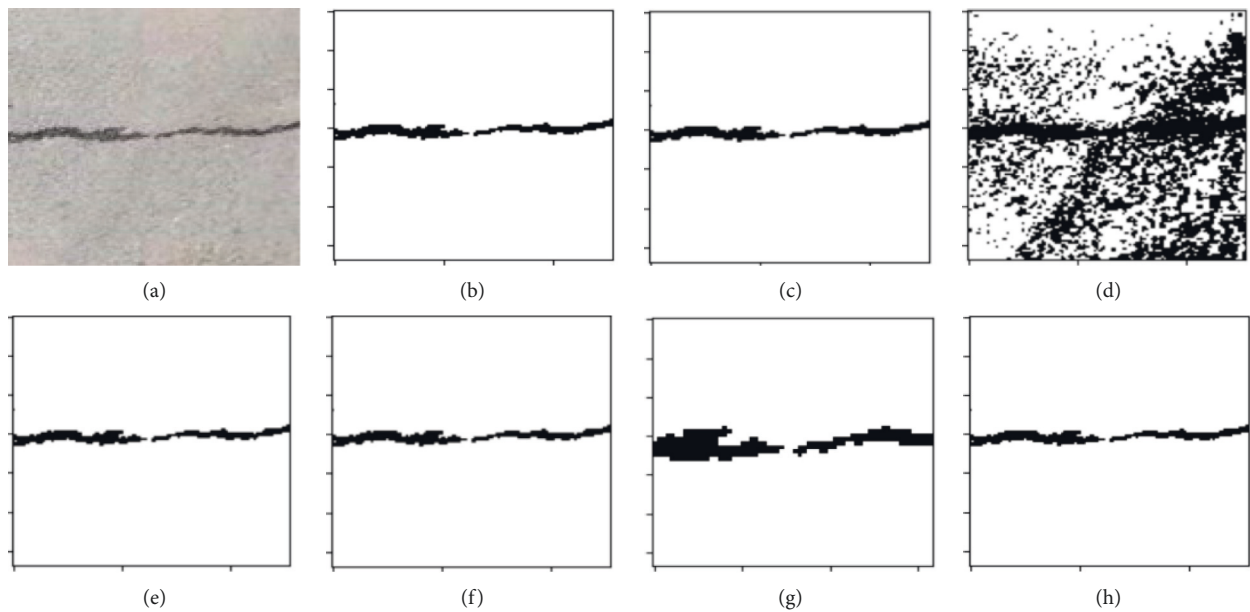


FIGURE 5: Performance comparison for different cluster algorithms. (a) RAW images; (b) K-means; (c) FCM; (d) MEC; (e) GMM; (f) mean-shift; (g) HC; (h) ground-truth.

feature distribution of cracks and completes the segmentation of crack images through adaptive thresholds, but the selection of parameter values calculated by the EM algorithm directly affects the segmentation results. Although the segmented cracks are relatively continuous, it is easy to mistakenly identify noncrack areas as crack area, showing an over-segmentation phenomenon, and there are a small number of noise points around the cracks, as shown in Figure 5(d). In a single crack, it is easy to misjudge the crack target as a noncrack with under-segmentation phenomenon.

Figure 6(g) shows the segmentation result of the algorithm in this paper. The segmentation effect is significantly improved, and the crack fracture phenomenon is improved. The mean-shift algorithm in this paper can completely segment the crack structure and effectively solve the detailed problems of crack image segmentation, such as fine cracks, peripheral cracks, and branch cracks. At the same time, the hierarchical clustering algorithm can extract clear crack area, which can well overcome the segmentation difficulties caused by uneven gray scale, noise, and

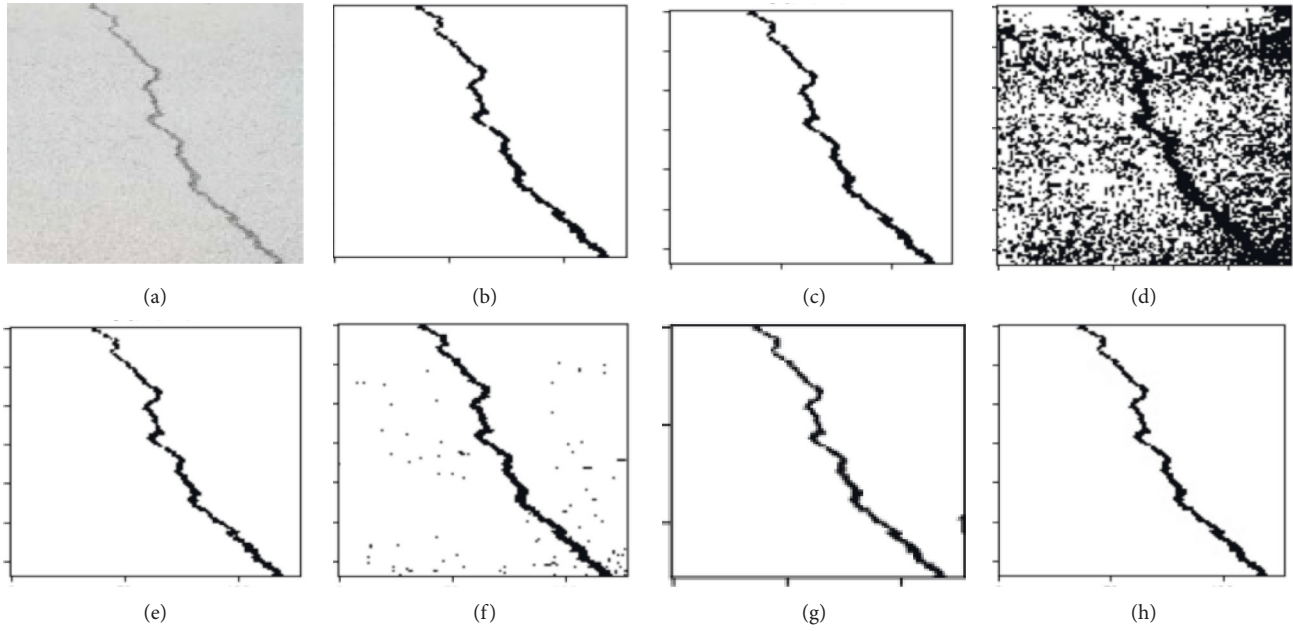


FIGURE 6: Performance comparison for different cluster algorithms. (a) RAW images; (b) K-means; (c) FCM; (d) MEC; (e) GMM; (f) mean-shift; (g) HC; (h) ground-truth.

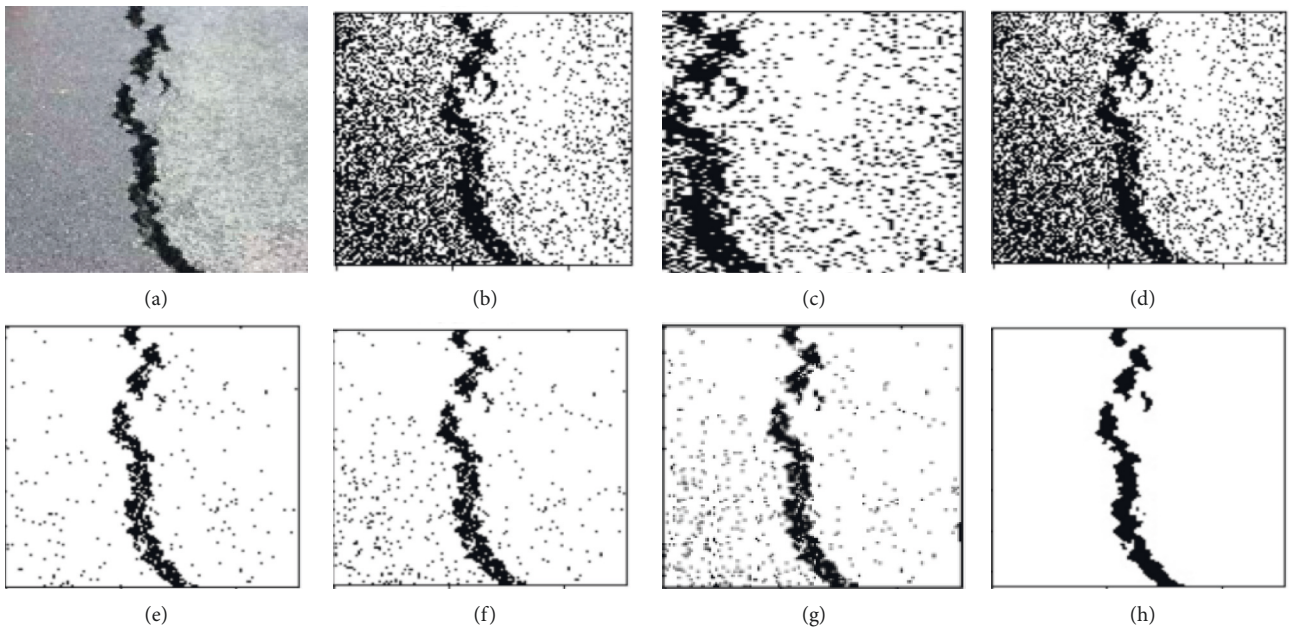


FIGURE 7: Performance comparison for different cluster algorithms with noise ((zero mean and 0.01 variance)). (a) RAW images; (b) K-means; (c) FCM; (d) MEC; (e) GMM; (f) mean-shift; (g) HC; (h) ground-truth.

overlapping edge. The crack segmentation results can provide high-quality reference for subsequent data analysis and crack classification.

Since this paper mainly analyzes the processing effects of different clustering algorithms on different noisy images, Figure 7 shows the result of adding Gaussian noise with zero mean and 0.01 variance to the raw image and clustering directly. It can be seen that the introduction of noise significantly changes the segmentation performance. Noise

interference appeared in all results, and some background areas were mistaken for cracks. Although the image enhancement function is added in the preprocessing stage of this paper, there are still some interferences, resulting in errors in the results, but the introduction of preprocessing is also significantly better than the results of direct clustering. It is worth noting that although all comparison algorithms have some scattered speckle areas disturbed by noise, the interference can be eliminated by edge detection.

TABLE 1: Performance comparison of different clustering algorithms for the same crack image without noise.

Images	Indexes	Algorithms					
		K-means	FCM	MEC	GMM	HC	Mean-shift
1	RI	0.9637	0.9639	0.9358	0.9637	0.9720	0.9639
	NMI	0.8371	0.8379	0.7707	0.8366	0.8792	0.8379
2	RI	0.9985	0.9985	0.5521	0.9973	0.9979	0.9985
	NMI	0.9621	0.9621	0.0792	0.9402	0.9517	0.9630
3	RI	0.9980	0.9996	0.5286	0.9944	0.9756	0.9995
	NMI	0.9516	0.9886	0.0624	0.8821	0.6861	0.9853
4	RI	0.9650	0.9664	0.6653	0.9655	0.9655	0.9659
	NMI	0.7014	0.7082	0.2110	0.7079	0.7090	0.7101
5	RI	0.9374	0.9383	0.8837	0.9381	0.9300	0.9386
	NMI	0.7311	0.7379	0.5630	0.7371	0.6902	0.7448
6	RI	0.5918	0.5933	0.5192	0.9885	0.9837	0.9867
	NMI	0.0669	0.0673	0.0442	0.7092	0.6192	0.6842
7	RI	0.9615	0.9606	0.5342	0.9696	0.9681	0.9702
	NMI	0.5424	0.5377	0.0585	0.6143	0.6015	0.6194
8	RI	0.9880	0.9880	0.5187	0.9916	0.9919	0.9922
	NMI	0.8694	0.8694	0.1042	0.8941	0.8989	0.9018
9	RI	0.9874	0.9874	0.5338	0.9843	0.9857	0.9956
	NMI	0.8827	0.8827	0.1383	0.8668	0.8718	0.9495
Average	RI-mean	0.9324	0.9329	0.6302	0.9770	0.9745	0.9790
	RI-std	0.1219	0.1216	0.1558	0.0182	0.0188	0.0195
	NMI-mean	0.7272	0.7324	0.2257	0.7987	0.7675	0.8218
	NMI-std	0.2652	0.2691	0.2455	0.1034	0.1248	0.1282

TABLE 2: Performance comparison of different clustering algorithms for the same crack image with noise (zero mean and 0.01 variance).

Images	Indexes	Algorithms					
		K-means	FCM	MEC	GMM	HC	Mean-shift
1	RI	0.8358	0.8358	0.8358	0.8513	0.8508	0.8633
	NMI	0.4663	0.4663	0.4663	0.5134	0.5116	0.5555
2	RI	0.5230	0.5177	0.5177	0.7176	0.8753	0.9565
	NMI	0.0620	0.0600	0.0600	0.1379	0.2645	0.4370
3	RI	0.5104	0.5065	0.5065	0.5486	0.9542	0.6283
	NMI	0.0367	0.0350	0.0350	0.0450	0.1907	0.0817
4	RI	0.5653	0.5653	0.5648	0.9200	0.8567	0.9049
	NMI	0.1035	0.1035	0.0992	0.4207	0.3250	0.3901
5	RI	0.8528	0.8528	0.8528	0.8580	0.8506	0.8578
	NMI	0.4496	0.4496	0.4496	0.4759	0.4422	0.4741
6	RI	0.5063	0.5039	0.5035	0.5533	0.9703	0.9569
	NMI	0.0297	0.0289	0.0169	0.0403	0.3236	0.2967
7	RI	0.5172	0.5134	0.5006	0.6674	0.9459	0.8982
	NMI	0.0267	0.0252	0.0022	0.0688	0.3584	0.2318
8	RI	0.5957	0.5885	0.5885	0.9532	0.9345	0.9233
	NMI	0.1124	0.1083	0.1083	0.5882	0.5234	0.4900
9	RI	0.5366	0.5432	0.5130	0.7466	0.6992	0.7747
	NMI	0.0952	0.0996	0.1155	0.2266	0.1921	0.2704
Average	RI-mean	0.6048	0.6030	0.5981	0.7573	0.8820	0.8627
	RI-std	0.1309	0.1318	0.1346	0.1409	0.0788	0.0984
	NMI-mean	0.1536	0.1529	0.1503	0.2796	0.3479	0.3586
	NMI-std	0.1655	0.1659	0.1687	0.2076	0.1172	0.1418

The statistics in Tables 1 and 2 show the quantitative indicators of different clustering algorithms for the same crack image with/without noise. In Table 1, the NMI index of mean-shift algorithm reaches 82.86%, which is 8.83%,

2.06%, and 3.85% higher than that of K-means algorithm, MEC algorithm, and FCM algorithm, respectively; as for the RI index in Table 2, K-means algorithm, MEC algorithm, and FCM algorithm are 9.52%, 6.27%, and 1.23% higher,

respectively. Therefore, for pavement crack cluster application, the measurement index of mean-shift algorithm is better than the comparison algorithm and meets the needs of engineering detection.

According to the NMI-mean comparison of each clustering algorithm, the effect of MEC is the worst, the effect of other algorithms is not bad, and the gap is small, where the mean-shift clustering algorithm has the best effect. Various clustering algorithms have been adjusted and optimized to achieve the optimal segmentation effect. In this experiment, it seems that the NMI-based evaluation results are more in line with the performance of segmentation effects, while the segmentation results that seem to have poor effects can also have RI value above 0.5. The segmentation effect of these cluster algorithms is significantly worse after adding Gaussian noise. According to the NMI value, the mean-shift clustering algorithm has the best de-noise effect, while the performance of K-means algorithm and FCM algorithm decreases significantly after adding noise.

5. Conclusion

The purpose of this study is to evaluate and analyze the performance of complex crack images in a variety of different environments using a number of different standard clustering algorithms. NMI and RI have been selected to serve as assessment indices for the purpose of determining whether or not the comparison algorithm is effective. The performance of a comparison analysis on the noisy photos is an additional component of the experiment that needs to be carried out. The NMI value indicates that the mean-shift clustering algorithm has the best de-noise effect, whereas the performance of the K-means algorithm and the FCM algorithm significantly decreases after noise is added to the data. The findings of the experiments indicate that the addition of Gaussian noise makes these cluster algorithms significantly less effective at segmenting the data than they were before.

Despite the fact that many different clustering algorithms have been proposed and are continually being improved, there has not been developed a single algorithm that is appropriate for a variety of data features. This is primarily the result of the fact that the algorithms used for clustering put an excessive amount of emphasis on the compactness that exists within clusters as well as the differences that exist between clusters. The arrival of the era of big data has resulted not only in an increase in the total amount of data, but also in an increase in the complexity of the data structure. This is due to the fact that the total amount of data has increased. Finding a way to build a clustering algorithm and evaluation index that is flexible enough to accommodate a wide range of different scenarios will be an essential goal for the work that will be done in the future.

Data Availability

The 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 by the National Natural Science Foundation of China (No.61862051), the Science and Technology Foundation of Guizhou Province (No.ZK[2022]549), the Top-notch Talent Program of Guizhou Province (No.KY[2018]080), the Natural Science Foundation of Education of Guizhou Province(No.[2019]203), the Industrial Science and Technology Foundation of Qiannan (No.[2017]16), and the Funds of Qiannan Normal University for Nationalities (Nos. qnsy2018003, qnsy2019rc09, qnsy2018JS013, and qnsyrc201715).

References

- [1] X. Feng, L. Xiao, W. Li, L. Pei, Z. Sun, and Z. Ma, "Pavement crack detection and segmentation method based on improved deep learning fusion model[J]," *Mathematical Problems in Engineering*, vol. 12, no. 12, pp. 2020–2132, 2020.
- [2] W. Qiao, Q. Liu, X. Wu, B. Ma, and G. Li, "Automatic pixel-level pavement crack recognition using a deep feature aggregation segmentation network with a scSE attention mechanism module," *Sensors*, vol. 21, no. 9, pp. 2902–2911, 2021.
- [3] R. Augustauskas, "Improved pixel-level pavement-defect segmentation using a deep autoencoder[J]," *Sensors*, vol. 20, no. 9, pp. 567–587, 2020.
- [4] D. Mazzini, P. Napoletano, F. Piccoli, and R. Schettini, "A novel approach to data augmentation for pavement distress segmentation," *Computers in Industry*, vol. 121, Article ID 103225, 2020.
- [5] K. Zhang, Y. Zhang, and H. D. Cheng, "CrackGAN: pavement crack detection using partially accurate ground truths based on generative adversarial learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1306–1319, 2021.
- [6] Z. Fan, H. Lin, C. Li, J. Su, S. Bruno, and G. Loprencipe, "Use of parallel ResNet for high-performance pavement crack detection and measurement," *Sustainability*, vol. 14, no. 3, p. 1825, 2022.
- [7] A. Ji, X. Xue, Y. Wang, X. Luo, and W. Xue, "An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement," *Automation in Construction*, vol. 114, Article ID 103176, 2020.
- [8] H. Ju, W. Li, S. Tighe et al., "Three-dimensional pavement crack detection based on primary surface profile innovation optimized dual-phase computing[J]," *Engineering Applications of Artificial Intelligence*, vol. 89, no. Mar, pp. 103376.1–103376.11, 2020.
- [9] G. X. Hu, B. L. Hu, Z. Yang, L. Huang, and P. Li, "Pavement crack detection method based on deep learning models," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, Article ID 5573590, pp. 1–13, 2021.
- [10] L. Pei, Z. Sun, L. Xiao, W. Li, J. Sun, and H. Zhang, "Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network," *Engineering Applications of Artificial Intelligence*, vol. 104, pp. 104376–104387, 2021.

- [11] W. Song, G. Jia, H. Zhu, D. Jia, and L. Gao, "Automated pavement crack damage detection using deep multiscale convolutional features," *Journal of Advanced Transportation*, vol. 2020, Article ID 6412562, pp. 1–11, 2020.
- [12] D. Kaya Ozdemir, A. Topal, B. Kacmaz, and B. Sengoz, "Evaluating the asphalt pavement's surface characteristics by field testing," *Revista de la Construcción*, vol. 19, no. 3, pp. 474–485, 2020.
- [13] P. C. Cong, W. Jian, Z. Lei, and S. Chen, "Underwater dam image crack segmentation based on mathematical morpholog [J]," *Applied Mechanics and Materials*, vol. 220, pp. 1315–1319, 2012.
- [14] W. Kaddah, M. Elbouz, Y. Ouerhani, V. Baltazart, M. Desthieux, and A. Alfalou, "Optimized minimal path selection (omps) method for automatic and unsupervised crack segmentation within two-dimensional pavement images," *The Visual Computer*, vol. 35, no. 9, pp. 1293–1309, 2019.
- [15] H. B. Guo, R. Vaßen, and D. Stöver, "Atmospheric plasma sprayed thick thermal barrier coatings with high segmentation crack density," *Surface and Coatings Technology*, vol. 186, no. 3, pp. 353–363, 2004.
- [16] 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.
- [17] M. Li, J. Zhou, D. Wang, P. Peng, and Y. Yu, "Application of clustering-based analysis in MRI brain tissue segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2022, Article ID 7401184, 1 page, 2022.

Research Article

Analysis of Balance of Income and Expenditure and Optimal Retirement Age of Pension Insurance Co-Ordination Account Based on Improved Machine Learning Algorithm

Shi Yan, Yaodong Zhou, and Youlu Zhang 

School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Youlu Zhang; 18113024@bjtu.edu.cn

Received 18 July 2022; Revised 4 August 2022; Accepted 5 August 2022; Published 31 August 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Shi Yan 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.

Since the turn of the twenty-first century, the issue of aging has gained international attention. Both developed and developing nations are currently dealing with this issue. To ensure the sustained and healthy growth of the economy and society in the face of an aging society, it is especially important to establish a scientific old-age insurance system and a reasonable retirement system. We are all aware that the key indicators for the state to control the old-age insurance system in the old-age insurance system are the income and expenditure balance of the old-age insurance pooling account and the analysis of the ideal retirement age. In this paper, a better machine algorithm is used. By independently learning the rules present in a large amount of data and gaining new experience and knowledge, machine learning (ML) can increase computer intelligence and give computers decision-making abilities comparable to those of humans. In general, a machine learning algorithm uses the laws it derives from data to predict unknown data after automatically analysing the data. This study's findings suggest that the ideal retirement age and life expectancy are positively correlated, with the ideal retirement age's growth rate 12.57 percent higher than that of life expectancy.

1. Introduction

Before the 18th century, the population was mainly concentrated in villages and towns, with poor living environment, underdeveloped medical conditions, diseases, wars, and other reasons, and the population growth was very slow [1]. With the sudden rise of the first industrial revolution, science and technique were born, medical level and grain output were greatly improved, and the global population began to grow slowly. After entering the 19th century, the world ushered in a century-long period of relative peace, and with the help of the second industrial revolution, the population surged, forming the second growth peak in human history [2]. At the same time, due to the declining fertility rate in some developed countries in Europe, some countries began to show signs of aging [3]. Since the 20th century, quite a few countries, including China, have been facing the aging of population. Global challenges have been gradually introduced as aging has spread continuously. The

old-age security system is under tremendous strain due to the large population, and the level of contributions and replacements within the system not only affects how well workers will live after they retire but also indirectly affects how dependent retirees will be on old-age insurance [4]. A reasonable contribution rate and substitution rate will help to reduce the financial strain that the aging population is putting on the old-age insurance fund. In addition, a reasonable contribution rate and substitution rate will aid in the establishment and improvement of a multilevel old-age insurance system with Chinese characteristics, even the social security system, as well as in the promotion of the old-age insurance system of governmental institutions and agencies. Therefore, it is important from both a theoretical and practical standpoint to dynamically modify the contribution rate and substitution rate of people and businesses. Government officials and academics have expressed interest in determining a reasonable contribution rate and substitution rate. The income and expenditure of old-age

insurance are particularly impacted by the multiple levels and angles of the aging process [5]. Affected by the rapid improvement of aging, the on-the-job workers' support rate for the elderly population has increased, resulting in a decrease in the number of contributors to their pension funds but an increase in the burden. Pension income and expenditure will be unbalanced, which indicates that there will be a gap in the future, and the gap will be bigger and bigger [6]. From a global perspective, the future 21st century will be the century of global population aging. In the degree of aging, there are significant differences among continents in the world [7]. The proportion of the aging population in Asia has remained at a high level for a long time, and the aging growth rate ranks first in the world. The proportion of the elderly population in Europe shows a downward trend; "Silver tsunami" will sweep across most areas; only South Africa will be spared temporarily, but the progress is slow, rather than completely reversing the trend of population aging [8]. Old-age insurance systems in many nations have gradually run into financial difficulties as a result of the rapid increase in the world's aging population. Since their own old-age insurance systems and welfare policies differ from those of the United States, many foreign nations have experimented with adjusting the retirement age and have seen some success [9]. An index that is frequently used to reflect the overall health and quality of life of people in a nation or region is the average life expectancy of the population [10]. The average lifespan of the population has steadily increased, thanks to improvements in living standards, medical and health services, and the infant mortality rate (which has been steadily declining) [11]. Many workers consider the retirement age from their own point of view, and whether it is postponed or not is also highly controversial. For example, people with high education level want to postpone the retirement age in order to give full play to human resources. Many women think that their jobs are not limited by too many physiological conditions and think that they should retire at the same age as men. Some people also think that with the increase of age, their physical condition is not as good as that of their young age, their working passion drops sharply, their working efficiency drops, and it is inappropriate to postpone their retirement age and so on [12]. Undoubtedly, the increasing number of elderly people, the expanding coverage of the pension system, a large number of "empty accounts" running personal accounts, the incipient fiscal deficit, the serious shortage of pension reserves, the gradually obvious "hidden debt," and the reform cost of the transition between the old and new systems have all increased the financial payment crisis [13]. From this point of view, the pension system is facing huge financial payment pressure in the improvement pattern of "getting old before getting rich" [14].

The innovation of this paper lies in the following:

- (1) This paper introduces the old-age insurance. Because it is a part of the research topic of this paper, we should discuss it. Part of the pension is the focus of this article, so it was described from the beginning. Scholars generally believe that the basic old-age

insurance system is the main component of the modern social security system and an indispensable component. The government adopts laws, regulations, and other means to require enterprises and employees to pay endowment insurance in proportion and give corresponding pensions to workers after their legal retirement to maintain their life after retirement. This policy is the basic endowment insurance system.

- (2) This article introduces the retirement age. This paper has also studied it, so it is necessary to discuss it. With the rapid spread of population aging, although different countries have implemented different statutory retirement age systems according to their own national conditions, facing this global question, many countries have already or will soon choose to extend their retirement time. The flexible retirement system should be implemented, that is, different people can choose their retirement time independently according to their job nature, health, economic situation, family background, and employment intention, so as to meet the pursuit of retirement life quality.
- (3) ML can increase computer intelligence by allowing computers to independently learn the rules present in a large amount of data, gain new experience, and learn new things. This allows computers to make decisions that are comparable to those of humans. In order to give computers the ability to learn on their own, ML theory primarily involves designing and examining specific algorithms. In general, ML algorithms use the laws they derive from data to predict unknown data. They automatically analyse the data and derive the laws from it.

2. Related Works

Fitzpatrick suggested a complete OLG model idea, modified it, and used it as a vital analysis tool [15]. Lu and Tang suggested the standard to measure economic welfare and the way to improve it and advocated that income distribution should be equalized and the government should play an intervention role in income redistribution [16]. Hatcher suggested to choose ways to increase the contribution rate of individual pension to ensure the adequacy of individual pension accounts [17]. Turner suggested that theory of welfare economics should be based on cardinal utility theory, assuming that consumption can be measured concretely and summed up, also known as "old welfare economics" [18]. A B suggested two OLG models in order to analyse the optimal retirement age better [19]. Yue suggested the combination of unified account and personal account. The purpose of introducing personal account is to accumulate some funds to cope with the peak of retirement, which indicates that the old-age insurance system has achieved a major breakthrough of the "partial accumulation system" [20]. Chen suggested the basic idea of "small step by step," but the reform opportunity, rhythm grasp, the path of

delaying retirement, and the optimal retirement age of different social groups have all become difficult questions to be solved in the policy reform of delaying retirement [21]. Den et al. suggested two options of combining social pooling with individual accounts, allowing them to choose according to the actual situation and even modifying them appropriately [22]. Meng et al. suggested a two-stage reform plan: first, the retirement age of women should be adjusted to 55 without distinction. Then, men will retire at the age of 65 at the rate of 1 year/6 years and women at the rate of twice that of men [23]. Steiber and Kohli suggested a new idea of supplementing enterprise insurance and personal savings endowment insurance, but overall, it is still the product of the times under the planned economy system [24].

With the advancement of science and technology, ongoing economic growth, ongoing medical and health conditions improvement, and a significant rise in life expectancy, the population aging trend is becoming more and more obvious, and the global question of how to balance the supply and demand of crisis endowment insurance funds has emerged. To varying degrees, every nation in the world is improving and reforming its endowment insurance system. Old-age insurance is a type of insurance designed to reduce the risk that an elderly person's income will be interrupted or reduced, which will result in a decline in living standards. Endowment insurance plays a crucial role in China's national economy and social advancement as a crucial component of the social system. The need to restructure and enhance the old-age insurance system is urgent. The old-age insurance reform is currently in a phase of transition, with the goal of implementing the changeover between the old and new systems. The analysis of the balance of payments and the ideal retirement age for the pension insurance pooling account, which is of great significance, is based on an improved machine learning algorithm in this paper.

3. Endowment Insurance and Optimal Retirement Age

3.1. Endowment Insurance. The disparity between public pension revenue and expenditure in various nations is getting progressively worse as the proportion of the elderly population rises. The operating mechanism of endowment insurance systems is the same across all countries, despite the variations in system types. Most academics concur that the fundamental old-age insurance system is both the main and most essential part of the contemporary social security system. The government adopts laws, regulations, and other means to require enterprises and employees to pay endowment insurance in proportion and give corresponding pensions to workers after their legal retirement to maintain their life after retirement. This policy is the basic endowment insurance system.

Let us say $S(t)$ is a discrete and discontinuous function, but the increase or decrease of personal normal savings is relatively small compared with the total personal savings, especially after working for a period of time. In order to study the method, $S(t)$ is regarded as continuously

differentiable's function, which does not affect its economic and practical significance, so it can be studied by the differential method.

Set the time interval of $[t, t + \Delta t]$, of which Δt is as small as possible. The increase (decrease) of personal savings is

$$S(t + \Delta t) - S(t) = \frac{\Delta S(t)}{\Delta t} \Delta t. \quad (1)$$

Let $r(t)$ be the risk-free interest rate at t hours (replaced by the bank interest rate), α be the wage growth rate, $u_1(t)$ be the personal contribution rate, β be the consumption change rate, $q(t)$ be the replacement rate of savings replaced by personal accounts, $X_1(t)$ be the personal consumption at t hours, and $I(t)$ be the personal account pension at t hours. Assuming that wage increases, personal pension contribution increase and consumption increase are proportional to wage $W(t)$ and time Δt , and there are

$$\begin{aligned} \frac{\Delta S(t)}{\Delta t} \cdot \Delta t &= (r(t) - q(t))S(t)\Delta t \\ &+ \alpha(1 - u_1(t))W(t)\Delta t - \beta X_1(t)\Delta t. \end{aligned} \quad (2)$$

When $\Delta t \rightarrow 0$, $S(t)$ satisfies the differential equation,

$$\frac{dS(t)}{dt} = (r(t) - q(t))S(t) + \alpha(1 - u_1(t))W(t) - \beta X_1(t). \quad (3)$$

On the other hand, wages are distributed among consumption, savings, and personal account pensions:

$$X_1(t) = W(t) - I(t) - S(t). \quad (4)$$

Take derivatives of t on both the sides, and there are

$$\frac{dX_1(t)}{dt} = \frac{dW(t)}{dt} - \frac{dI(t)}{dt} - \frac{dS(t)}{dt}. \quad (5)$$

For retirement consumption, assume the individual's consumption at τ hours in the retirement period is $X_2(\tau)$, and there are

$$X_2(\tau) = u_3(\tau) \cdot W(\tau) + \frac{(1 + r_1)S_{I_1}}{T - t_1} + \frac{(1 + r_1)I_{I_1}}{T - t_1}. \quad (6)$$

Among them, S_{I_1} is the total amount of savings during the working period, I_{I_1} is the total amount of personal account principal during the working period, r_1 is the storage interest rate during the whole working period, and $u_3(\tau)$ is the replacement rate at τ hours.

The greatest utility of the basic old-age insurance is to maintain the life of retired workers. Its implementation is beneficial to the society and individuals. Its main contents are as follows: it helps to strengthen the promotion of total social productivity, mobilize the participation enthusiasm of individual workers, and continuously provide the required labor supply for social reproduction. Delaying retirement can alleviate the financial pressure of pension to a certain extent, but it cannot fundamentally solve the account gap of pension, and there may be a pension deficit question in the future. There is no inevitable connection between the pension gap and the aging of the population, and there is

great uncertainty about whether delaying retirement can make up for the pension gap. Although delaying retirement can achieve the effect of increasing income, the insured can receive more pensions in a longer life, which may offset the degradation effect of delaying retirement policy on the pension deficit. The number of employees participating in the basic old-age insurance system has been increasing, and their basic old-age insurance income has been increasing, and the total pension is gradually increasing, but the growth rate has slowed down. The appearance of prosperity cannot hide the hidden crisis of basic old-age insurance.

3.2. Optimal Retirement Age. If delaying retirement is an effective plan to solve China's pension dilemma at present, how should the best time for reform be determined? With the rapid spread of population aging, although different countries have implemented different statutory retirement age systems according to their own national conditions, facing this global question, many countries have already or will soon choose to extend their retirement time. Some scholars believe that the formulation and implementation of the delayed retirement policy cannot be rushed for a while, but it needs to wait until the pressure of old-age care becomes increasingly urgent, and the employment pressure is not obviously aggravated.

Considering that the policy of delaying retirement involves a wide range of issues and there are big differences of opinions, it is inevitable that there will be many obstacles in its implementation, and it is impossible to achieve it in one step. Therefore, most of the schemes are designed with flexible adjustment and gradual steps, so as to minimize the adverse effects caused by the reform and consider social fairness and efficiency to a greater extent. First of all, there is a mismatch between the rising life expectancy and the stagnant retirement system. Second, the education level of residents is gradually rising, and the number of years of education continues to increase, resulting in the working hours of school-age workers shrinking. Finally, in order to maintain the domestic basic pension balance, it is possible to increase the overall social pension payment rate of enterprises, which will inevitably reduce the interests of enterprises, weaken the competitiveness of the enterprises where they are located, and stabilize the average pension replacement rate.

From the international experience, most developed countries adopt a gradual and steady reform scheme based on the actual age structure of the population, which is easy for the society to digest gradually and reduce the resistance to reform. It is best to avoid raising the contribution rate of social endowment insurance for enterprises and extending the retirement age may be a more reasonable way. Some scholars also believe that the flexible retirement system should be implemented, that is, different groups of people can choose their retirement time independently according to their job nature, health, economic status, family background, and employment intention, in order to meet the pursuit of retirement life quality. The flexible retirement policy is more recognized because it makes the labor market more flexible,

and at the same time, it is conducive to the optimization of individual resource allocation and the stability of the economic environment.

4. ML and Model Building

4.1. ML. As science and technology have advanced, artificial intelligence techniques have been used more and more, and machine learning (ML) has always been the subject of people's attention as its primary component. This article provides a definition of Hadoop based on ongoing research into the technology: To have a large number of high-growth message assets with better process optimization and discovery capabilities, Hadoop needs a new processing mode. The practical use of data mining techniques can advance social science and methodology, enhance people's capacity for message processing, and be extremely important in the information age. Hadoop has clear examples of the 4 Vs, which stand for low value density, large data capacity, diverse data types, and quick data processing. Previously, the traditional data mining algorithm involved optimising the ML algorithm using the dataset. It has been challenging for this traditional ML method to meet the demand for data mining in the current large amount of heterogeneous data due to the current aspects of collection, retrieval, storage, sharing, analysis, and processing. Learning is an essential human skill, and as technology has advanced, computers are now capable of gradually learning new things. The classical neural network model is shown in Figure 1.

One goal of learning enhancement is to obtain reward function. Reward function is usually a scalar, which is an evaluation of the concrete behavior of an entity. The general reward title is set to a positive number, whereas the scalar of punishment is negative. Reward functions are usually described as the following function:

$$R_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}. \quad (7)$$

The gathering of parameters γ has a great influence on the convergence speed of the algorithm. Generally speaking, the parameter γ takes a value between 0 and 1.

$T D(0)$: The update formula of algorithm value function is as follows:

$$V(S_t) \leftarrow V(S_t) + \alpha [r_{t+1} + \gamma V(S_{t+1}) - V(S_t)], \quad (8)$$

where α represents the learning factor, γ represents the discount rate, $V(S_t)$ represents the value function of the entity at time t and state S_t , $V(S_{t+1})$ represents the predicted state value function of the entity at time $t + 1$ and state S_{t+1} , and r_{t+1} represents the instantaneous reward value obtained after the entity transitions from state S_t to state S_{t+1} after performing the behavior.

In order to improve its convergence speed, the $T D(\lambda)$ algorithm is developed, which can roll back the instant reward value obtained by the entity by any step. $T D(\lambda)$: The iterative formula of the algorithm is as follows:

$$V(S_t) \leftarrow V(S_t) + \alpha [r_{t+1} + \gamma V(S_{t+1}) - V(S_t)] e(S), \quad (9)$$

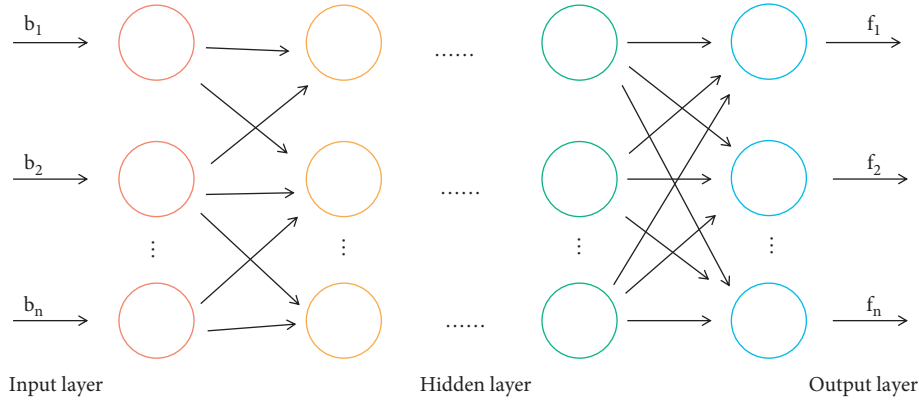


FIGURE 1: Classical neural network model.

where $e(S)$ indicates the degree of election in the state s , and λ represents a number between 0 and 1. When $\lambda = 0$, the algorithm degenerates to $T D(0)$.

$$e(S) = \begin{cases} \gamma\lambda e(S) + 1, & \text{if } s \text{ is current state,} \\ \gamma\lambda e(S), & \text{otherwise.} \end{cases} \quad (10)$$

What is ML? By independently learning the rules present in a large amount of data and gaining new experience and knowledge, ML can increase computer intelligence and give computers decision-making abilities comparable to those of humans. Designing and examining various algorithms that give computers the ability to learn on their own is the main focus of machine learning theory. In general, a machine learning algorithm uses the laws it derives from data to predict unknown data after automatically analysing the data. In reality, data mining's limited data processing ability makes it more and more challenging to analyse data. Due to its interdisciplinary nature, machine learning algorithms are able to simulate human behavior and automatically pick up new skills and knowledge. The ability to analyse data can be significantly improved by applying ML algorithms to data mining. People can learn and imitate well, but the process of learning is very difficult. ML theory underlies this alleged procedure. The practise of using computer simulation techniques to study how humans learn, innovate on previously held knowledge, enhance analysis, and find answers to problems is known as machine learning (ML).

The main goal of machine learning is to gain knowledge through extensive data analysis. The Hadoop technique has attracted the attention of an increasing number of experts in recent years. ML has developed quickly, thanks to the advancements in hardware technology, Hadoop's computing power, and storage capacity. The ability of computers and human brains to answer questions is still very different at this point in machine learning's development, which is still in its early stages. The ML technique is the main source for the study of learning mechanisms. In the current Hadoop environment, data analysis has drawn attention from a variety of industries. ML can absorb knowledge more quickly, which can successfully encourage the advancement of the machine technique. In the field of ML, reinforcement learning is a crucial algorithm. The reinforcement algorithm

has the advantage of essentially not requiring prior knowledge of the environment as it learns the best course of action for the dynamic system under consideration based on perception of the surrounding environment and primarily corrects its own behavior strategy through trial and error.

With regard to ML, there are two main research directions in its improvement: the first is the study of the learning mechanism, which pays more attention to the exploration of human learning mechanism simulation. The second is the study of effective use of message, which focuses on finding and discovering valuable and cognizable potential knowledge from huge databases. The research direction of learning mechanism originated from the ML technique. With the advent of Hadoop environment, there is a strong demand for data analysis in all walks of life. ML can acquire knowledge more quickly, which makes ML become a booster for the improvement of the machine technique. The main purpose of ML is to get corresponding conclusions from data independently. The ML algorithm is mainly the process of solving the optimal solution of a question by mathematical and statistical methods. In the current Hadoop environment, how to adopt effective learning methods is the significance of ML at present, and ML will also become a widely respected and popular learning and service technique.

4.2. Model Building. A country or society typically creates the old-age insurance system through mandatory legislation in accordance with its unique national circumstances. Workers will be guaranteed a minimal standard of living if they leave their jobs or reach the legal age for which they are required to perform compulsory labor, as specified by the applicable laws. There are two fundamental market-based pension models: one involves providing home care for the elderly and the other involves enrolling in a social pension institution. They both share service content that is market-oriented and socialised to start. Second, there are some differences between them in terms of how old-age resources are provided and how old-age care is provided. Urban employees' basic endowment insurance fund is split into individual and overall accounts. The structure diagram of the old-age insurance system is shown in Figure 2.

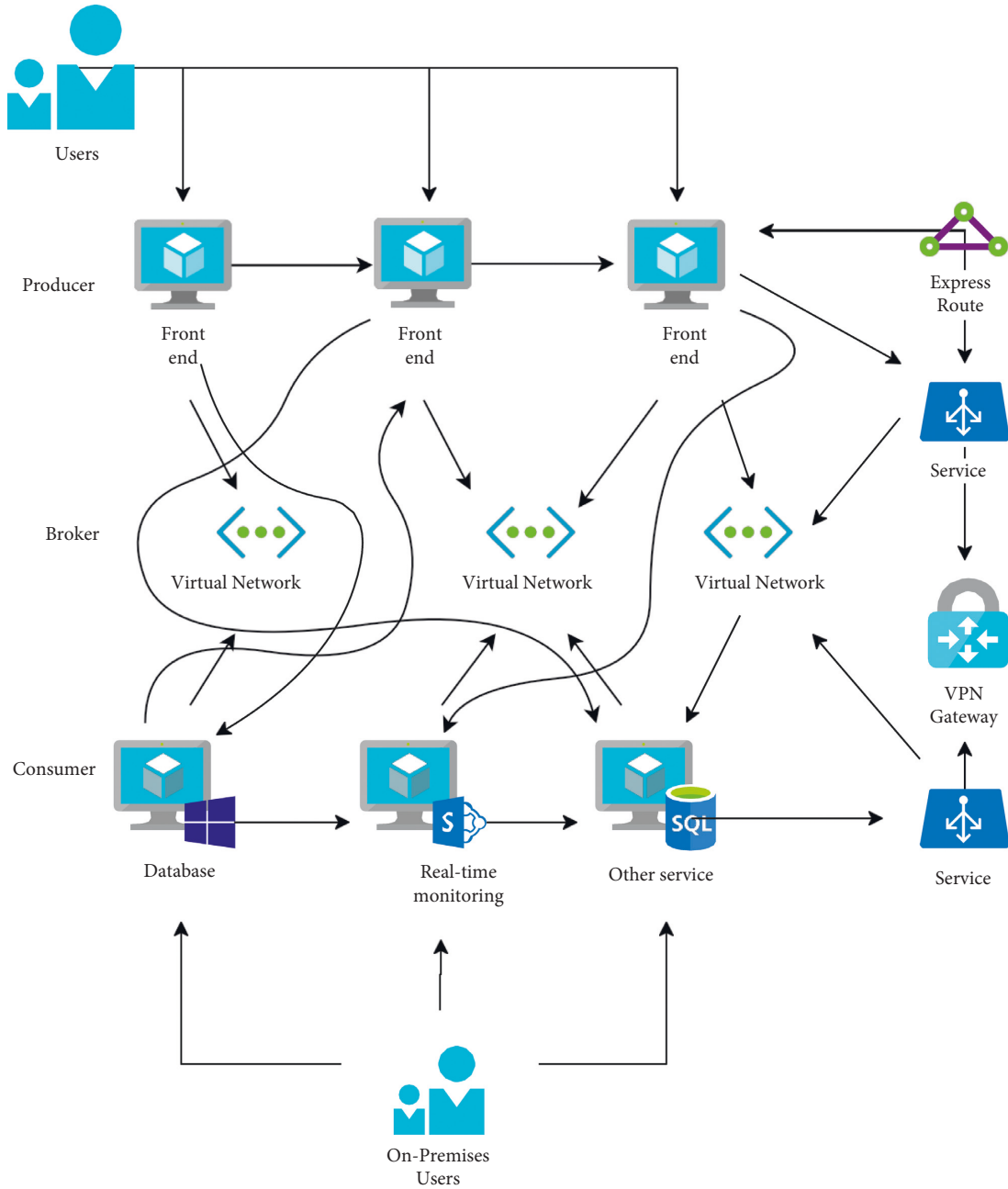


FIGURE 2: Architecture diagram of the endowment insurance system.

The overall account fund implements the pay-as-you-go system, that is, it pays retirees' pensions with the endowment insurance premiums paid by on-the-job employees in the same period, which has the function of income redistribution between generations. However, the personal account fund implements the full accumulation system, that is, individuals accumulate funds by saving during the whole employment period, and after retirement, they get all the interest income of the fund, which does not have the function of income redistribution between generations. The pension system structure is shown in Figure 3.

Social pooling specifically refers to the fact that in the same period, the pension fees paid by workers will be used for the pension collection of the retired elderly, which will be

explained by the formula in the subsequent government departments. Moreover, considering that the source of social pooling is paid by enterprises, social pension pooling will not be included in individual behavior, but in corporate behavior.

Assuming that the salary of employees in the first year of employment is w , the annual growth rate of individual wages is g , the individual contribution rate is c ($c = 8\%$), the employment period is m ($m \geq 15$), the predetermined interest rate of individual account is i ($i \neq g$), the average life expectancy is l , and the average retirement age is r and the payment time is at the beginning of each year, then the final value of each year's contribution at the end of employment is as follows.

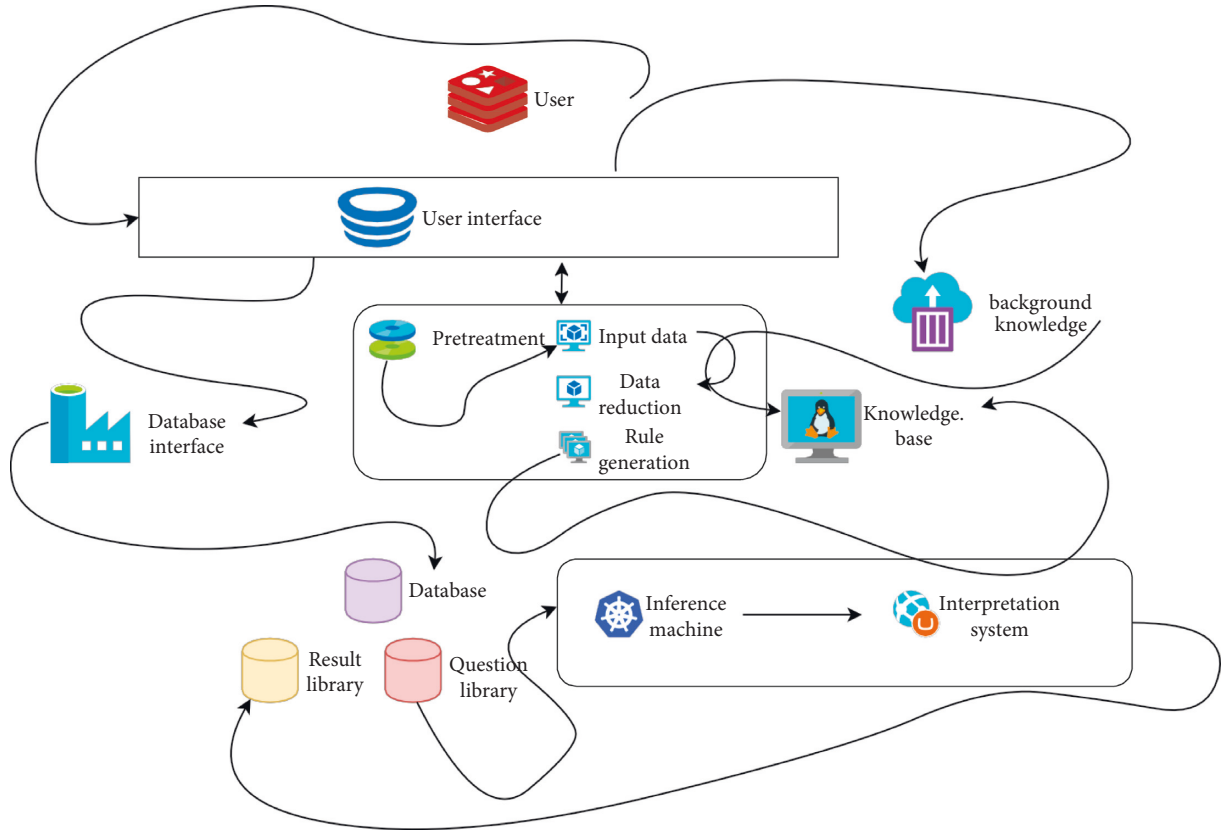


FIGURE 3: Structure of the pension system.

Final value from the beginning of the first year to the end of m : $cw(1+i)^m$.

Final value from the beginning of the second year to the end of m : $cw(1+g)(1+i)^{m-1}$.

.....

Final value from the beginning of m to the end of m : $cw(1+g)^{m-1}(1+i)$:

$$\begin{aligned}
 S &= cw(1+i)^m + cw(1+g)(1+i)^{m-1} \\
 &+ \dots + cw(1+g)^{m-1}(1+i) \\
 &= cw(1+i) \frac{(1+i)^m - (1+g)^m}{i-g}.
 \end{aligned} \tag{11}$$

The salary of an employee in the year before retirement can be expressed as $w(1-g)^{m-1}$.

By

$$c = 8\%, b = \frac{s}{(l-r)}. \tag{12}$$

The replacement rate of personal account pension is

$$\begin{aligned}
 T &= \frac{b}{w(1+g)^{m-1}} = \frac{S/l-r}{w(1+g)^{m-1}} \\
 &= \frac{0.08w(1+i)(1+i)^m - (1+g)^m/i-g}{(l-r)w(1+g)^{m-1}} \\
 &= \frac{0.08(1+i)}{l-r} \frac{[(1+i)^m - (1+g)^m]}{(1+g)^{m-1}(i-g)}.
 \end{aligned} \tag{13}$$

The goal of home-based care for the elderly is to mobilize all social forces and create a home-based care system that is most in line with the preferences of the elderly, places the family at its center, is supported by the community care service network, and is insured by the old-age insurance system. However, the institutional pension is entirely provided by society in terms of the funding source, the nature of the service, and the method of operation. This means that the financial support and life services for the elderly in their later years are provided by society, including pension, medical expenses, welfare expenses, relief expenses, and life care for the elderly, which are provided by social security

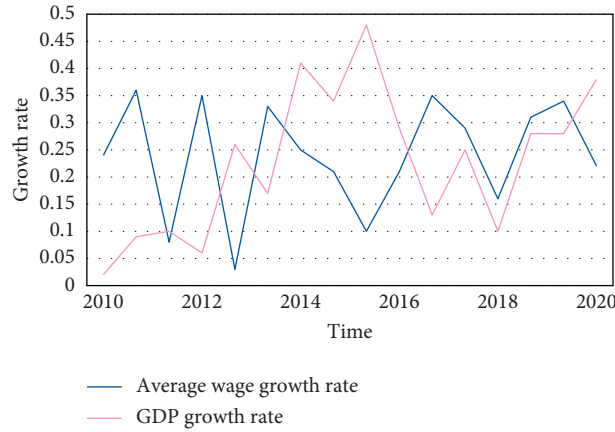


FIGURE 4: Trends of GDP growth rate and average wage growth rate.

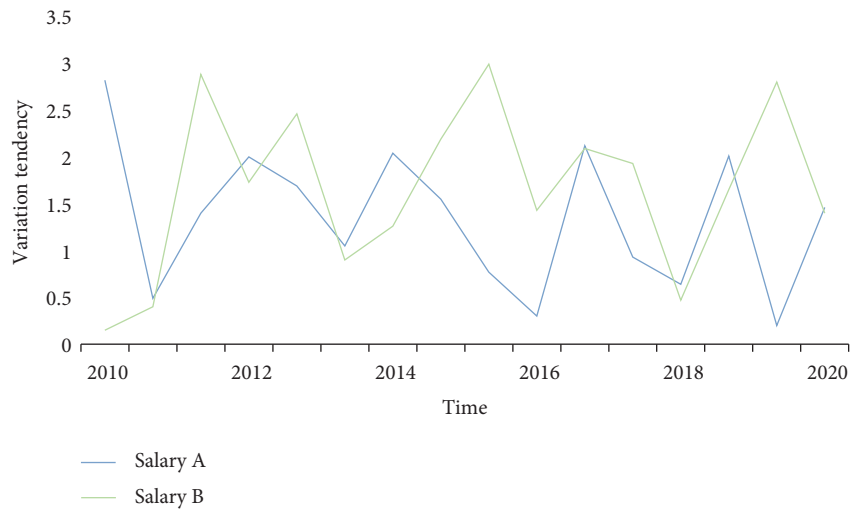


FIGURE 5: Salary scatter chart.

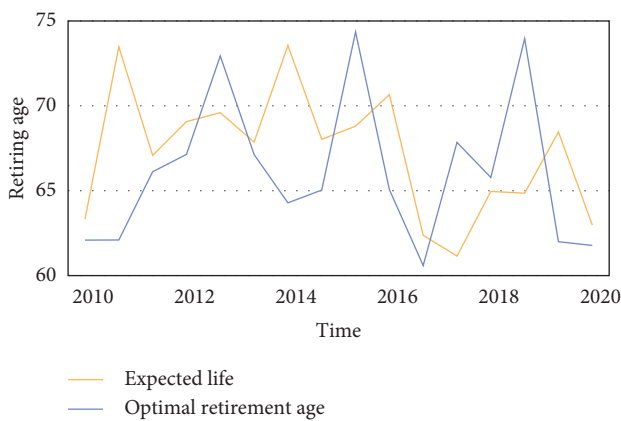


FIGURE 6: Comparison of optimal retirement age.

institutions, governments at all levels, businesses, and institutions. Figures 4, 5, and 6 show how the continuous rise in life expectancy results in a more serious aging situation. The ideal retirement age and life expectancy are positively correlated, with the ideal retirement age growing at a rate 12.57 percent higher than life expectancy.

Living forms include nursing homes, nursing homes, and care for the elderly. It is characterized by centralized pension. The old-age insurance system is mainly about the management and operation of pensions, which are all based on mathematical knowledge and operated through actuarial management. The old-age insurance system is a comprehensive subject integrating economy, society, and management, involving many vital concepts. The construction of social welfare function is mainly based on the maximization of social welfare. As a policy maker, the government should draw up a retirement system with reference to this premise. The basic old-age insurance adopts the mode of unified accounting, which forms the “hidden debt” of the old-age insurance for various historical reasons. From Tables 1, 2, and 3 and Figures 7, 8 and 9, it can be seen that although the implementation of the two-child policy has improved the total fertility rate as a whole, the proportion of women of childbearing age continues to decline and China’s population growth rate will slow down in the future. At the same time, the changing trend of the optimal retirement age is consistent with the population growth rate, so the optimal retirement age is positively related to the population growth rate.

TABLE 1: Population growth rate under the two-child policy.

	2010	2012	2014	2016	2018	2020
Estimated growth rate	0.69	0.09	0.34	0.59	0.84	0.43
Actual rate of growth	0.11	0.69	0.54	0.05	0.78	0.57

TABLE 2: GDP ratio comparison.

	2010	2012	2014	2016	2018	2020
Estimated GDP	0.91	0.06	0.94	0.15	0.4	0.54
Real GDP	0.3	0.11	0.7	0.07	0.06	0.13

TABLE 3: Simulation of optimal retirement age.

	2010	2012	2014	2016	2018	2020
Optimal retirement age 1	63.46	63.96	64.5	61.27	60.67	63.21
Optimal retirement age 2	62.76	64.89	60.28	61.65	63.01	62.17

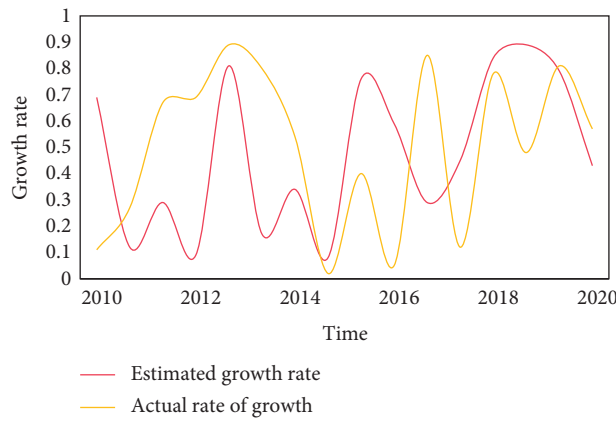


FIGURE 7: Population growth rate under the two-child policy.

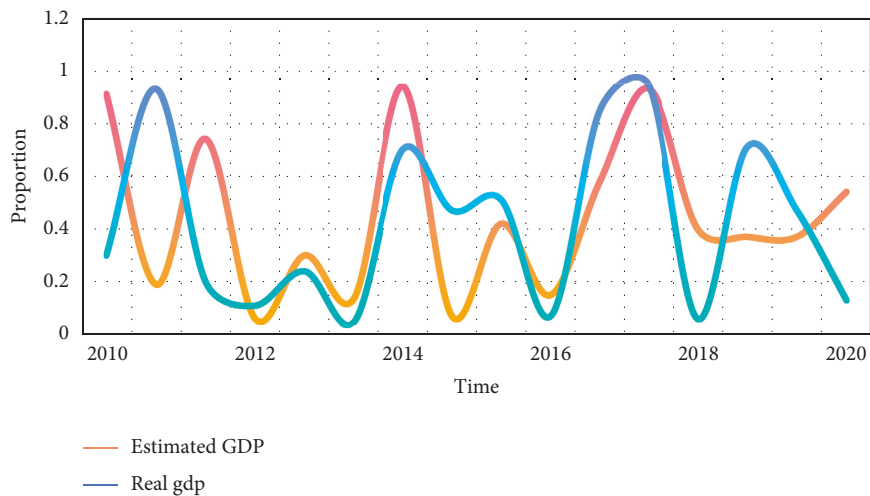


FIGURE 8: GDP ratio comparison.

In order to make up for the gap of endowment insurance fund, the government misappropriated the fund of individual account, resulting in serious deficit of individual account and a series of “mixed accounts” questions, which caused great concern of the state and provinces and cities.

The principle of raising pension funds must follow the balance between collection and payment; in other words, the raising of pension funds should keep the basic financial balance with the payment of pension expenses according to regulations.

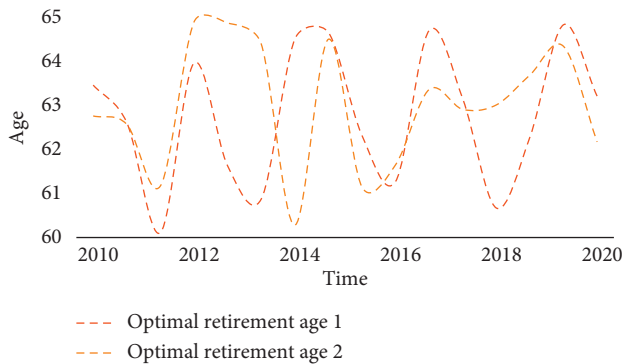


FIGURE 9: Simulation of optimal retirement age.

5. Conclusions

Many nations run the risk of widening the gap between income and pension expenditures as a result of the population's rapid aging. The average wage growth rate is a factor in both the income and expenditure models of the overall pension account. Raising the wage growth rate will raise the pension level for retirees while also raising the payment standard for employees who are still working. However, because the pension benefits received by retirees are only increased to the extent that the wages of the on-the-job employees have increased, after raising the average wage growth rate, the increased income of the pension fund is naturally higher than the expenditure, which not only significantly lessens the pressure on the pension fund's ability to make payments but also serves as an obvious buffer during the implementation of the deferred retirement policy. The ideal retirement age is a dynamic equilibrium age that depends on a variety of factors rather than being a fixed age. People who started working later and those who saved less early on tend to retire later, given the assumption that the ideal retirement age is earlier than the voluntary retirement age. This study's findings suggest that the ideal retirement age and life expectancy are positively correlated, with the ideal retirement age's growth rate 12.57 percent higher than that of life expectancy.

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 there are no conflicts of interest.

References

- [1] "The health implications of social pensions: evidence from China's new rural pension scheme[J]," *Journal of Comparative Economics*, vol. 46, no. 1, pp. 53–77, 2018.
- [2] F. Menoncin and E. Vigna, "Mean-variance target-based optimisation for defined contribution pension schemes in a stochastic framework," *Insurance: Mathematics and Economics*, vol. 76, no. 4, pp. 172–184, 2017.
- [3] M. Knell, "Increasing life expectancy and NDC pension systems," *Journal of Pension Economics and Finance*, vol. 17, no. 2, pp. 170–199, 2018.
- [4] A. F. Gouveia, "Political support for reforms of the pension system: two experiments," *Journal of Pension Economics and Finance*, vol. 16, no. 3, pp. 371–394, 2017.
- [5] I. Abinzano, L. Muga, and R. Santamaría, "Bad company. The indirect effect of differences in corporate governance in the pension plan industry," *International Review of Financial Analysis*, vol. 54, no. November, pp. 63–75, 2017.
- [6] Y. Cheng, C. Li, and L. A. Johnston, "The intergenerational education spillovers of pension reform in China," *Journal of Population Economics*, vol. 31, no. 3, pp. 671–701, 2018.
- [7] J. Brzezczynski, M. T. Bohl, and D. Serwa, "Pension funds, large capital inflows and stock returns in a thin market," *Journal of Pension Economics and Finance*, vol. 18, no. 3, pp. 347–387, 2019.
- [8] D. Blake, L. Sarno, and G. Zinna, "The market for lemmings: the herding behavior of pension funds," *Journal of Financial Markets*, vol. 36, no. nov, pp. 17–39, 2017.
- [9] D. H. Chen, R. M. Beetsma, D. W. G. A. Broeders, and A. A. Pelsner, "Sustainability of participation in collective pension schemes: an option pricing approach," *Insurance: Mathematics and Economics*, vol. 74, no. MAY, pp. 182–196, 2017.
- [10] Z. Chen, Z. Li, Y. Zeng, and J. Sun, "Asset allocation under loss aversion and minimum performance constraint in a DC pension plan with inflation risk," *Insurance: Mathematics and Economics*, vol. 75, no. jul, pp. 137–150, 2017.
- [11] C. T. Kreiner, S. Leth-Petersen, and P. E. Skov, "Pension saving responses to anticipated tax changes: evidence from monthly pension contribution records," *Economics Letters*, vol. 150, no. 1, pp. 104–107, 2017.
- [12] E. Naumann, "Do increasing reform pressures change welfare state attitudes? An experimental study on population ageing, pension reform preferences, political knowledge and ideology," *Ageing and Society*, vol. 37, no. 2, pp. 266–294, 2017.
- [13] M. L. Tang, S. N. Chen, G. C. Lai, and T. P. Wu, "Asset allocation for a DC pension fund under stochastic interest rates and inflation-protected guarantee," *Insurance: Mathematics and Economics*, vol. 78, no. JAN, pp. 87–104, 2018.
- [14] A. Kerner, "Pension returns and popular support for neoliberalism in post-pension reform Latin America," *British Journal of Physical Education*, vol. 50, no. 2, pp. 1–36, 2018.
- [15] M. D. Fitzpatrick, "Pension-spiking, free-riding, and the effects of pension reform on teachers' earnings," *Journal of Public Economics*, vol. 148, no. APR, pp. 57–74, 2017.
- [16] T. J. Lu and N. Tang, "Social interactions in asset allocation decisions: e," *Journal of Economic Behavior & Organization*, vol. 159, no. MAR, pp. 1–14, 2019.
- [17] M. Hatcher, "Should a pension reform be announced? A reply [J]," *Economics Letters*, vol. 183, no. 9, Article ID 108583, 2019.
- [18] J. A. Turner, G. Hughes, and A. Chlon-Dominczak, "Improving pension income and reducing poverty at advanced older ages: longevity insurance benefits in Ireland and Poland as models for the United States," *Journal of Portfolio Management*, vol. 44, no. 7App, pp. 46–51, 2018.
- [19] M. Bertoni, G. Brunello, and G. Mazzarella, "Does postponing minimum retirement age improve healthy behaviors before retirement? Evidence from middle-aged Italian workers - ScienceDirect," *Journal of Health Economics*, vol. 58, no. 1, pp. 215–227, 2018.

- [20] L. Yue, "Paradoxical effects of increasing the normal retirement age: a prospective evaluation - ScienceDirect," *European Economic Review*, vol. 101, no. 4, pp. 512–527, 2018.
- [21] H. J. Chen, "Fertility, retirement age, and pay-as-you-go pensions," *Journal of Public Economic Theory*, vol. 20, no. 6, pp. 944–961, 2018.
- [22] B. Den, S. A. Zijderveld, and J. Bruers, "Preferred and actual retirement age of oral and maxillofacial surgeons aged 55 and older in The Netherlands: a longitudinal study from 2003 to 2016," *Human Resources for Health*, vol. 16, no. 1, p. 25, 2018.
- [23] A. Meng, M. A. Nexø, and V. Borg, "The impact of retirement on age related cognitive decline – a systematic review," *BMC Geriatrics*, vol. 17, no. 1, p. 160, 2017.
- [24] N. Steiber and M. Kohli, "You can't always get what you want: actual and preferred ages of retirement in Europe," *Ageing and Society*, vol. 37, no. 2, pp. 352–385, 2017.

Research Article

A Novel Adaptive Affective Cognition Analysis Model for College Students Using a Deep Convolution Neural Network and Deep Features

Huali Feng 

Data and Information Center, Wuxi Vocational Institute of Commerce, Wuxi 214153, Jiangsu, China

Correspondence should be addressed to Huali Feng; fenghuali@wxic.edu.cn

Received 21 July 2022; Revised 11 August 2022; Accepted 17 August 2022; Published 27 August 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Huali 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.

Currently, under the impact of the COVID-19, college students are facing increasingly elevated employment pressure and higher education pressure. This can easily cause a huge psychological burden on them, causing affective cognition problems such as anxiety and depression. In the long run, this is not conducive to students' physical and mental health, nor is it conducive to the healthy development of the school and even the whole society. Therefore, it is imperative to build a novel adaptive affective cognition analysis model for college students. In particular, in the context of smart cities and smart China, many universities have opened the smart campus mode, which provides a huge data resource for our research. Due to problems of the low real-time evaluation and single data source in traditional questionnaire evaluation methods, evaluation errors are prone to occur, which in turn interferes with subsequent treatment. Therefore, for the purpose of alleviating the above deficiencies and improving the efficiency and accuracy of the affective cognition analysis model of college students, this paper studies the adaptive affective cognition analysis method of college students on basis of deep learning. First, because students' psychological problems are often not sudden, on the contrary, most of these abnormalities will leave traces in their daily activities. Therefore, this paper constructs a multisource dataset with the access control data, network data, and learning data collected from the smart campus platform to describe the affective cognition status of students. Second, the multisource dataset is divided into two categories: image and text, and the CNN model is introduced to mine the psychological characteristics of college students, so as to provide a reference for the subsequent affective cognition state assessment. Finally, simulation tests are developed to confirm the viability of the technique suggested in this research. The experiments demonstrate that the accuracy of the assessment model is significantly increased because it can fully reflect the heterogeneity and comprehensiveness of the data. This also highlights that the new method has a wide range of potential applications in the modern campus setting and is also helpful in fostering the accuracy and depth of college students' work on their affective cognition.

1. Introduction

Because life is moving at an increasingly rapid speed, people's mental health is deteriorating, and more and more people are facing depression, autism, insomnia, and other psychological problems [1]. Simultaneously, under the impact of the COVID-19, college students are facing increasing pressure for higher education and employment, leading to an increasing number of college students suffering from different degrees of depression and anxiety disorders. In other words, various mental health problems have slowly

penetrated into all aspects of college students' daily life, so that their bodies and psyches are suffering from different degrees of destruction [2]. Currently, according to reports, there have been an increasing number of extremely dangerous occurrences involving college students brought on by psychological issues, such as suicide, self-harm, and fights and brawls occur frequently [3]. Therefore, paying attention to the college students' affective cognition and studying a set of scientific and effective affective cognition analysis methods is one of the problems that colleges and universities urgently need to solve. This not only helps ensure the safety

of students' lives during school, but also helps them complete their studies smoothly. It can also provide support for the promotion of smart mental health education in universities.

With the aid of various digital and intelligent technologies, universities are gradually transitioning to the age of the smart campus. Relying on these technologies, a smart campus is an effort to create an intelligent teaching environment, intelligent management mode, and digital teaching resources to form a new ecology of learner-centered education and teaching, which is an inevitable trend for digital school reform [4, 5]. In the intelligent campus environment, students' life and learning data will be recorded into various management systems. These massive data contain a huge amount of information and provide a better opportunity to grasp their psychological and affective cognition state. It can be said that the flourishing development of smart campus will open a new door for the study of college students' affective cognition. At the present stage, related research still rely on the modes of questionnaires and structured interviews, which have many shortcomings, such as the insufficient amount of data, single data source, easy to conceal the truth and mislead the assessment results, lack of timeliness of assessment results, and passive assessment work [6]. In response to the drawbacks of traditional assessment methods, some scholars have tried to do some improvement work. They used social network data in affective cognition assessment studies and realized automatic affective cognition analysis with the help of trained models, and achieved relatively satisfactory assessment results [7]. However, the shortcoming is that these methods still face problems such as difficulties in collecting data on network behaviors and incomplete information on single-modal data, which make it difficult to judge the affective cognition status of the assessed person accurately and comprehensively.

Technology empowers education, and the ability to autonomously assess students' affective cognition state has been made available by the deep learning technology's quick development. After years of building information application systems, schools have accumulated massive amounts of data on teaching, consumption, and student behavior. With the continuous accumulation of these data, if the sentiment tendencies of students can be mined from these multisource data, it will help to know the affective cognition status of students more accurately. If the manual labeling method is still used to deal with such a large amount of data, it is indeed labor and time intensive. Moreover, the accuracy of the information obtained is low. In view of this, for the purpose of improving the accuracy and effectiveness of college students' affective cognition analysis model, it is very important to use deep learning architecture to classify and predict related data in practical applications. Deep learning has been one of the key directions of academic research recently. Additionally, it has made significant strides in the research of college students' affective cognition [8–10]. Many scholars have proposed to use deep neural network models such as CNN [11] and Bi-LSTM [12] to analyze and train the

influencing factors affecting students' affective cognition problems and establish analytical models to classify and predict students' affective cognition problems. In addition, the literature [13] and the literature [14] applied the massive data recorded by the relevant systems on campus to mine the behavioral data related to students' affective cognition. Meanwhile, artificial intelligence algorithms are used to intelligently identify students' abnormal behavior and construct early warning models. The affective cognition problems of college students are detected in time and precise psychological counseling is provided for them.

The above research shows that deep learning technology provides a new perspective for students' affective cognition work to take the road of precision and wisdom. In view of this, this study closely combines college students' affective cognition work with the construction work of smart campus. In the research, the advantages of deep learning technology are fully utilized and based on the multisource data such as access control data, learning behavior data, and network data recorded in the smart campus platform, an adaptive affective cognition analysis model for college students is constructed for the purpose of improving the effectiveness and accuracy of psychological state assessment of college students. The research work mainly includes: first, carefully studying relevant domestic and foreign literature, closely combining the multisource data recorded in the smart campus environment, including access control data, learning behavior data, and network data, and building a multisource dataset. Second, on the premise of information security, the multisource data set is divided into two categories: image and text, and the CNN model is introduced to process and analyze these data. In this way, the hidden affective cognition problems of students in these data can be excavated, and the mental health status of college students can be accurately identified. Finally, simulation tests are developed to confirm the viability of the technique suggested in this research. The experiments demonstrate that the novel technique has a higher identification accuracy when assessing the affective cognition state of college students compared with the data of a single modality. It can accurately grasp students' mental health status and has broad application prospects in the smart campus environment.

2. Related Theory

2.1. Artificial Intelligence and Deep Learning. Artificial intelligence (referred as "AI") allows machines to simulate or realize the process of human learning behavior by computing and analyzing some human thinking, consciousness, and behavior with the help of mathematical tools [15]. Currently, as new technologies and industries arise, AI has been given more expectations and heavy responsibilities. At present, it has made great achievements in iris recognition, face recognition, intelligent search, and other fields [16, 17]. The scope of artificial intelligence is shown in Figure 1.

As one of the representative technologies in the field of AI, machine learning studies human learning behavior and

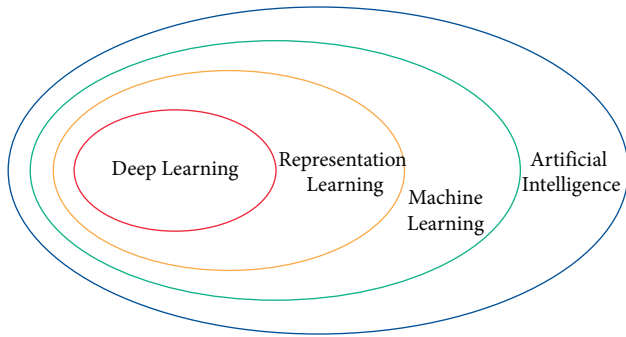


FIGURE 1: Scope of artificial intelligence.

is the core of artificial intelligence. It mainly refers to the computer through the automatic analysis of data, facts, or its own experience, and then comprehensively acquire new knowledge or skills, so that the computer can have the ability to automatically learn specific knowledge and skills, and establish a task-oriented learning system with specific applications [18]. It can be said that it is the fundamental way to make computers intelligent. As a subset of machine learning, deep learning aims to achieve the universalization of artificial intelligence. Deep learning was first proposed by Hinton et al. to study the optimal representation of information and its acquisition method. In the case of neural networks or belief networks, it is the process of machine learning of the mapping between inputs and outputs based on deep structures or network representations [19]. Currently, deep learning is a hold expression for a group of pattern recognition techniques. Convolutional neural networks, multilayer neuron-based autoencoder neural networks, and deep belief networks are basically the three forms of convolutional neural networks that are relevant to the specific study material. Because deep learning makes it possible for machines to accurately simulate some aspects of human society and is useful for many complex recognition patterns, it has greatly contributed to the development of related fields such as artificial intelligence. In the last 30 years or so, research on this topic has attracted the close attention of many scholars, and significant progress has been made in areas such as natural language processing [20], image processing [21], data mining [22], and machine translation [23]. Figure 2 shows the difference in principle between deep learning and several other machine learning techniques.

2.2. Deep Neural Network. The specific process of deep learning can be briefly described as: train a multilayer neural network to mine the inherent laws and connections of the given sample data. Extract and analyze the characteristic information of samples, such as images, texts, and sounds. Process data information and issue instructions to control the behavior of the machine, so that the machine has the capabilities of learning, analysis, recognition, and processing similar to humans. As a representative of deep learning technology, Deep Neural Network (referred as DNN) has received wide attention since its introduction. Based on the

traditional neural network, it expands the hidden layer to multiple layers, which can better deal with the problem of the network in dealing with complex functions, and greatly improve the performance of the model [24]. The network structure of DNN is shown in Figure 3.

It has been shown that the training process of traditional neural networks can be essentially divided into two key steps forward computation and backward propagation. And through repeated loops, iterates continuously until the final convergence condition is reached, the model training ends, and the training results are output. The difference between DNN and traditional neural networks is that the former introduces several hidden network learning units between the input layer and the output layer, thereby greatly improving the training efficiency of the model and the ability to process complex data.

3. An Adaptive Affective Cognition Analysis Model Based on Deep Learning

The affective cognition analysis model for college students based on deep learning covers the following two core steps: (1) On the basis of the smart campus platform, we collect multiple sources of student behavior data to build an adaptive affective cognition analysis system for college students, including access control data, learning data and network data, so as to more comprehensively portray the psychological indicators of students. (2) The multisource data are divided into two categories: image and text, and then the CNN model is applied to mine the students' emotions and psychological changes hidden in these data, so as to accurately assess the students' affective cognition level. Figure 4 illustrates the framework of the assessment model based on multisource data.

As shown in Figure 4, the general process of our research model consists of three main parts, which are sentiment calculation based on text data, sentiment calculation based on image data, and sentiment calculation of image-text fusion data. First, in this study, we collected three types of data sources access control data, learning data, and network data and preprocessed these multisource data. Second, we divide the collected multisource data into two categories image and text and then compute the affective values using CNN models, respectively. Finally, in the model training and recognition stage, the text and image affective values are fused and calculated using the maximum value rule. At the same time, it is compared with the constructed affective cognition analysis model, and finally, the mental health level is output.

3.1. Multisource Data. With the rise of smart campuses, a large amount of student behavior data has been preserved, which records the bits and pieces of students' study and life during school. These data are both in the form of images and texts, which can feedback a student's psychological changes in a certain period of time from different perspectives. By mining and analyzing these data, it helps to assess students' current affective cognition conditions more comprehensively and

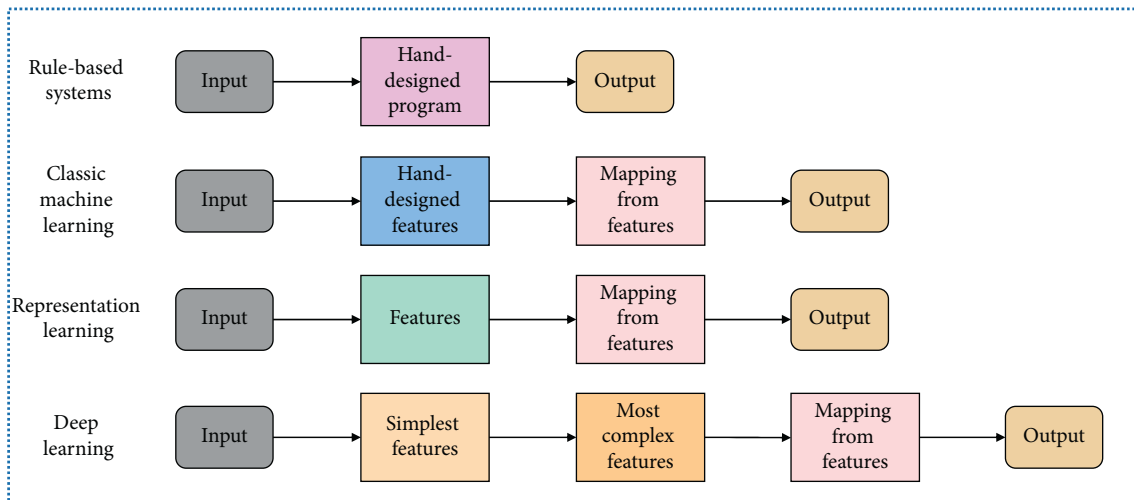


FIGURE 2: Differences between deep learning and other machine learning algorithms.

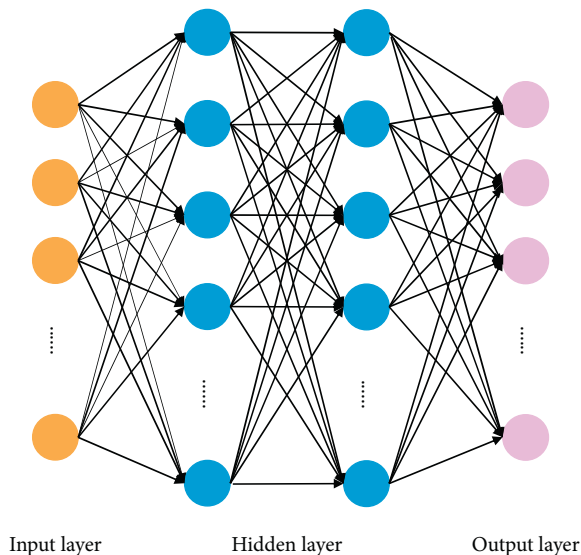


FIGURE 3: Deep neural network structure.

systematically. Therefore, collecting data that can accurately represent the features of college students' affective cognition is a necessary first step in analyzing their mental health. Figure 5 depicts the complete multisource data system.

As shown in Figure 5, our research focuses on the following data sources:

3.2. Network Data. Research has shown that there is a relationship between online behavior and psychological traits such as personality, mood, and depression. Using students' online data can graphically portray their psychological and affective cognition characteristics. From a psychological point of view, the information posted by college students on forums, the content of social chats, the music playlists they listen to, and the students' online active time are all affected by the students' affective cognition. Web logs record these data of students' online characteristics, and how to dig out

students' affective cognition characteristics from these records is the problem we want to solve.

3.3. Earning Behavior Data. In the field of education research, significant negative correlations were found between mild, moderate, and severe depression and academic performance. The student's learning behavior data reflect the student's past learning state and process. Introducing these data can be helpful to our research. In a smart campus environment, cameras in the classroom can be utilized to capture data on students' learning behaviors in the classroom, such as sleeping in class, playing with cell phones, depressed mood, and other abnormal behaviors.

3.4. Access Control Data. Similarly, some surveys found that there is a strong link between the level of interpersonal relationships and the level of affective cognition. As a result, we can identify students who have mental health issues by extracting indicators that reflect students' social relationships. Access control data show more than just the timing of students entering and exiting the dormitory. Often, using the image data recorded by the access control system of students entering and leaving the dormitory with their friends or roommates, mining in this direction may also reveal information of interest to us such as their interpersonal relationships. Therefore, our research also benefits from the mining of students' social relationship features using access control data.

In summary, with the help of the data center in the smart campus, the multisource data that may contain the psychological and affective cognition characteristics of students can be collected and integrated to form a 360-degree portrait of students. Based on the multisource data, we can establish an affective cognition analysis model for college students to detect problems in time and provide proactive early warning to teachers and administrators, so as to realize the development of students' psychological problems from "passive management" to "active prevention".

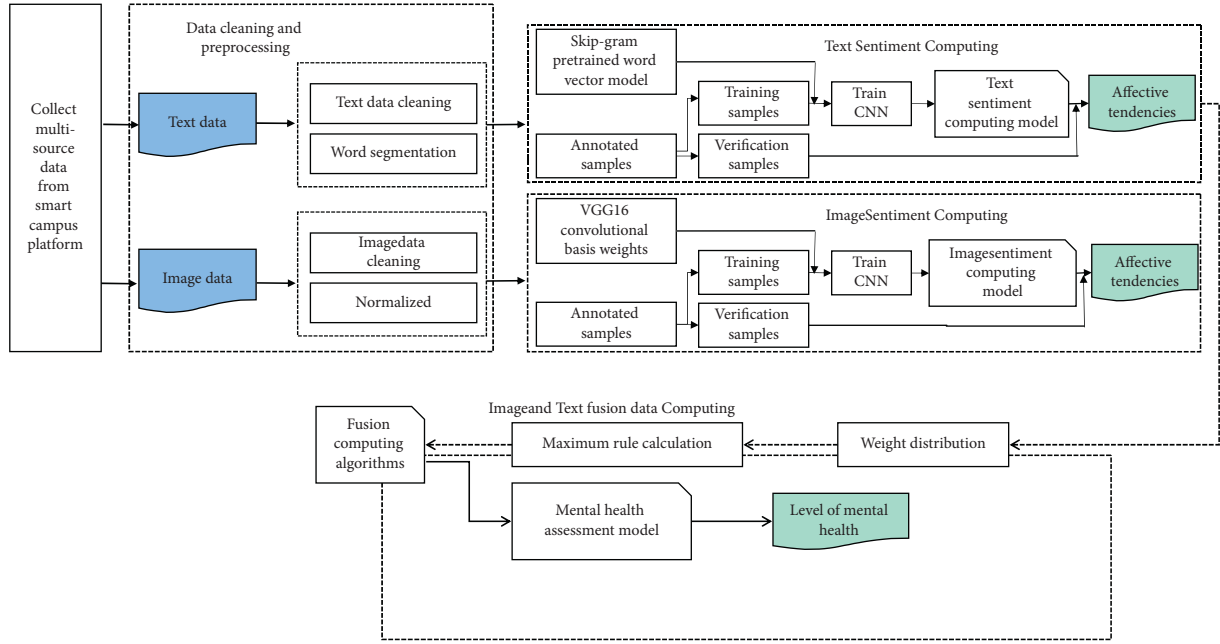


FIGURE 4: Framework of affective cognition analysis model based on multisource data.

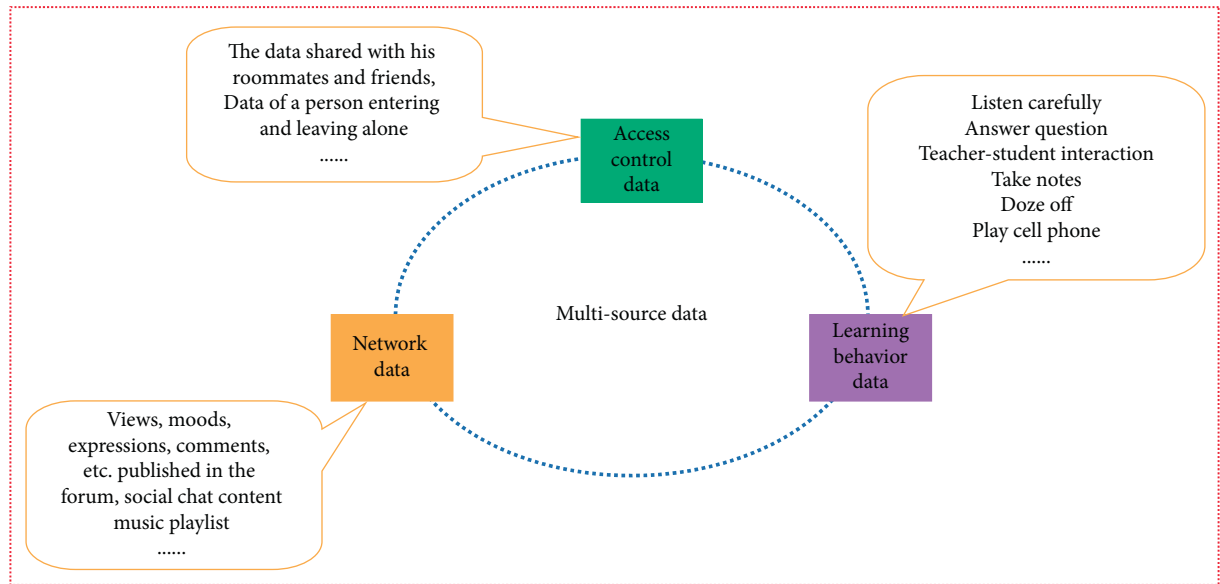


FIGURE 5: Multisource data system.

3.5. CNN Network Model

3.5.1. Model Training Based on Text Data. The CNN training model based on text data applied in this study is shown in Figure 6. It contains several convolutional layers, pooling layers, and a fully connected layers.

First, in the input layer, the collected text data needs to be word partitioned. That is, the input text content is segmented into word vectors of length N using Word2Vec technique [25], as shown in formula (1).

$$T = [t_1, t_2, \dots, t_N], t_i \in R^{l \times h}. \quad (1)$$

Then, in the convolution layer, the word vectors in formula (1) are further processed one by one using the convolution operation, and the detailed process is illustrated in formula (2) and formula (3).

$$\langle t_{0:k-1}, t_{1:k}, \dots, t_{N-k+1:N} \rangle, \quad (2)$$

$$f_j = \partial(\varphi(\omega \bullet T_{i-j+k-1} + \beta)), \quad (3)$$

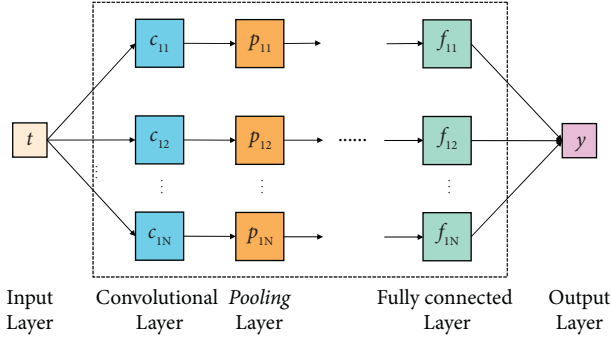


FIGURE 6: CNN network structure.

where φ represents the convolution kernel function and f_j indicates the eigenvalues obtained after convolution.

The convolution result is further de-linearized, and the link requires the introduction of an activation function, as shown in formula (4). The processed structure is then stitched to obtain the feature matrix F , as shown in formula (5).

$$\delta(t) = \frac{1}{1 + e^{-t}}, \quad (4)$$

$$F = f_1 \oplus f_2 \oplus \dots \oplus f_{N-k+1}. \quad (5)$$

In the pooling layer, this study utilizes the maximum pooling method to reduce the dimensionality of the feature matrix obtained from the convolutional layer to prevent the overfitting problem and to increase the operational effectiveness of the network simultaneously. The method is shown in formula (6).

$$F = \{\max(f_1), \max(f_2), \dots, \max(f_{N-k+1})\}. \quad (6)$$

The fully connected layer further processes the features extracted by the pooling layer to obtain the final feature vector, as shown in formula (7).

$$H = \sum_{i=1}^{N-k+1} (\omega \bullet F_i + \beta_i). \quad (7)$$

Finally, in the output layer, the final feature vector is predicted for classification using the softmax function and the computed results are output. The detailed process is shown in formula (8) and formula (9).

$$y = \text{soft max}(V \bullet H + \lambda), \quad (8)$$

$$\text{soft max}(t) = \frac{e^t}{\sum_{k=1}^K e^t}. \quad (9)$$

3.5.2. Model Training Based on Image Data. As an important supplement to textual information, accurate identification of the psychological and affective cognition characteristics of students embedded in image data is also very important. Image-based affective computing is a challenging visual problem. Only by learning a large number

of parameters and effective features, the CNN model can precisely calculate the psychological sentiment contained in image data. This study utilizes the VGG16 convolutional base as a pretrained model for image affective tendencies computing to learn general feature representations and abstract feature representations for images. Meanwhile, the original dense connection layer settings are changed to suit the task of image affective computing. This can significantly increase the CNN model's applicability and reduce the chance of overfitting during model training.

We initially utilize the capabilities of several convolutional layers in the VGG16 convolutional base to learn the sentiment representation of images during the training of the CNN model. Then, the extracted image affective features are integrated and classified by the dense connection layer. Finally, the affective tendency value of the image is output.

3.5.3. Computation of Multisource Data Using the Maximum Value Rule. Finally, this paper adopts the maximum value rule to compute the affective tendency value of text and image data in order to precisely analyze the level of college students' affective cognition. In this way, it fully takes into account the affective characteristics of these two types of data. Formulas (10) and (11) illustrate the precise computation procedure.

$$Y'_j = \max_i (Y_{ij}(N)), \quad i = 1, 2, j = 1, 2, \quad (10)$$

$$Y_j(N) = \frac{Y'_j(N)}{\sum_j Y'_j(N)}, \quad (11)$$

where i and j indicate the number of classifiers and categories, respectively, and $Y_j(N)$ represents the probability value of the j th sentiment category.

4. Experimental Test

4.1. Data Set and Data Processing. To evaluate how well the method put forward in this work performs when used in practice, we recruited subjects from college students in a university in Wuxi to design simulation tests. Based on the principle of voluntary enrollment, a total of 200 subjects were recruited for the experiment. They were then given depression self-assessment questionnaires designed according to the CES-D scale. Then, the collected questionnaire data was cleaned, and the affective cognition grades of the experimental subjects were marked according to the scores of the questionnaires, which were divided into three categories: healthy, possible depression, and existing depression. The distribution of the three cases was 135 (67.5%), 42 (21%), and 23 (11.5%). In addition, with the consent of 200 subjects, a confidentiality agreement was signed with them to collect the behavioral data of the subjects from the school's smart platform, including access control data, learning data, and Internet access data, to construct a multisource data set. What's more, experts in related fields were invited to label these multisource data according to their emotional disposition and classify them

into two categories: positive emotions and negative emotions. Thus, a joint annotation dataset for psychological assessment is formed, and Table 1 illustrates the detailed data distribution.

Whether it is the data of students' affective cognition level collected in the form of questionnaires or the multisource data about students' behavior collected by the smart platform, they cannot be directly used in the model for psychological assessment. Therefore, data cleaning needs to be performed on these raw data, including questionnaire data processing and text and image data processing in multisource data. Finally, the cleaned data are preprocessed, i.e., they are transformed into data types that can be recognized by the CNN model. For example, for text data, the irrelevant symbols are removed and the fonts are converted. For image data, format conversion, size adjustment, and normalization are performed.

5. Results and Analysis

In the simulation experiments, three representative evaluation metrics are utilized to verify the evaluation effect of the model, which are Precision, Recall, and F1-Measure.

First, to verify the effectiveness of the proposed method in the application of affective cognition assessment for college students in this paper, the following simulation experiments are designed: (1) Calculate the assessment results based on text data using CNN model. (2) Calculate the assessment results based on image data using CNN model. (3) Calculate the assessment results based on the fusion of text and image using the maximum rule. Figure 7 illustrates the assessment results for different data forms.

Observing Figure 7, we can conclude the following:

- (1) Overall, the proposed method of fusing multisource data in this paper achieves relatively satisfactory assessment results in all three evaluation indexes, and the results reach more than 85%. This indicates that, due to the utilization of the maximum value rule calculation in the new method, the advantages of text data and image data can be fully utilized, allowing the two to form a complementary relationship, thus overcoming the drawback of single data in the traditional assessment method and greatly improving the assessment accuracy of the model.
- (2) Second, among the assessment results of three different data forms, the assessment result based on fusion data has the best performance, which is significantly better than that of text data and image data. Moreover, the assessment results based on text data are also more desirable than those based on image data. However, despite this, the influence of image data on the assessment results cannot be ignored. On the contrary, because image data can be a key supplement to text data, it can sometimes express students' emotions and feelings more directly and graphically, and even convey some mental states that cannot be depicted by textual language. In this study, we integrate them with text data to fully exploit the

TABLE 1: Distribution of data in multisource dataset.

Data form	Training set	Verification set	Test set
Text	2000	250	250
Images	2000	250	250

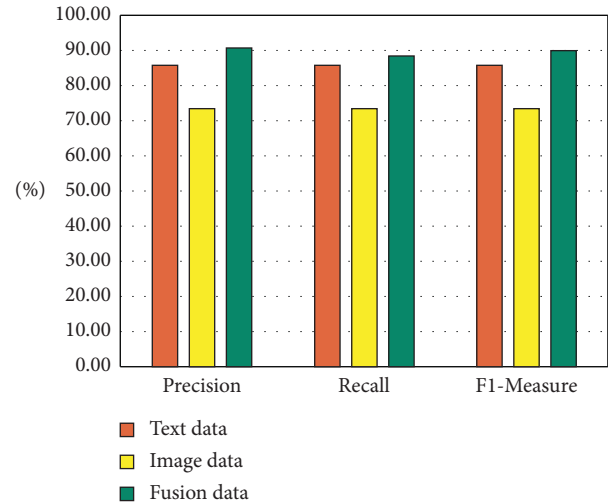


FIGURE 7: Assessment results of the model for different data forms.

benefits of various data types and to produce a more thorough assessment and analysis of students' affective cognition conditions.

In addition, to further verify the feasibility and usefulness of our research model in real life, the scores based on the questionnaire scales were compared with the assessment results of the model to produce the final mental health assessment accuracy. Figure 8 displays the detailed assessment results.

The results in Figure 8 reveal that among the assessment results of the three mental health levels, the recognition accuracy for the health level is the highest, reaching 91.6%. Followed by the existing depression level, its recognition accuracy was 85.3%. The lowest accuracy was obtained for the possible depression level, which was 80.6%. The difference between the assessment accuracy of the health level and the assessment accuracy of the possible depression level is nearly 10%, and the gap is relatively obvious. The reason is that, whether students in the healthy level or in the existing depression level, their affective cognition in these two states is in a relatively stable positive state or negative state most of the time. As a result, the model could more precisely pinpoint the psychological traits of students in the above two states and determine if they are more likely to experience depression. In contrast, the possible depression level is in the middle of the above two conditions, and the psychological and affective cognition characteristics of students in this condition are vague and not easy to identify, thus leading the model to be prone to misjudgment in the assessment, which affects the accuracy of the assessment. Overall, the average assessment accuracy of the algorithm proposed in this paper reached 85.8%, indicating that the model has strong

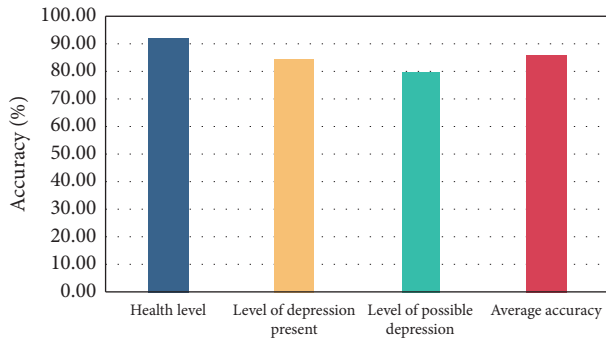


FIGURE 8: Assessment results of different health levels.

discriminative power in determining whether students have depressive tendencies and can accurately detect students in abnormal psychological states, which has some practical application value.

Compared with the traditional assessment methods based on a single data source, this study uses the multisource student behavior data recorded in the smart campus platform as the basis and uses a deep learning algorithm to extract the deep semantic knowledge and psychological and affective cognition characteristics of students contained in these data, which is more advantageous than traditional machine learning algorithms in terms of assessment accuracy. In real life, this research also helps to achieve rapid positioning and continuous tracking of students' affective cognition status. In this way, problems can be discovered in time, and active push warnings can be provided for school teachers and administrators to realize the development of students' psychological problems from "passive management" to "active prevention".

6. Conclusions

This work seeks to combine deep learning technology to investigate a novel affective cognition analysis technique for college students associated with the smart campus in order to realize intelligent assessment and early warning of college students' mental health concerns. Most of the existing related research studies rely on the questionnaire survey mode, which has shortcomings such as single data source, easy to hide problems, and low efficiency. These are not conducive to providing precise and real-time intervention treatment for students with the abnormal psychological state. Therefore, for the purpose of improving the comprehensiveness and accuracy of the evaluation, this research makes improvements in two aspects of data source and model training. First, relying on the intelligent campus platform of universities, this study uses multisource behavioral data to portray the affective cognition portrait of students, including access control data, learning behavior data, and network data, so as to provide more comprehensive data sources for the affective cognition assessment of college students. Second, the collected multisource data is divided into two categories: texts and images, and CNN technology is utilized for model training, which greatly improves the accuracy and efficiency of the model. The simulation results show that the

novel method has better accuracy in assessing the affective cognition status of college students compared with the data of a single modality. It can accurately grasp the mental health level of students and also provides a reference for the precise and intelligent development of mental health education in colleges and universities. In the study, we use the multi-source data recorded by the smart campus platform to identify the affective cognition status of students and achieve relatively satisfactory assessment results, but there are still some problems to be further solved. In this research, both image and text data from multiple-source dataset were trained using CNN models, without considering the differences between the two. In the future, it can be studied to use suitable deep neural network models for training for different modal data. For example, for images, a CNN model can be used, while for text data, an LSTM model, which is better at text analysis, can be adopted.

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 study was sponsored by 2022 University Philosophy and Social Science Research Project (Fund No. 2022SJYB1050).

References

- [1] R. J. Mcquaid, "Transdiagnostic biomarker approaches to mental health disorders: consideration of symptom complexity, comorbidity and context," *Brain, Behavior, & Immunity-Health*, vol. 16, Article ID 100303, 2021.
- [2] C. R. Wilks, R. P. Auerbach, J. Alonso et al., "The importance of physical and mental health in explaining health-related academic role impairment among college students," *Journal of Psychiatric Research*, vol. 123, pp. 54–61, 2020.
- [3] Q. Tang, Y. Zhao, Y. Wei, and L. Jiang, "Research on the mental health of college students based on fuzzy clustering algorithm," *Security and Communication Networks*, vol. 2021, no. 3, pp. 1–8, 2021.
- [4] Z. Behi, K. T. W. Ng, A. Richter, N. Karimi, A. Ghosh, and L. Zhang, "Exploring the untapped potential of solar photovoltaic energy at a smart campus: shadow and cloud analyses," *Energy & Environment*, vol. 33, no. 3, pp. 511–526, 2022.
- [5] A. M. Eltamaly, M. A. Alotaibi, A. I. Alolah, and A. A. Mohamed, "IoT-based hybrid renewable energy system for smart campus," *Sustainability*, vol. 13, no. 15, pp. 1–18, 2021.
- [6] M. Kuwabara, K. Oba, N. Takano et al., "An exploratory questionnaire survey about overwork on mental health of Japanese elementary and junior high school teachers," *The Journal of Mental Health Training, Education and Practice*, vol. 16, no. 3, pp. 181–186, 2021.

- [7] E. Sva, C. Jvoab, and D. Pda, "Ecological Momentary Assessment and Other Digital Technologies for Capturing Daily Life in Mental health," *Mental Health in a Digital World*, Academic Press, Cambridge, MA, USA, 2022.
- [8] P. Wang, X. Chi, and Y. Yu, "The application of deep learning in college students' sports cognition and health concept," *Journal of Intelligent and Fuzzy Systems*, vol. 24, pp. 1–13, 2021.
- [9] K. Dheeraj and T. Ramakrishnu, "Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model," *Expert Systems with Applications*, vol. 182, Article ID 115265, 2021.
- [10] C. Zang-Hee, P. Sun-Ha, K. Young-Bo, and C. Taigyoun, "Neural adaption of deep learning for human cognition and thinking hypothesis – allied with Engram, Langram, and Neural Lexicon hypothesis," *IBRO Reports*, vol. 6, p. S73, 2019.
- [11] A. Chanaa and N. Faddouli, "An Analysis of Learners' Affective and Cognitive Traits in Context-Aware Recommender Systems (CARS) Using Feature Interactions and Factorization Machines (FMs)," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, 2021.
- [12] P. Zhang, J. Lin, J. He, X. Rong, C. Li, and Z. Zeng, "Agricultural machinery virtual assembly system using dynamic gesture recognitive interaction based on a CNN and LSTM network," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5256940, 16 pages, 2021.
- [13] A. S. Uban, B. Chulvi, and P. Rosso, "An emotion and cognitive based analysis of mental health disorders from social media data," *Future Generation Computer Systems*, vol. 124, pp. 480–494, 2021.
- [14] S. Liu, F. Vahedian, D. Hachen et al., "Heterogeneous network approach to predict individuals' mental health," *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 2, pp. 1–26, 2021.
- [15] H. Fuketa and K. Uchiyama, "Edge artificial intelligence chips for the cyberphysical systems era," *Computer*, vol. 54, no. 1, pp. 84–88, 2021.
- [16] A. Sadeghzadeh and H. Ebrahimnezhad, "Pose-invariant face recognition based on matching the occlusion free regions aligned by 3D generic model," *IET Computer Vision*, vol. 14, no. 5, pp. 268–277, 2020.
- [17] A. Sasa, A. Vm, C. Nmb, M. Milan, Š. Marko, and S. Muzafer, "An efficient novel approach for iris recognition based on stylometric features and machine learning techniques," *Future Generation Computer Systems*, vol. 107, pp. 144–157, 2020.
- [18] K. Suthar and Q. P. He, "Multiclass moisture classification in woodchips using IIoT Wi-Fi and machine learning techniques," *Computers & Chemical Engineering*, vol. 154, Article ID 107445, 2021.
- [19] M. Bazulin, D. Sabitov, and M. Charara, "Determination of the elastic parameters of a VTI medium from sonic logging data using deep learning," *Computers & Geosciences*, vol. 152, no. 12, Article ID 104759, 2021.
- [20] A. Jr, A. Sp, K. K. Bo, O. Sang-Hoon, and L. Soo-Young, "Unsupervised multi-sense language models for natural language processing tasks," *Neural Networks*, vol. 142, pp. 397–409, 2021.
- [21] J. Juszczuk, P. Badura, J. Czajkowska et al., "Automated size-specific dose estimates using deep learning image processing," *Medical Image Analysis*, vol. 68, Article ID 101898, 2021.
- [22] I. Kaur, M. N. Doja, and T. Ahmad, "Data mining and machine learning in cancer survival research: an overview and future recommendations," *Journal of Biomedical Informatics*, vol. 128, Article ID 104026, 2022.
- [23] V. Mann and V. Venkatasubramanian, "Retrosynthesis prediction using grammar-based neural machine translation: an information-theoretic approach," *Computers & Chemical Engineering*, vol. 155, Article ID 107533, 2021.
- [24] M. J. Shafiee, A. Jeddi, A. Nazemi, P. Fieguth, and A. Wong, "Deep neural network perception models and robust autonomous driving systems: practical solutions for mitigation and improvement," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 22–30, 2021.
- [25] K. Jia, "Chinese sentiment classification based on Word2vec and vector arithmetic in human–robot conversation," *Computers & Electrical Engineering*, vol. 95, Article ID 107423, 2021.

Research Article

Application of Cognitive System Model and Gestalt Psychology in Residential Healthy Environment Design

Jicheng Yang ¹ and Chao Yuan ^{2,3}

¹School of Design, Beijing Normal University, Zhuhai, Guangdong 519087, China

²College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China

³School of Design, Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Chao Yuan; circle@jiangnan.edu.cn

Received 21 July 2022; Revised 2 August 2022; Accepted 3 August 2022; Published 21 August 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Jicheng Yang 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.

The influence of the environment on people is very large. According to the general rules of people's daily routines, people spend at least half of their time in residential areas. Therefore, the environment of residential areas has a great impact on people's emotions, behaviors, and even the development of living habits. Residential area refers to a residential area where residents live in clusters and form a certain scale, including buildings where people live, public buildings and facilities for rest, education, fitness, work, and even communication between people, green space, and traffic roads. Usually, the environmental design of urban residential areas usually meets the requirements of diverse functions, strong compatibility, and convenient travel, so as to facilitate residents' living. In an ideal state, the environmental design of a residential area should not only meet the basic living conditions but also improve the living comfort of the residents. Because there is a close relationship between people's psychology and behavior and the living environment of the community, environmental design that is in line with people's positive and optimistic psychology helps residents maintain a happy mood. Therefore, from the perspective of environment to residents' living comfort, this paper introduces a cognitive system model and Gestalt psychology to optimize the design of residential healthy environment. A more harmonious and comfortable living environment is established by using the principle of bottom, grouping, and simplification in Gestalt.

1. Introduction

The living environment refers to the accommodation within the scope of the residential area, the surrounding things related to the living behavior, and the corresponding psychological feelings generated by people. The residential area environment includes both the tangible environment, that is, natural elements, artificial elements, and social elements. Its specific performance is the space formed by the combination of various entities, including residential buildings, public service facilities, structures, roads, squares, green spaces, and various activity places, which together constitute the physical environment of the residential area. The living environment also includes intangible and spiritual things. For example, residents' interest in life, information exchange and communication, social order, neighbor relationship, spiritual outlook, moral cultivation, customs, security, and

sense of belonging. The improvement of people's living standards makes modern residents have higher requirements for the living environment. From the beginning, people only need a comfortable house, and now they are pursuing a better and better external environment. Based on this, the environmental design of the residential area began to enter people's sight and attract people's attention. From a macro perspective, the environmental design of the residential area is a part of the urban and even regional environmental design. Therefore, it is closely related to the development of cities, the changes in urban life, and the protection and optimization of natural ecology. From a microscopic point of view, the environmental design of residential areas serves people, and is closely related to people's behaviors and psychological needs. The environmental landscape of the residential area is not only for people to see but also for people to use.

The domestic environmental design of residential areas has risen from the traditional material level design to the spiritual level. The high-quality residential healthy environment design can make the occupants feel comfortable and satisfied psychologically. The relationship between the living environment and people's psychology has become a problem that scholars are very concerned about. Therefore, the discipline of environmental psychology was also born one after another [1, 2]. Reference [3] introduces environmental psychology into the design of buildings. This study shows that the control of spatial communication in architectural design leads to the control of behavioral communication in interpersonal communication is very important. Reference [4] introduces environmental psychology and social psychology into all aspects of human life and discusses their impact on people's well-being. Reference [5] explores the relationship between man and architecture, nature, and living environment. The study pointed out that the purpose of environmental psychology is to explore the relationship between man and nature. Reference [6] pointed out that environmental psychology should give priority to the reciprocal relationship between people and the environment. Reference [7] introduces psychology into social environment and life. Reference [8] presents a comprehensive framework for promoting healthy habits based on environmental psychology. Reference [9] discusses the relationship between feng shui and psychology, and studies the influence of feng shui on people's lives. Reference [10] assesses how the built environment is influenced by human life as well as the urban landscape. The above research studies the relationship between the environment and psychology from the perspective of psychology, and proves that people's psychological activities are affected by the environment.

It is based on the above research conclusions, in order to improve the environmental comfort of the residential area, this paper introduces the psychological theory into the environmental design of the residential area. Typical psychological theories mainly include cognitive system model and Gestalt psychology [11–13], color psychology [14–16], cognitive psychology [17–19], ecological self-consciousness theory [20], and probability self-consciousness theory [21]. Gestalt psychology is most widely and successfully used in interaction design. Therefore, this paper applies the cognitive system model and Gestalt psychology to the environmental design of residential areas. Gestalt principles of organization theoretically clarify the relationship between perceptual integrity and form. In the residential healthy environment design, the use of Gestalt's organizational principles can increase the richness and observability of the landscape, and it is convenient for residents to increase their awareness and understanding of the landscape from the perspective of formal beauty.

2. Related Concepts

2.1. Introduction to Residential Area and Environment. The characteristic of the residential area is that the residences are concentrated, and there are a certain number and corresponding scale of public service facilities. It can provide

residents with housing, recreation, and daily life services. Residential areas have a specific scale in urban planning. Usually the population is about 40,000 people, and the land area is about 80 hectares. The components of the residential area are shown in Figure 1:

The environment of the residential area refers to the living place of the residents, the surrounding things related to the behavior and the corresponding psychological cognition generated by the people within the scope of the residential area. The elements of residential area mainly include geographical elements, economic elements, social elements, and psychological elements. Geographical elements mainly refer to the physical environment such as buildings, public facilities, green plants, roads. Economic factors refer to shopping and consumption in the community. The social element refers to the social interaction between residents such as the interaction between neighbors. Psychological elements refer to the overall division of the residential area presented by the environment of the residential area such as the mood of the residents, the relationship between neighbors, and the spiritual outlook.

2.2. The Composition of the Environmental Landscape of the Residential Area. The living environment can be divided into different types according to different standards. One divides the living environment into material space environment and nonmaterial space environment, and the other divides the living environment into outdoor environment and indoor environment.

The physical space environment refers to the basic needs environment on which people live, including buildings and engineering equipment such as housing, squares, green space, commercial, service, and cultural facilities. A good material space environment has good air quality, sufficient sunlight, water environment, and green space for plants, as well as perfect and convenient facilities. These can meet people's living needs and behavioral needs. The material space environment design not only includes the buildings and greening of the residential area but also includes green space decoration, walls, gates, activity facilities, various signs, water features, sculptures, lighting facilities, audio facilities, etc. These contents must be organically integrated with the residential building to form a harmonious whole.

The immaterial space environment is reflected in spiritual, informational, and psychological belonging. It mainly includes mental space, information space, and psychological space. Spiritual space can bring people beautiful sentiments and creative thinking ability. Information space can shorten the distance between people and enable information sharing. Mental space is built on people's concept of home. Home is a shelter for people both physically and psychologically. Nowadays, many communities build a sense of community identity and build themed homes from this point. The components of the outdoor environment and the indoor environment are shown in Figure 2:

Through the classification of the environmental landscape of the residential area, it can be concluded that the constituent elements of the residential area landscape can be

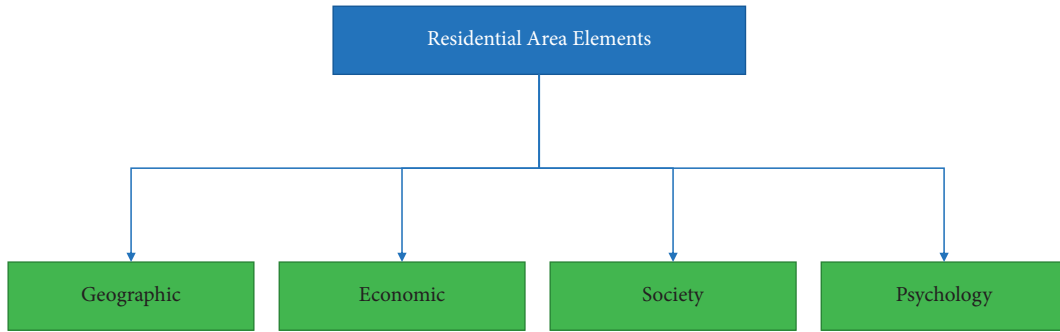


FIGURE 1: Composition of residential area.

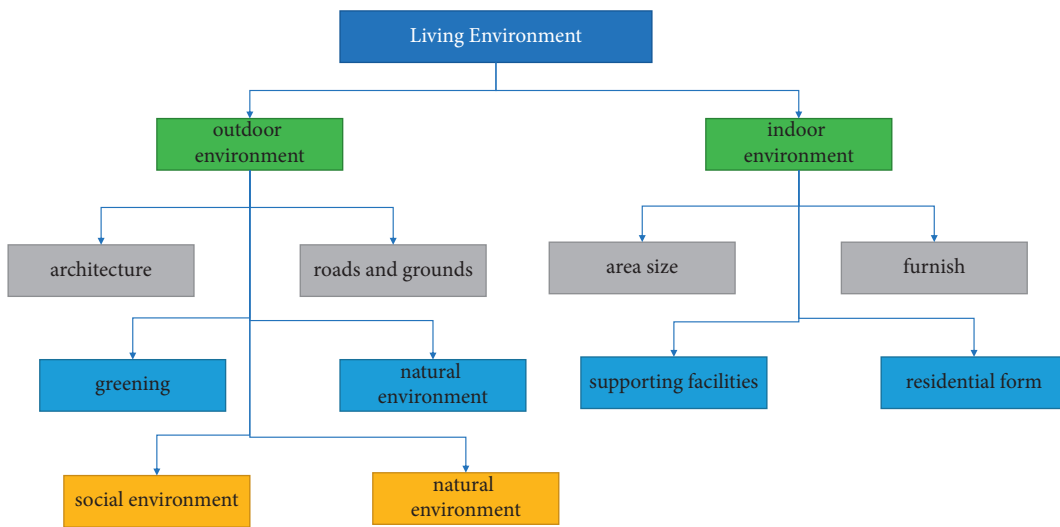


FIGURE 2: Composition of indoor and outdoor environments.

divided into natural elements, artificial elements, and social elements. Natural elements mainly include native landscapes such as hills, woods, and lakes. Artificial elements mainly include rockeries, flower beds, fountains, statues, etc., in the residential area. The social elements mainly include activities held in residential areas and places for leisure and entertainment.

2.3. Cognitive System Model and Gestalt Psychology Theory. In 1912, a German scholar put forward the theory of Gestalt psychology. Gestalt psychology mainly contains two meanings: one refers to the general attribute of the transaction, namely the form; the other refers to the individual attribute of the transaction, that is, the separated whole, and the form is only one of its attributes. This theory mainly emphasizes the relationship between the whole and the parts of things. A whole may be composed of multiple parts, and the parts are combined to form a whole. And the reason why each of these parts play their own role is because they do not exist alone, but in the whole. Gestalt psychology does not look at things from separate parts, but studies problems as a whole. In daily life, when people perceive an object, they like

to organize and order the object, so that it is easier to enhance the understanding and adaptation of the object. In Gestalt psychology, there are three main organizational principles, namely, the principle of map-ground relationship, the principle of grouping, and the principle of simplification. The map-ground relationship means that when people observe an object, they cannot fully perceive all the information of the image, but can only perceive part of the information. The perceived information forms the image, and the unperceived information is called the background. The schematic diagram of the map principle is shown in Figure 3.

When we observe things, Gestalt psychology believes that human perception can link similar elements together so that they form an organic whole. This law of perceiving similar elements as a unified whole is called the grouping principle. The grouping principle is shown in Figure 4.

Gestalt psychology believes that when people perceive graphics, in order to be more conducive to their own understanding and cognition, they usually prefer to simplify operations on graphics, which is called the principle of simplification. The simplification principle is shown in Figure 5.



FIGURE 3: Principles of map and map.

3. The Relationship between Residents and the Environment in Which They Live

3.1. The Relationship between Residents and the Environment of the place of Residence. The living environment can be divided into two parts: hard environment and soft environment. The hard environment refers to the hardware facilities of the community, such as fitness facilities, activity rooms. Soft environment refers to the cultural literacy, civility and politeness of community living. The two must be interdependent and constrained. A good living environment shows that the soft environment and the hard environment complement each other and promote each other. A bad living environment either has only one aspect, or even lacks both aspects. At present, the hard environment of each residence has reached the basic needs of residents, but the soft environment needs to be further improved. There is a close relationship between the living behavior of residents and the living environment. The construction of the soft environment in the residential area mainly depends on the behavior of the residents. People-oriented is a common requirement for the upgrading of contemporary living environment. Human life needs are divided into material needs and spiritual needs. Living needs are embodied in various activities of residents, which are divided into routine activities, occasional activities, and social activities. Each activity has different requirements for the environment.

Routine activities refer to the activities that residents need to complete every day, such as commuting to get off work, commuting to school, grocery shopping, cooking, washing dishes, mopping the floor. These are activities that must be done on a daily basis. These activities are seriously affected by the environment of the residential area, especially the effect of the material environment. For example, whether the public facilities in the residential area are complete, and

whether the transportation is convenient. Whether it is convenient to pick up and drop off children on the way to and from get off work, whether it is convenient to shop and park vehicles, etc. Once these physical environments cannot meet the conditions for engaging in essential activities, it means that the residential area does not have convenience and safety. Accidental activities are activities that only occur under the right circumstances. In contrast to routine activities, it occurs only when people have the will and when and where it is permitted. Such activities include mountain climbing, running, and watching movies. The frequent occurrence of contingent activities is an important criterion to test whether the environmental planning is in place. For example, the greening of the community is well done, and the active atmosphere of the running track and promenade design can stimulate people's willingness to exercise. Social activities are behaviors that require the participation of other people. Such as marathon, community encyclopedia knowledge contest.

To sum up, a good residential healthy environment can not only satisfy people's regular activities, but also promote residents' contingency and social activities. So as to promote interpersonal communication among residents, fitness exercise. The development of outdoor activities among residents also enhances the soft environment of the residential area, reflecting that the residential area environment is full of vitality and the neighborhood relationship is good. By analyzing the relationship between the residents' behavior and the environment of the residential area, it is concluded that a good residential healthy environment can promote the residents to have a sense of intimacy and dependence on the living environment. And this generalization of intimacy and dependence leads to a sense of belonging to the environment. For residents, there are many factors that affect the sense of belonging, such as the sense of security, comfort, and privacy mentioned in the ecological perception theory. In the design of the residential landscape environment, we must pay great attention to these psychological needs of people.

3.2. Residents' Psychological Needs and Residential Healthy Environment Design. For the environment of the residential area, due to the different ages, personalities, and hobbies of the residents, the requirements for environmental design are also different. According to age, residents can be roughly divided into the elderly, young and middle-aged, and children.

The elderly pay more attention to the comfort and leisure of the body, and require more interest in life, so as to pass the time. The elderly pay more attention to the accessibility of communication, and at the same time have higher requirements for safety and convenience. Therefore, in the construction of the garden landscape design in the residential area, a special activity place should be opened up for the elderly. This is not only convenient for the elderly to exercise and exercise but also a platform for the elderly to communicate with each other. Thereby it can ensure the fun

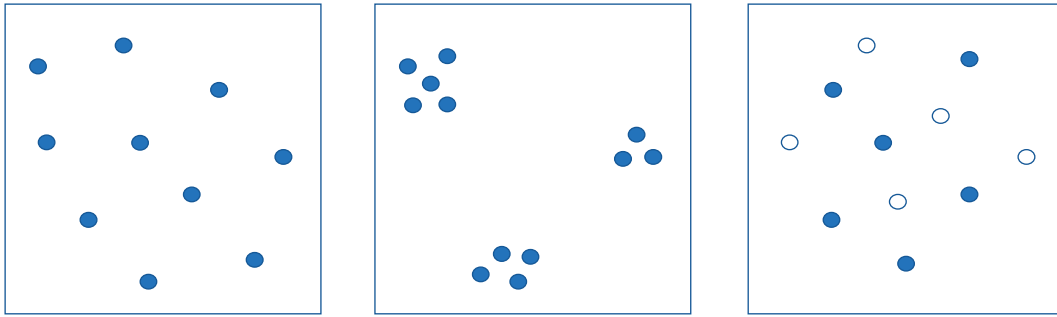


FIGURE 4: Grouping principle.

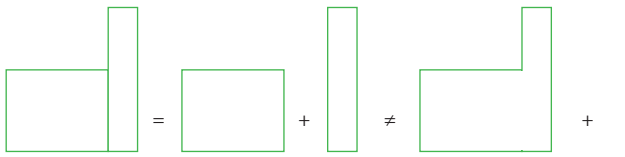


FIGURE 5: Simplification principle.

of communication for the elderly and satisfy the emotional lack of psychological aspects.

For young and middle-aged groups, they pay more attention to the quality of the environment and the design of ecological green plants, because a comfortable and green environment can make people more physically and mentally relaxed. Based on this demand, first of all, we should increase the coverage of green plants as much as possible, and set up some leisure places such as coffee houses, reading bars, which can enable young people to communicate better. Secondly, young and middle-aged groups like to exercise, and some running tracks should be set up in the residential area as much as possible to facilitate walking and running, and there should be no shortage of sports venues, such as table tennis tables or basketball courts. In this way, young and middle-aged groups can smoothly carry out sports and exercise in the residential area. It should be noted here that sports venues should maintain a certain distance from residential buildings so as not to affect the rest of residents. Because different groups of people have different time periods for rest and exercise, some people like to get up early to exercise, and some people like to exercise in the evening. Not disturbing the daily life of other residents is the most basic requirement.

For groups of children, they mainly focus on entertainment and fun in the residential area. For example, for the design of garden landscapes, children like the bright colors of the vegetation, the contrast is large, the shapes are varied, interesting, and lovely. In the rest area of the community, on both sides of the promenade, and in the children's playground, some cartoon and animation elements can be set up to increase the fun of the place. Second, child-optimized places are a top priority for groups of children. First of all, openness should be considered, because usually young children are accompanied by adults to play. The place is too small to fit, the place is too closed and uncomfortable. Most children like to slide and play with sand. When purchasing

amusement equipment, there should be a priority. Children also like to be close to nature. For garden trees, you can plant some tree species with bright flowers or leaves and seasonal changes to enhance the ornamental effect of the trees and increase the interest of the garden. These improvement measures can give full play to the attractiveness of residential gardens to children.

In summary, we can conclude the following rules. First, no matter what type of people, there is a demand for green plants. The general requirements for green plant design are bright colors and interesting shapes. The second is to have suitable sports venues. In addition to considering residents' preferences for sports, outdoor sports should also be taken into account. Outdoor exercise can ensure enough sunlight, which is very important for children and the elderly. Proper exposure to sunlight can supplement calcium and is good for physical and mental health. The third is to have leisure and entertainment venues. The purpose of this place is to facilitate communication between residents. Therefore, a casual, comfortable, slow-paced atmosphere should be created.

4. Residential Healthy Environment Design Based on Cognitive System Model and Gestalt Psychology

4.1. Interior Design. The overall theory is that the components of the indoor space have different styles, but they tend to be consistent as a whole. Cognitive system model and Gestalt psychology believes that perception is not the addition of multiple individual elements, nor is it a specific element, but the perception of the whole (See Figure 6). That is to say, the whole is greater than the sum of its parts. The interior space design can use different materials and different styles of decorations to match together, and finally let people perceive an overall decoration style. Gestalt psychology believes that the addition of parts does not equal the whole, and the whole is greater than the sum of the parts. The combination of psychological field and physical field enables us to perceive more connotations than things themselves, which is also the "design style" that interior designers have been trying to create. For example, Chinese style interior design is mainly reflected in traditional furniture, decorations, and black and red-based decorative colors. Symmetrical layouts are often used in the interior. The style is elegant, the shape is simple and beautiful, and the color is

strong and mature. Indoors generally have calligraphy and painting, plaques, hanging screens, bonsai, porcelain, antiques, screens, Bogu frames, and so on. The Chinese style is shown in Figure 7.

Gestalt psychologists believe that the observer's perception is a "field" concept. The mind-object field is the combination of the psychological field and the physical field. The identity, status, and age of the space users are considered in the design, and the physical field is designed according to different psychological needs. Due to the theoretical basis of isotype theory, the psychological field and the physical field will correspond one by one, and finally create an interior space that makes users feel warm, which is also the essence of interior design.

There are many "physical fields" in interior design, such as materials, lighting, and functions. Different "psychological fields" will be directly reflected in "physical fields". For example, the interior design and decoration of kindergartens are different from ordinary home decoration. The choice of color system should be based on the physical and psychological needs of children. The right color is in line with the growth process of children. To create a healthy learning and living environment for them, cool colors should be used carefully, and warm colors can bring a warm home-like feeling to children. For example, green will bring imaginative space to children, and the decoration of kindergarten should be combined with the region and the selected theme to determine the decoration style. A kindergarten uses orange, pink, yellow, green, purple, blue, and other different colors as the theme colors of each classroom, including wallpaper, table, and chair colors. This is to take full advantage of the influence of different colors on mental activity. Therefore, if the physical field is designed according to specific psychological needs, it can fill our indoor space with "temperature" and make people feel more comfortable.

4.2. Outdoor Environment Design. Outdoor environment design mainly includes ecological landscape design, activity place design, rest place design, and public space design. Ecological landscape design is mainly divided into green design, water environment design such as fountains and pools. For the greening design, it is mainly designed from the arrangement, shape, height, and size of the plants. Based on the theory of Gestalt psychology, it is known that the design of green plants needs to avoid a single color tone, and the combination of plants with contrasting colors will be more eye-catching. An example of the design of green plants is shown in Figure 8.

One visual perception forms a figure, and it is highlighted on another background figure, this phenomenon is one of the basic principles of Gestalt graphics principle. The relationship between "shape" and "bottom" is dialectical, they are interrelated and interdependent. In the visual realm nothing is negative, positive can be negative and negative can be positive. Therefore, it is necessary to make overall plans and make full use of the changing relationship between "shape" and "bottom" in the design to obtain a perfect and interesting visual effect. In the community, residential

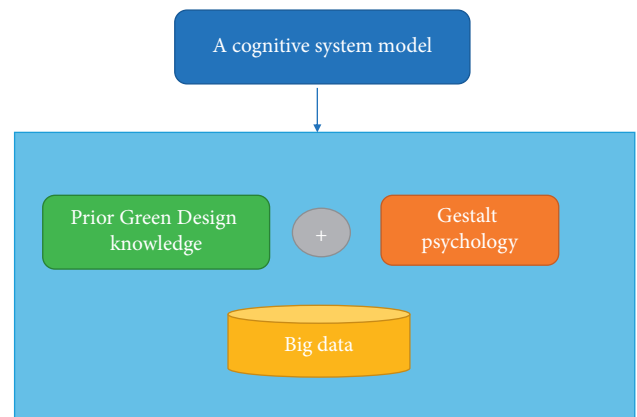


FIGURE 6: A cognitive system model for healthy environment design.



FIGURE 7: Chinese style interior design.



FIGURE 8: Green plant design.

buildings are regular, small in size, with relatively clear outlines, and are in a dominant position. It has a great influence on the layout and landscape design of the whole community. Therefore, designers often regard residential buildings as "pictures" and green landscapes as "bottoms", in order to more clearly distinguish the relationship between the picture and the bottom in the community. Figure 9 depicts a schematic diagram of the community's artificial pool. The pool is placed not far from the entrance of the community. The reflection of the blue sky and white clouds on the water surface, as well as the surrounding buildings and green plants, makes the courtyard present a colorful scene and enhances the integrity of the community. This wholeness makes the courtyard itself take on the character of a "picture". The arrangement of the buildings forms the "bottom".



FIGURE 9: Water environment design.

4.3. Case Analysis

4.3.1. Residential Area Description. In a community in an economic development zone, 47% of the residents are young people, 26% are the elderly, and 27% are children. There are many young people in the residents of this community, and they are stingy and full of vigor and vitality on the whole. The community is far from the city and close to the suburbs. The air is fresh and the vegetation is abundant. The area outside the community is wide, but public transportation is not very developed. There are schools and parks within 3 km. The community is rectangular and covers an area of 55,997 square meters. An aerial view of the community is shown in Figure 10. The gate of the community is set in the middle of two square grids, and a road divides the community into two parts.

4.3.2. Environmental Design. Some attractive landscapes should be set up at the entrance of the community to make people feel happy as soon as they enter the community. In addition, more than one small square should be set up in the community to facilitate residents' communication and sunbathing. The design of the small square should be based on the principles of Gestalt psychology, with the surrounding green plants as the bottom and the square in the middle as the map. In the design of small roads in the community, there are various ground paving forms, including regular and rigorous paving forms, as well as irregular and interesting paving forms. In view of the youthful characteristics of this community, some irregular and clustered floor tiles can be laid. In the design of sitting and resting space, the number of seats in the community is relatively large. The basic seats are arranged near the pedestrian path, which is convenient for many residents with limited mobility or physical fatigue to sit at any time. Other auxiliary seats such as steps, low walls, flower terraces, etc., are combined with landscape design to provide services for residents who use it occasionally.

Due to the large number of children in the community, the design of the children's play area is very important. The playground of the community is placed on the side of the main pedestrian street. Through game activities, residents can be attracted to stop and watch, and the vitality and vitality of the street can be increased. Dense plants are



FIGURE 10: A bird's-eye view of the community.

planted between the site and the road, which not only prevents children from accidentally rushing into the main road but also forms a good boundary, increasing the layering and richness of the space. There are seats for parents of young children to rest around the playground. The floor coverings come in a variety of colors. In the community, the elderly activity facilities are mainly distributed in the form of dots. These facilities include chess and card tables and chairs, fitness equipment, etc. This distribution pattern is conducive to the spontaneous activities of the elderly. For example, a few neighbors and friends who live nearby can easily reach the activity area and play chess and card leisure activities or exercise in good weather. In the community, because there are many seats, it is convenient for the elderly to do it at any time. Through the combination of seats, pavilions, corridors, and other facilities, the elderly are given more sitting and leisure space, which basically meets the needs of the elderly. The various needs of the elderly to sit, lie down, stop, and stay.

5. Conclusion

With the development of social economy, residents' demands for the environment of residential areas are increasing day by day. For modern residents, they not only require the living environment of the community to be comfortable and clean but also require that the living environment can give people a warm and decompressed feeling from the soul. This demand has a qualitative change, which highlights the effect of environmental design on the psychology of residents. Inspired by this, in order to qualitatively improve the residential healthy environment, this paper introduces cognitive system model and Gestalt psychology, a typical theory of psychology, into the design of the residential healthy environment. This paper first expounds the definition of residential healthy environment and Gestalt psychological theory. Second, it analyzes the relationship between psychological changes and environmental design, including how the environment affects residents' psychology and how residents' behavior affects the environment. Finally, based on the Gestalt theory, the design suggestions for the residential area environment are given, respectively. Including interior design, ecological green plant design, water

environment design. In the design of each module, relevant theories in Gestalt psychology are integrated. The effectiveness of the proposed design method is also demonstrated by relevant examples. However, there are still some problems in this paper that need to be further optimized. For example, the environmental design of the residential area is not complete, including the lighting, leisure space, and sports venues of the residential area. In addition, some other psychological theories are also considered to be introduced into the residential healthy environment design, such as color psychology, cognitive psychology, and so on. This study will carry out in-depth research on these aspects in the follow-up work.

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 by the Beijing Normal University.

References

- [1] EK. Bryant and C. C. Sonn, "Learning through life narratives in environmental and community psychology," *Australian Community Psychologist*, vol. 31, no. 2, pp. 97–106, 2022.
- [2] JM. Wiener and F. Pazzaglia, "Ageing- and dementia-friendly design: theory and evidence from cognitive psychology, neuropsychology and environmental psychology can contribute to design guidelines that minimise spatial disorientation," *Cognitive Processing*, vol. 22, no. 4, pp. 715–730, 2021.
- [3] SB. Oskouei, S. Toofan, and S. Jamali, "Promoting theoretical foundations of privacy concept in contemporary housing from perspective of environmental psychology, A Case study of Milad Tabriz Residential tower," *Bagh-E Nazar*, vol. 16, no. 79, pp. 69–82, 2020.
- [4] M. Ciszek, "The philosophy of perceiving the human environment from the perspective of environmental social psychology and environmental sociology (implications for sustainable environmental and health security)," *Problemy Ekorozwoju*, vol. 15, no. 2, pp. 211–222, 2020.
- [5] R. Gifford, "Environmental psychology: manifold visions, unity of purpose," *Journal of Environmental Psychology*, vol. 29, no. 3, pp. 387–389, 2009.
- [6] D. Uzzell and N. Rathzel, "Transforming environmental psychology," *Journal of Environmental Psychology*, vol. 29, no. 3, pp. 340–350, 2009.
- [7] S. Riger, "Transforming community psychology," *American Journal of Community Psychology*, vol. 29, no. 1, pp. 69–81, 2001.
- [8] O. Santos, A. Virgolino, A. Vaz Carneiro, and M. G. de Matos, "Health behavior and planetary health A multi-level environmental health approach," *European Psychologist*, vol. 26, no. 3, pp. 212–218, 2021.
- [9] S. Kryzanowski, "A comparative analysis of selected recommendations of the feng shui school of form, Alexander et al.'s pattern language, and findings of environmental psychology," *Urbani Izziv*, vol. 2, no. 30, pp. 124–134, 2019.
- [10] HH. Molana and R. E. Adams, "Evaluating sense of community in the residential environment from the perspectives of symbolic interactionism and architectural design," *Journal of Community Psychology*, vol. 47, no. 7, pp. 1591–1602, 2019.
- [11] B. Skowron and K. Wojtowicz, "Throwing spatial light: on topological explanations in Gestalt psychology," *Phenomenology and the Cognitive Sciences*, vol. 20, no. 3, pp. 537–558, 2020.
- [12] JD. Greenwood, "On two foundational principles of the berlin school of Gestalt psychology," *Review of General Psychology*, vol. 24, no. 3, pp. 284–294, 2020.
- [13] M. Sinico, "Scientific phenomenology in design pedagogy: the legacy of walter gropius and Gestalt psychology," *International Journal of Art and Design Education*, vol. 40, no. 1, pp. 99–107, 2021.
- [14] S. Akbay and N. A. G. Z. Borekci, "Construing colours using repertory grid technique: an idiographic approach in colour perception," *Color Research & Application*, vol. 47, no. 2, pp. 329–351, 2021.
- [15] EA. Hunter, MA. Hanks, A. Holman et al., "The hurdles are high: women of color leaders in counseling psychology," *Journal of Counseling Psychology*, vol. 68, no. 6, pp. 642–656, 2021.
- [16] H. J. Kim, VA. McNeil-Young, DN. Wang, RD. Duffy, and B. D. Underill, "Women of color and decent work: an examination of psychology of working theory," *The Career Development Quarterly*, vol. 70, no. 2, pp. 125–137, 2022.
- [17] N. Scott, "Cognitive psychology and tourism - surfing the "cognitive wave": a perspective article," *Tourism Review*, vol. 75, no. 1, pp. 49–51, 2020.
- [18] M. Drazil, "Lidová psychologie v současné neurovědě: diskuse o kognitivních ontologiích," *Filozofia*, vol. 77, no. 6, pp. 442–455, 2022.
- [19] EM. DeRobertis, "The humanistic revolution in psychology: its inaugural vision," *Journal of Humanistic Psychology*, vol. 61, no. 1, pp. 8–32, 2020.
- [20] DA. Friedman and E. Sovik, "The ant colony as a test for scientific theories of consciousness," *Synthese*, vol. 198, no. 2, pp. 1457–1480, 2021.
- [21] E. Vityaev, "Consciousness as a logically consistent and prognostic model of reality," *Cognitive Systems Research*, vol. 59, pp. 231–246, 2020.

Research Article

Optimization of Railway Mixed Goods Loading Layout considering Stability

Juan Wang,¹ Yinghua Yao,² Yinggui Zhang ,² and Ximing Wang³

¹School of Logistics and Transportation, Central South University of Forestry and Technology, Changsha 410004, Hunan, China

²School of Traffic and Transportation Engineering, Central South University, Changsha 410075, Hunan, China

³School of Intelligent Equipment Technology, Hunan Vocational College of Science and Technology, Changsha 410004, Hunan, China

Correspondence should be addressed to Yinggui Zhang; ygzhang@csu.edu.cn

Received 23 May 2022; Revised 6 June 2022; Accepted 17 June 2022; Published 21 August 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Juan 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.

It is well known that stability, center-of-gravity balance, and concentrated-weight are key factors of the transportation safety. The reasonable formulation of the loading layout scheme ensures the safety of shipment based on fully utilizing the effective volume and load capacity of freight vehicles. This paper takes the railway mixed goods loading layout as the research object, considering the constraints such as goods loading center-of-gravity balance, the allowable moment of concentrated-weight, supporting and goods placement mode, and taking the maximum comprehensive utilization rates for both effective volume and load capacity of freight vehicle as the optimization objective, an optimization model of railway mixed goods balanced and anticoncentrated-weight loading layout considering stability is built. Additionally, this paper designs mixed goods classified and simple/general goods block composition methods. We improve the representation and selection of layout space, construct goods block selection algorithm based on the greedy d-step lookahead tree search and goods block evaluation function and propose a goods block placement strategy and update rules of layout space after goods block placement. An optimization algorithm of railway mixed goods balanced and anticoncentrated-weight load layout considering stability is designed. The results show that the formulated scheme not only ensures that the goods meet the full support constraints, but also the comprehensive utilization rate of the effective volume and load capacity of the vehicle is not less than 89%, and the probability of meeting the loading center-of-gravity balance and allowable moment of concentrated-weight are as high as 99% and 99.47%, respectively. The proposed method realizes the balanced and anticoncentrated-weight loading of railway mixed goods, ensures the safe, stable, and efficient goods loading, and provides decision support for the safe loading layout of railway goods.

1. Introduction

The transportation safety of railway mixed goods of different types, sizes, and weights is closely related to the scientific and reasonable loading layout scheme. How to efficiently prepare the loading layout scheme of railway mixed goods, make full use of the effective volume and load capacity of freight cars, while ensuring the balance of goods loading center of gravity, anticoncentrated-weight, and stability support, has important practical significance for reducing costs and increasing efficiency and safe transportation of railway goods [1].

Many scholars have carried out some research related to goods loading layout problems, mainly focusing on

optimization objectives, constraints, goods combination methods, spatial representation, loading layout algorithm design, etc. For optimization objectives and constraints, Huang et al. for the single container loading problem, with the objective of minimizing the length of occupied space in a container, used a 0–1 mixed integer linear programming model to describe the problem and proposed a load distribution heuristic algorithm [2]. Costa et al. divided the center-of-gravity balance constraint into three subconstraints: longitudinal, transverse, and vertical [3]. Ramos et al. proposed a variety of group biased random key genetic algorithms based on static mechanical equilibrium conditions [4], and considering the limitation of loading

efficiency, a physical loading order algorithm is proposed [5]. In terms of the way the goods are combined, Bortfeldt et al. combined goods into vertical layers and proposed a hybrid genetic algorithm for strongly heterogeneous goods [6]. Zhu Xiang et al. constructed goods as towers and proposed an optimization method for loading multiple pieces of concentrated-weight goods in one vehicle [7]; however, for high-density nonconcentrated-weight goods, when the distribution of goods is unreasonable, it will also produce concentrated-weight. In addition, aiming at the problem of balanced loading of multiple vehicles and multiple pieces of goods, they put forward the idea of symmetrical loading of multiple vehicles and algorithm optimization strategy [8]. Fanslau et al. extended the way of goods combination and proposed a general block generation method [9]. In terms of spatial representation, Jiang Yidong et al. proposed the method of using a trinomial tree data structure to represent the partitioning of rectangular goods layout space by putting a suitable goods block in the lower left corner of each small space of the partition [10]. Moura et al. proved that using the maximum coverage method to represent the layout space can obtain better solutions than the partitioning method [11]. In addition, there are many studies on load layout algorithms and strategies. Araya et al. determined the parameters of goods evaluation function by the control variable method to improve the efficiency of goods unit search [12]. Wang Zhe et al. constructed and designed an optimization model and algorithm for the similar goods [13]. Liu Xiaoqun et al. constructed and designed an optimization model and algorithm based on different benchmarks for multivariety goods [14]. Bischoff proposed a constructive heuristic embedding search algorithm and placement rules for the packing problem of goods with different bearing strength [15]. Lei Dingyou et al. proposed a mixed and balanced loading method of railway containers for light and heavy goods based on the central skeleton idea, but they did not include the load capacity utilization rate into the optimization objective [16]. Huang et al. proposed a novel technique that combines a differential evolution algorithm with a ternary search tree model to solve the three-dimensional container loading problem [17]. Zhang et al. proposed an optimization method for balanced loading layout of railway container mixed goods, but stability constraints such as full support are not involved [18].

Existing studies have provided strong support for the loading layout of railway mixed goods. However, most of them only take the volume of the loading space as the optimization objective, ignoring the optimization of the load capacity utilization. In addition, the constraints related to concentrated-weight and stability are less considered. Motivated by the above considerations, to reflect the utilization rate of freight vehicles more objectively and comprehensively, and ensure the loading efficiency and safety, this paper studies the balanced and anticoncentrated-weight loading layout of railway mixed goods considering stability. Specifically, this problem uses railway freight vehicles (such as gondola cars and boxcars) to load mixed goods of different types, sizes, and weights, with the optimization goal of maximizing the comprehensive utilization of effective

volume and load capacity of freight vehicles, and considering the practical constraints such as loading center-of-gravity balance, the allowable moment of concentrated-weight and full support. For this problem, based on the process of goods loading layout, we designed algorithms and rules such as mixed cargo classification, fully supported goods block unit generation, goods block unit selection and placement, and update of remaining available layout space. Furthermore, an optimization method of railway mixed goods balanced and anticoncentrated-weight load layout considering stability is proposed. Finally, the feasibility and effectiveness of the method are verified by combining the international standard example and the mixed goods example generated based on the standard example.

The rest of the paper is organized as follows. In Section 2, the problem is defined, the practical constraints involved in loading are analysed, and the mathematical model of the problem is established. Section 3 designs a heuristic algorithm for railway mixed goods balanced and anticoncentrated-weight loading layout optimization considering stability based on the six elements of goods loading layout. In Section 4, the test and comparative analysis of the international standard example and the improved mixed goods examples are carried out. Finally, conclusions are drawn in Section 5.

2. Problem Description and Mathematical Model

2.1. Problem Description. Given a set of rigid cuboid goods set $C = \{C_1, \dots, C_i, \dots, C_n\}$ with uniform density of class n , whose total volume and total weight are V_n and Q_n , respectively; denote by $C_i = \{c_{i1}, \dots, c_{ij}, \dots, c_{iu}\}$ the goods of class i containing u goods, the volume of class i goods is V_i , and the weight of class i goods is Q_i ; l_{ij} , w_{ij} , h_{ij} , v_{ij} , q_{ij} indicate the length, width, height, volume, and weight of individual goods, respectively. Assuming that the effective space inside the vehicle is rectangular, and the goods can be loaded and stacked, not special cargo such as dangerous goods, allowed to have surplus to be left for the next assembly. There is no loading and unloading operation on the way, and the goods all arrive at the station together [19]. Take the left front inner lower corner of the vehicle as the coordinate origin O , take the floor of the vehicle as the X - Y plane, and establish the space rectangular coordinate system in Figure 1.

2.2. Constraint Analysis. Set L , W , H , V for the internal length, width, height, and effective volume of the railway vehicle, respectively, Q , Q_v for the maximum allowable weight and self-weight of the vehicle, respectively, D for the center distance of bogie, H_v for the vehicle floor to rail height, and H_o for the empty vehicle center of gravity high. In order to ensure transport safety, in accordance with the current "Rules for railway cargo loading and reinforcement" [20], the longitudinal deviation of the combined center of gravity after loading relative to the intersection of the vertical and horizontal center line of the vehicle floor shall be

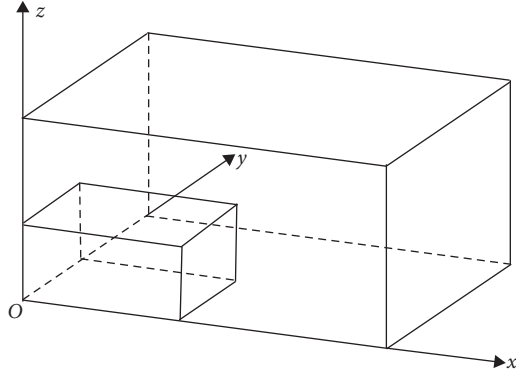


FIGURE 1: Space rectangular coordinate system inside the freight vehicle.

calculated according to $\Delta_1 = \min\{(Q - Q_C) \times D/2Q_C, 5 D/Q_C\}$, where $Q_C = \sum_{i=1}^n \sum_{j=1}^u \mu_{ij} q_{ij}$ for the total weight of the loaded goods, μ_{ij} for 0 or 1 that the goods c_{ij} is not loaded or has been loaded; the lateral deviation $\Delta_2 \leq 100mm$; the height of the center of gravity of the vehicle from the rail surface is $\Delta_3 \leq 2000mm$, more than that should be speed limit [20].

Using the X-Y plane rectangular coordinate system of Figure 1, setting $(G_1, G_2, 0)$ for the coordinates of the intersection G of the vehicle floor longitudinal and transverse center line and $(c_{ij}^x, c_{ij}^y, c_{ij}^z)$ for the coordinates of the center of gravity of the goods c_{ij} after it is put into the vehicle. Then, the coordinates (G_x, G_y, G_z) of the goods combined center of gravity in the vehicle are $(\sum_{i=1}^n \sum_{j=1}^u \mu_{ij} c_{ij}^x q_{ij} / Q_C, \sum_{i=1}^n \sum_{j=1}^u \mu_{ij} c_{ij}^y q_{ij} / Q_C, \sum_{i=1}^n \sum_{j=1}^u \mu_{ij} c_{ij}^z q_{ij} / Q_C)$. The actual longitudinal deviation of the combined center of gravity after loading is $S_1 = |G_x - G_1|$, the transverse deviation is $S_2 = |G_y - G_2|$, and the height of the center of gravity of the vehicle is $S_3 = G_z + H_v$. When $S_1 \leq \Delta_1$, $S_2 \leq \Delta_2$, $S_3 \leq \Delta_3$, the goods loading center-of-gravity balance constraint is satisfied.

Using the X-Z plane rectangular coordinate system of Figure 1, the balance or not of the forces in the length direction of the vehicle is the main reason whether the concentrated-weight is generated or not [7]. Here, only the force in the length direction of the vehicle is analysed, as shown in Figure 2, the front and rear bogies are supported by R_A and R_B and the coordinate positions are X_A and X_B , respectively. According to the cargo boundary, the vehicle floor is divided into N layers $L_1, L_2, L_3 \dots L_N$ that is perpendicular to the longitudinal center line, $G_1, G_2, G_3 \dots G_N$ are the corresponding goods combined gravity of each layer, respectively, and $X_1, X_2, X_3 \dots X_N$ are the distances from each combined gravity force to any position x_0 on the floor length. According to the moment balance equation, the point B as a balance point to find out the vehicle floor in x_0 position bending moment M_0 is $R_A \cdot X_A - \sum_{l=1}^N G_l \cdot X_l$. For different loading schemes, the maximum working bending moment M_{\max} of the vehicle floor and its coordinate position can be determined

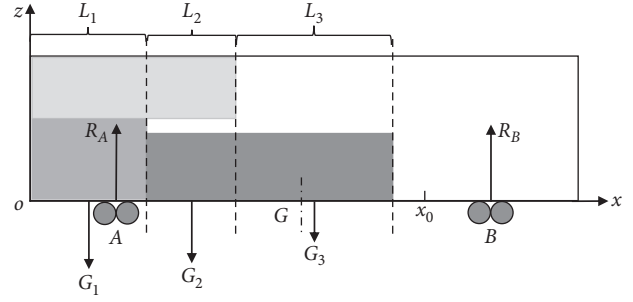


FIGURE 2: Diagram of longitudinal stress on vehicle floor.

according to the above method. When not more than the allowable bending moment $[M]$, that is, $M_{\max} \leq [M]$ is to meet the concentrated-weight allowable bending moment constraint.

Loading stability generally requires that the goods in the space can be placed stably without tilting and collapsing, mostly by restricting the supported area of the bottom surface of the goods in the vertical direction, and limiting the supported area of the bottom surface of the goods not less than the specified proportion of its own bottom surface areas, such as 55%, 75%, or 100%, that is, partial support constraint or complete support constraint. In this paper, the full support constraint is adopted to ensure that the bottom surface of the goods is fully supported by the vehicle floor or other loaded goods.

2.3. Model Construction. In summary, this paper considers the goods loading center-of-gravity balance, the allowable moment of concentrated-weight, full support, and other constraints, with the maximum comprehensive utilization rates for both effective volume and load capacity of freight vehicle as the optimization objectives. Finally, the optimization model of railway mixed goods balanced and anticentered-weight loading layout considering stability are built.

$$\max Z = \frac{\alpha \sum_{i=1}^n \sum_{j=1}^u \mu_{ij} l_{ij} w_{ij} h_{ij}}{V} + \frac{\beta \sum_{i=1}^n \sum_{j=1}^u \mu_{ij} q_{ij}}{Q}, \quad (1)$$

s.t.

$$S_1 \leq \Delta_1, \quad (2)$$

$$S_2 \leq \Delta_2, \quad (3)$$

$$S_3 \leq \Delta_3, \quad (4)$$

$$M_{\max} \leq [M], \quad (5)$$

$$l_{ij}^* \leq c_{ij}^x \leq L - (l_{ij} - l_{ij}^*), \quad (6)$$

$$w_{ij}^* \leq c_{ij}^y \leq W - (w_{ij} - w_{ij}^*), \quad (7)$$

$$h_{ij}^* \leq c_{ij}^z \leq H - (h_{ij} - h_{ij}^*), \quad (8)$$

$$|c_{ij}^x - c_{st}^x| \geq \frac{(l_{ij} + l_{st})}{2}, \quad (9)$$

$$|c_{ij}^y - c_{st}^y| \geq \frac{(w_{ij} + w_{st})}{2}, \quad (10)$$

$$|c_{ij}^z - c_{st}^z| \geq \frac{(h_{ij} + h_{st})}{2}, \quad (11)$$

$$\begin{aligned} &(((l_{ij}^x, w_{ij}^y, h_{ij}^z) = (l_{ij}, w_{ij}, h_{ij})) \vee, \\ &((l_{ij}^x, w_{ij}^y, h_{ij}^z) = (w_{ij}, l_{ij}, h_{ij})) \vee, \\ &((l_{ij}^x, w_{ij}^y, h_{ij}^z) = (h_{ij}, w_{ij}, l_{ij})) \vee, \\ &((l_{ij}^x, w_{ij}^y, h_{ij}^z) = (w_{ij}, h_{ij}, l_{ij})) \vee, \\ &((l_{ij}^x, w_{ij}^y, h_{ij}^z) = (l_{ij}, h_{ij}, w_{ij})) \vee, \\ &(((l_{ij}^x, w_{ij}^y, h_{ij}^z) = (h_{ij}, l_{ij}, w_{ij}))\%) = 1. \end{aligned} \quad (12)$$

$$S_{ij} = l_{ij}^x w_{ij}^y \cdot 100\%, \quad (13)$$

$$\sum_{i=1}^n \sum_{j=1}^u \mu_{ij} l_{ij} w_{ij} h_{ij} \leq V, \quad (14)$$

$$\sum_{i=1}^n \sum_{j=1}^u \mu_{ij} q_{ij} \leq Q, \quad i, s \in \{1, 2, 3, \dots, n\}; \quad j, t \in \{1, 2, 3, \dots, u\}, \quad (15)$$

where $(l_{ij}^*, w_{ij}^*, h_{ij}^*)$ is the position of the center-of-gravity of goods c_{ij} , that is, its geometric center, (x_{ij}, y_{ij}, z_{ij}) is the coordinates of the vertex of the left front lower corner of c_{ij} after it is put into the vehicle, and $l_{ij}^x, w_{ij}^y, h_{ij}^z$ is the dimensional length of the projection in the direction of x, y, z axis. Formula (1) is the objective function, which indicates maximizing the comprehensive utilization rates for both effective volume and load capacity of freight vehicle, such that $\alpha + \beta = 1, 0 \leq \alpha \leq 1, 0 \leq \beta \leq 1, \alpha = Q_{vw}'/2Q_{vw}$, where the volume-weight of the vehicle is $Q_{vw} = Q/V$, the volume-weight of the loaded goods is $Q_{vw}' = \sum q_{ij} / \sum \mu_{ij} l_{ij} w_{ij} h_{ij}$. Formulas (2)–(4) are the goods loading center-of-gravity balance constraints; formula (5) is the allowable moment of concentrated-weight constraint; formulas (6)–(8) indicate that the loaded goods do not exceed the vehicle boundary constraint; formulas (9)–(11) indicate that any two goods c_{ij}, c_{st} cannot overlap each other constraints; formula (12) for the placement of goods constraints; formula (13) for the full support constraints; formula (14) for the vehicle effective volume constraints; and formula (15) for the vehicle maximum allowable load constraints.

3. Algorithm Design

The studied problem is an NP-hard packing problem. Based on the six elements of goods loading layout [21], focusing on the internal space and goods of freight vehicle, the classification of mixed goods and the construction method of goods blocks, the selection and placement strategy of goods

blocks are designed, the representation, selection and update rules of layout space are given, and a heuristic algorithm for railway mixed goods balanced and anticoncentrated-weight loading layout optimization considering stability is formed.

3.1. Layout Space Representation. Using the maximum coverage method [11] to represent the layout space, as shown in Figure 3(a), after the goods are put into the space, three great rectangular layout spaces r_x, r_y, r_z are generated in the three directions of x, y, z , respectively. Considering the stability constraint, the bottom of all the generated layout spaces should be fully supported by the top surface of the placed goods or the floor of the vehicle. Therefore, it is kept unchanged in the x and y directions, while a rectangular space determined by the support area of the top of the loaded goods is generated in the z direction, as shown in Figure 3(b).

3.2. Goods Classification and Goods Blocks Construction. Constructing goods block is currently the most effective way to solve the packing problem [22]. The properties of mixed goods are complex. Taking the bubble weight ratio of goods 1: g_1 and 1: g_2 as the boundary, combined with the weight, volume, and density of goods, the mixed goods are divided into heavy goods, middle goods, and light goods. On this basis, different types of goods blocks are generated. Specific details are described in Algorithm 1.

Blocks can be divided into simple blocks and general blocks. Simple block is a goods block composed of identical goods. The details are as shown in Algorithm 2.

The types and placement methods of goods constituting the general block are not limited, and gaps are allowed in the block, but the ratio of the total volume of goods contained in the block to the minimum external cuboid volume of the block shall meet certain requirements, and the bottom of all goods in the block shall be fully supported by the top surface of goods or the bottom of the block to meet the stability constraints. The specific process is shown in Algorithm 3.

Where block b is constructed under the condition that block b_2 and block b_1 should be of equal height when combined along the x -axis and y -axis directions of block b_1 , respectively, as shown in Figure 4(a), and that the length of the rectangular area region is S_{pa} where the upper surface of the block that can provide effective support is $S_l = l_{b_1} + l_{b_2}$, and the width is $S_w = \min(w_{b_1}, w_{b_2})$ when along the x -axis. As shown in Figure 4(b), the length of S_{pa} is $S_l = \min(l_{b_1}, l_{b_2})$ and the width is $S_w = w_{b_1} + w_{b_2}$ when along the y -axis; when combined along the z -axis direction of block b_1 , as shown in Figure 4(c), the conditions that should be satisfied are $l_{b_1} \geq l_{b_2}, w_{b_1} \geq w_{b_2}$, the length of S_{pa} is $S_l = l_{b_2}$ and the width is $S_w = w_{b_2}$.

3.3. Selection of Layout Space. As shown in Figure 5, selecting the layout space by Manhattan distance [21]. Then, we set the coordinates of the vertex of a layout space r in the vehicle as (r_x, r_y, r_z) , and the coordinates of the corresponding vehicle vertex as (R_x, R_y, R_z) , then the Manhattan distance between the two points is $|R_x - r_x| + |R_y - r_y| + |R_z - r_z|$. Considering the stability constraints, it should be ensured

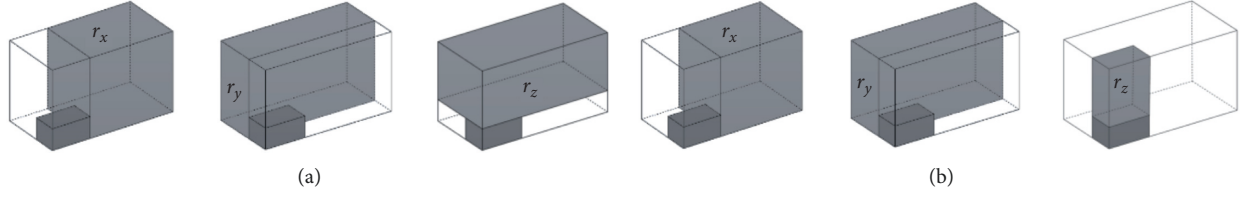


FIGURE 3: Layout space represented by maximum coverage method.

Input mixed goods set C ;

Output heavy goods set C_K , middle goods set C_M , light goods set C_S .

- (1) Arrange the goods set C in descending order according to the weight of class//If the weight is the same, the goods with fewer classes will be ranked first, the same as below
- (2) Take the former $K_{a1} = 0.5n_{c1} + 1 // n_{c1}$ as the total number of goods
- (3) The former K_{a1} goods are ranked as $C_{K1}, C_{K2}, \dots, C_{Ka1}$ in descending order of class density
- (4) Find K_{a2} so that the class density of $C_K = \{C_{K1}, C_{K2}, \dots, C_{Ka2}\}$ is greater than $g_1 kg/m^3$, and get the set C_K
- (5) Delete the goods contained in C_K from C
- (6) Arrange the remaining goods in C in descending order by category volume
- (7) $M_{a1} = 0.5n_{c2} + 1 // n_{c2}$ as the quantity of the remaining goods except heavy goods
- (8) The goods of the former M_{a1} category are arranged in descending order of class density as $C_{M1}, C_{M2}, \dots, C_{Ma1}$
- (9) Find M_{a2} so that the class density of $C_M = \{C_{M1}, C_{M2}, \dots, C_{Ma2}\}$ is greater than $g_1 kg/m^3$, and get the set C_M
- (10) Delete the goods contained in C_K and C_M from C
- (11) The remaining goods in C are arranged in descending order according to class density to obtain set $C_S = \{C_{S1}, C_{S2}, \dots, C_{S(n-K-M)}\}$
- (12) return C_K, C_M, C_S

ALGORITHM 1: Mixed goods classification algorithm.

Input heavy goods set C_K , the maximum number of blocks generated num_K ;

Output simple block set FSB_S .

- (1) Initialize $FSB_S = \emptyset$
- (2) for goods in each category i
- (3) $FSB_i = \emptyset$
- (4) for all permissible ways of placing each class of goods i
- (5) for $el = 1$ to e_i //arrange the goods el times along the x axis, where e_i is the quantity of each type of goods
- (6) for $ew = 1$ to e_i //rearrange the goods ew times along the y axis
- (7) for $eh = 1$ to e_i //rearrange the goods eh times along the z axis
- (8) if $el \cdot ew \cdot eh \leq e$
- (9) Generate block b_s //the sizes are $el \cdot l, ew \cdot w, eh \cdot h$, including $el \cdot ew \cdot eh$ pieces of class i goods
- (10) Put b_s into set FSB_i
- (11) $N = \sum_i N_i$ // N_i is the number of simple goods blocks in each set FSB_i , N is the total number of blocks eventually generated
- (12) if $N > num_K$
- (13) for goods in each category i
- (14) $num = \max(N_i/N \times num_K, 1)$
- (15) Select the top num blocks in each category of goods block set FSB_i and put them into set FSB_S
- (16) else put all the blocks in set FSB_i into set FSB_S
- (17) return FSB_S

ALGORITHM 2: Simple block construction algorithm (Taking heavy goods set as an example).

that the goods block is placed on the bottom of the fully supported layout space. Therefore, only the Manhattan distance between the four corners of the bottom of the layout space and the four corners of the vehicle bottom should be considered, and the goods block should be placed at the bottom corner of the space corresponding to the shortest Manhattan distance.

3.4. Selection of Goods Blocks. Using $V(b)$ as the volume of goods block b , $C(b, p)$ as the direct contact or indirect projection coverage, $L(b, r)$ as the spatial loss rate [16], $N(b)$ as the number of goods contained in the goods block, and $W(b)$ as the weight of the goods block, then the goods block selection evaluation function of formula (17) is constructed. When the parameter $\lambda' p' \varphi' \gamma' \delta' \varepsilon$ takes a suitable value, the

Input heavy goods set C_K , generate maximum number of blocks num_K and effective volume ratio $vol\%$;
 Output General block set FSB_G .

- (1) Initialize $FSB_G = \emptyset$, $FSB'_G = \emptyset$
- (2) $FSB_G = C_K$, $FSB'_G = C_K$
- (3) while $FSB_G < num_K$ and $FSB'_G \neq \emptyset$
- (4) $GB = \emptyset$
- (5) for block b_1 in each set FSB'_G
- (6) for block b_2 in each set FSB_G
- (7) A new block b is generated by lacing b_2 along the three axes of b_1 in all possible ways
- (8) If b meets the criteria and there is no identical block in the FSB_G , b is put into GB
- (9) $FSB_G = FSB_G \cup GB$ and the number does not exceed num_K
- (10) $FSB'_G = GB$
- (11) return FSB_G

ALGORITHM 3: General block construction algorithm (Taking heavy goods set as an example).

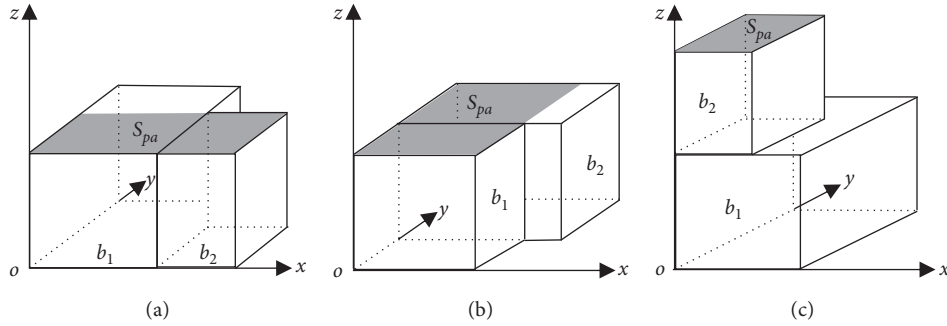


FIGURE 4: Diagram of general block construction conditions.

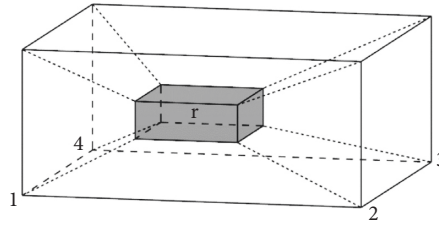


FIGURE 5: The Manhattan distance between the layout space and each vertex of the freight vehicle.

higher the value of the evaluation function, the more likely the goods block is selected in preference.

$$f(b, r) = V(b)^\lambda \cdot C(b, p)^\theta \cdot (1 - L(b, r))^\gamma \cdot N(b)^{-\delta} \cdot W(b)^\epsilon. \quad (17)$$

In order to load the local optimal goods block as much as possible, based on the greedy d-step lookahead tree search, the goods block is selected with formula (17) as the evaluation function, and the goods block selection process is constructed as Algorithm 4. In Figure 6, if the bold path is the optimal layout scheme, the local optimal goods block that should be put into the current r is b_2 .

3.5. Placement of Cargo Blocks and Updating of Layout Space. Firstly, place the heavy goods that have a great impact on the position of the total center of gravity, place the heavy goods

close to the corner of the vehicle, and then place the middle goods and light goods. Load the goods in the order from the corner of the vehicle to the center, so that the final remaining debris space is concentrated in the center of the vehicle, and the goods with large weight or density are distributed near both sides of the vehicle bogie, so as to reduce space waste, ensure the balance of center of gravity, and avoid concentrated-weight. When the goods are placed, there is no available space or the total weight of goods exceeds the vehicle maximum allowable load constraints, so the layout shall be ended.

When a goods block is placed into a space, a new layout space is created, and the space needs to be updated to delete the original space, the duplicate space, the contained small space, and the part where the goods block crosses and overlap with the existing space, and keep or create the remaining available and fully supported space. Since the

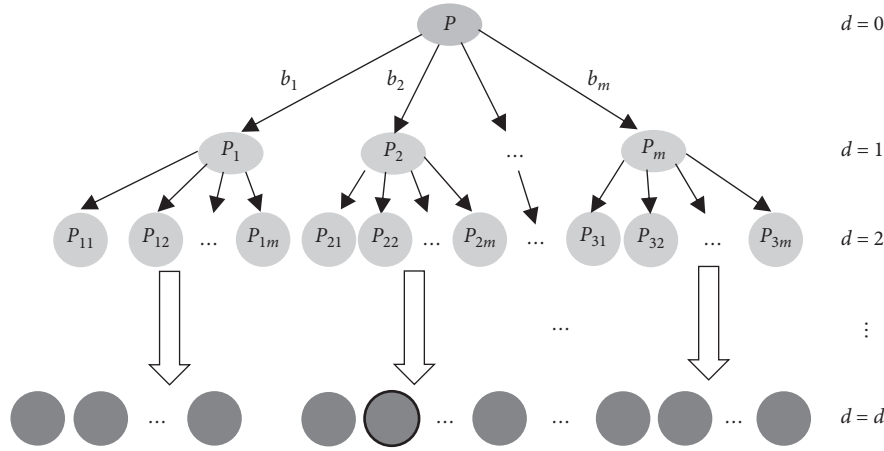


FIGURE 6: Diagram of cargo block selection based on greedy d-step lookahead tree search.

Input Space set \mathbf{R} , current layout space r , goods block set \mathbf{B}_S or \mathbf{B}_G , parameter $\lambda' p' \varphi' \gamma' \delta' \varepsilon$, the width coefficient m and depth coefficient d of the tree;

OutputPlace the locally optimal goods block b in the current layout space r .

- (1) Initialize $d = 0$, create an initial state P in the current layout space r as the root node
- (2) for each layer of tree// d takes $1, 2 \dots d$
- (3) Select m blocks with $f(b, r)$ values in the first m to generate m sub-states//generate m incomplete layout schemes
- (4) If $(d = d)$ //at layer d , each leaf node in this layer is a complete layout scheme, with $w = m^d$ schemes in total
- (5) Choose the scheme that produces the optimal solution out of w schemes
- (6) return place the locally optimal block b of the current layout space r

ALGORITHM 4: Goods block selection algorithm.

existing spaces all satisfy the stability constraint, there are 50 possible cases of cross overlap between goods blocks and spaces; 4, 8, 5, and 1 cases of 1, 2, 4, and 8 vertices of goods blocks in spaces; 8, 5, and 1 cases of 2, 4, and 8 vertices of spaces in goods blocks; 12, 3 cases of goods blocks running along the boundary of spaces along the direction of coordinate axes and from the middle of spaces, respectively; 3 cases where the goods block divides the space into two parts.

3.6. Overall Algorithm for Loading Layout. The above process is generally adjusted and optimized by introducing an indicator t , $t = \sum_n u/n$ to distinguish the goods structure. When $t > 6$, it is a weakly heterogeneous problem, and a simple block is constructed. When $t \leq 6$, it is a strongly heterogeneous problem, then a general block is constructed. When the computation time is left, a doubled greedy d-step lookahead tree search is performed. To sum up, the overall flow of railway mixed goods balanced and anticoncentrated-weight loading layout optimization considering stability is shown in Algorithm 5.

4. Example Analysis

To verify the validity and feasibility of the proposed method, 1600 international standard arithmetic cases used to evaluate the loading layout problem are adopted [23], on which the

goods weight is considered, and the range of goods density is set as $(Q_T/2V_T, Q_T(1/E[x] + 1)/2V_T)$, Q_T/V_T is the average density of the goods, $E[x]$ is the beta distribution expectation, and the goods density of the arithmetic cases is obtained according to $f(x, 2, 5)$ [24]. The arithmetic cases are numbered from BR0-BR15 with 16 groups, each group has 100 arithmetic cases, and the goods types are increased from 1 class of BR0 to 100 classes of BR15, where BR0 is a homogeneous class problem, BR1-7 is a weakly heterogeneous class problem, and BR8-BR15 is a strongly heterogeneous class problem. P_{60} vehicle is selected for shipment, and its relevant attribute parameters are as follows: $D = 11500$ mm, $H_v = 1144$ mm, $H_o = 1315$ mm, $Q = 60$ t, $Q_v = 22.2$ t, $L = 15470$ mm, $W = 2830$ mm, $H = 2750$ mm, and $V = 120$ m³. Since there is no unified standard for the soak weight ratio of land goods, referred to the standard of soak weight ratio of other transportation modes and combined with the actual railway transportation, 300 kg/m³ and 167 kg/m³ are adopted to g_1 and g_2 in the Algorithm 1, respectively. The values of parameters λ and ε in the selection evaluation function for the goods block are determined to be 0.4 and 0.2, respectively, using the control variable method; and the other parameter values are referred to the work studied by Araya et al. [8], which are $\varphi = 4$, $\gamma = 1$, $\delta = 0.2$, $p = 0.04$. Moreover, the search tree depth factor of $d = 2$ is used to provide a better trade-off between search accuracy and running time. The average

Input Space set \mathbf{R} , mixed goods set \mathbf{C} , computation time T , maximum number of blocks generated num_K , effective volume fraction $vol\%$, goods block evaluation function parameters $\lambda' p' \phi' \gamma' \delta' \varepsilon$, the width coefficient m and depth coefficient d of the tree; Output loading layout scheme S_{best} .

- (1) Initialize the space as the rectangular space inside the vehicle, and classify the mixed goods set//call Algorithm 1
- (2) Judge the problem type with index t and generate the set of heavy goods, middle goods and light goods blocks of the corresponding type//call algorithm 2 or 3
- (3) while there is computation time do
 - (4) $\mathbf{R}' = \mathbf{R}$, $\mathbf{C}' = \mathbf{C}$
 - (5) while $\mathbf{C}' \neq \emptyset$ do
 - (6) if there is available efficient space in \mathbf{R}'
 - (7) Choose the space r with the shortest distance in Manhattan
 - (8) According to the evaluation function, select the top m blocks from the block set//call Algorithm 4, place heavy goods blocks first, then middle goods blocks, then light goods blocks
 - (9) if b is the block that should be placed in the current layout space then
 - (10) Put it at the corresponding angle of r
 - (11) Update \mathbf{R}' , \mathbf{C}' , and goods block set
 - (12) else the flag r is no longer selected until \mathbf{R}' is updated again
 - (13) end if
 - (14) else break
 - (15) end if
- (16) end while
- (17) Generate a loading layout scheme s
- (18) if $s > S_{best}$ then $S_{best} = s$
- (19) end if
- (20) $w = 2w, m = \sqrt[4]{w}$ //Double the iterative search, let $m = \sqrt[4]{2}m$ continues to generate a new layout scheme
- (21) end while
- (22) return S_{best}

ALGORITHM 5: Loading layout optimization overall algorithm.

running time of 1600 cases is approximately 100 s (± 10 s), the test computer processor is Intel(R) Core(TM) i7-8565U CPU @1.80 GHz, the running memory is 8 GB, the operation system is Windows 10 (64-bit). And then the algorithm is implemented by Java language (JDK version 1.8, 64-bit), which is compiled by Eclipse software (Oxygen version).

4.1. Comparative Analysis of International Standard Examples. Based on the above, the scenarios of simple and general blocks are constructed by the test standard cases, and the results of that are shown in Table 1. The results are the average results of the same group including 100 cases, “Vol” is the effective volume utilization rate, “Wt” is the load capacity utilization rate, “Sum” is the comprehensive utilization rate, “AV” is the average of the results of certain groups of cases, “Unb” is the number of cases that dissatisfy the constraint of loading center-of-gravity balance, and “WtC” is the number of cases that dissatisfy the allowable moment of concentrated-weight constraint.

By comparing and analyzing the results of columns *SB* and *GB*, it can be seen that the overall results in constructing general blocks are better than those of simple blocks, and the average volume utilization rate, load capacity utilization rate, and comprehensive utilization rate of the proposed method can reach more than 92%, 87%, and 89%, respectively. From “Unb” and “WtC”, it can be inferred that the satisfaction rate of the loading center-of-gravity balance constraint can meet 99.8%, and the satisfaction rate of the concentrated-weight allowable bending moment constraint can reach to 99.87%, so

the loading effect is beneficial enough to meet the actual loading. In addition, for BR0-BR7, the results are better than general blocks when constructing simple blocks, which is the opposite for BR8-BR15. Thus, judging the goods structure type by indicator t to construct the corresponding goods blocks can combine advantages to make better loading results.

4.2. Comparative Analysis of Mixed Goods Examples Generated Based on Standard Examples. To further verify the effectiveness of the proposed method for mixed goods with greater differences in density, volume, and weight, we increase the goods size, weight interval, and beta distribution on the basis of international standard cases, and generate mixed goods examples. Simple blocks are constructed for BR0-BR7 and composite blocks are constructed for BR8-BR15. The test is carried out with and without consideration of stability.

From the results of Table 1, it is obvious that for the mixed goods examples considering stability, the average volume utilization rate, load utilization rate, and comprehensive utilization rate of the proposed method are lower than the standard examples, while they can reach more than 91%, 86%, and 89% respectively, and the satisfaction rate of the loading center-of-gravity balance constraint and the concentrated-weight allowable bending moment constraint can meet 99% and 99.47%, respectively. The average volume utilization rate, load capacity utilization rate, and comprehensive utilization rate are slightly higher for the mixed goods examples without considering the stability, but the

TABLE 1: Test results under different examples and conditions.

Examples	International standard goods calculation examples										Mixed goods calculation examples (improved on the basis of international standard calculation example)										
	SB (Simple block, consider stability)					GB (General block, consider stability)					Construct SB or GB from t (regardless of stability)					Construct SB or GB from t (regardless of stability)					
	Vol. (%)	Wt. (%)	Sum. (%)	Unb. (/)	WtC. (/)	Vol. (%)	Wt. (%)	Sum. (%)	Unb. (/)	WtC. (/)	Vol. (%)	Wt. (%)	Sum. (%)	Unb. (/)	WtC. (/)	Vol. (%)	Wt. (%)	Sum. (%)	Unb. (/)	WtC. (/)	
Homogeneous	BR0	89.67	82.35	85.86	1	1	89.90	82.17	85.86	2	2	88.82	81.96	85.25	1	1	89.59	82.34	85.81	5	5
	BR1	91.52	82.80	86.94	0	0	91.22	82.62	86.71	1	1	90.06	82.19	85.95	3	2	91.33	82.89	86.90	4	3
	BR2	91.74	83.75	87.56	0	0	91.33	83.57	87.28	0	0	90.92	83.34	86.97	2	2	91.89	83.76	87.64	9	4
	BR3	92.59	84.80	88.52	0	0	92.09	84.82	88.31	0	0	92.05	84.30	88.00	3	1	92.12	84.43	88.11	7	5
Weak heterogeneous	BR4	92.96	84.74	88.66	1	1	92.52	84.60	88.38	2	1	92.29	85.13	88.57	1	0	92.86	85.69	89.13	3	0
	BR5	93.14	85.92	89.38	2	1	92.85	85.73	89.15	0	0	92.43	85.36	88.75	0	0	93.04	86.88	89.85	5	2
	BR6	92.99	86.83	89.81	0	0	92.76	86.52	89.53	0	0	92.70	85.31	88.86	2	0	92.79	87.17	89.89	6	3
	BR7	93.39	87.95	90.59	1	0	93.06	87.83	90.37	0	0	92.91	86.58	89.63	1	1	93.43	87.32	90.27	4	0
	BR8	92.87	88.53	90.65	0	0	93.12	88.91	90.96	0	0	92.85	87.14	89.90	2	2	92.57	87.79	90.12	0	0
	BR9	91.93	89.39	90.64	0	0	92.38	89.87	91.11	0	0	92.31	88.62	90.43	0	0	92.83	88.41	90.57	3	2
	BR10	92.86	89.68	91.24	2	2	93.24	90.13	91.66	0	0	92.23	88.71	90.44	1	0	92.70	88.86	90.74	1	1
Strong heterogeneous	BR11	91.61	88.94	90.25	1	0	92.29	89.62	90.93	0	0	91.77	88.91	90.32	0	0	92.65	89.26	90.92	2	1
	BR12	91.26	89.53	90.39	0	0	92.43	89.84	91.12	0	0	91.56	88.57	90.04	0	0	91.42	89.74	90.57	0	0
	BR13	91.47	89.12	90.28	1	0	92.34	89.66	90.98	0	0	91.11	88.23	89.65	0	0	92.17	89.85	91.00	1	0
	BR14	90.81	88.61	89.69	0	1	91.16	89.76	90.45	0	0	91.06	87.40	89.19	0	0	91.64	89.91	90.77	0	0
	BR15	90.47	89.27	89.87	2	2	90.73	89.58	90.15	0	0	90.46	87.95	89.19	0	0	91.23	89.47	90.34	0	0
	AV.1-7	92.62	85.25	88.78	4	2	92.26	85.10	88.53	3	2	91.91	84.60	88.10	12	6	92.49	85.45	88.83	38	17
Mean	AV.8-15	91.66	89.13	90.38	6	5	92.21	89.67	90.92	0	0	91.67	88.19	89.89	3	2	92.15	89.16	90.63	7	4
	AV.1-15	92.11	87.32	89.63	10	7	92.23	87.54	89.81	3	2	91.78	86.52	89.06	15	8	92.31	87.43	89.79	45	21

satisfaction rate of the load center-of-gravity balance constraint and the allowable bending moment constraint of the concentrated-weight is much worse, so the comparison shows that the proposed method can achieve favorable loading effects on basis of ensuring the safety and stability of mixed goods transportation.

5. Conclusion

- (1) In this paper, several factors are considered, such as the balance of the combined center of gravity after goods loading, concentrated-weight and full support, and the effective volume and load capacity of the freight vehicle are included in the optimization objective. Then the problem optimization model is constructed. Next, we designed algorithms and rules such as mixed cargo classification, fully supported goods block unit generation, goods block unit selection and placement, and update of remaining available layout space. Furthermore, an optimization method of railway mixed goods balanced and anti-concentrated-weight load layout considering stability is proposed.
- (2) The proposed method not only meet the goods loading stability but also ensure the comprehensive utilization rate of effective volume and load capacity of the vehicle which is not less than 89%, loading center-of-gravity balance, concentrated-weight allowable bending moment constraints to meet the probability of up to 99% and 99.47%. The method effectively realizes the railway mixed goods balanced and anticoncentrated-weight efficient loading, ensures the safety of railway goods shipment, and provides decision-making reference for railway goods loading layout.
- (3) The problem of railway goods loading layout is complex and diverse. For simplicity, the interior of the freight vehicle is assumed to be a cuboid space, and the loading layout is carried out from the angle of the loading space. In the future, the problem algorithm should be further improved by considering the specific loading door position and shape attributes of the vehicle.

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 projects funded by the National Natural Science Foundation of China (Grant Nos. 71971220 and 71901093), the Natural Science Foundation of Hunan Province of China (Grant No.

2022JJ31020), the Research Foundation of Education Bureau of Hunan Province, China (Grant No. 20B597), and Hunan Provincial Fund for Philosophy and Social Sciences Youth Project, China (Grant No. 18YBQ139). Their support is gratefully acknowledged.

References

- [1] L. I. Peng and Y. Tang, "Review on three-dimensional bin packing problem," *Journal of Railway Science and Engineering*, vol. 12, no. 05, pp. 1232–1242, 2015.
- [2] Y. H. Huang, F. J. Hwang, and H. C. Lu, "An effective placement method for the single container loading problem," *Computers & Industrial Engineering*, vol. 97, pp. 212–221, 2016.
- [3] M. D. G. Costa and M. E. Captivo, "Weight distribution in container loading: a case study," *International Transactions in Operational Research*, vol. 23, no. 1-2, pp. 239–263, 2016.
- [4] A. G. Ramos, J. F. Oliveira, and M. P. Lopes, "A physical packing sequence algorithm for the container loading problem with static mechanical equilibrium conditions," *International Transactions in Operational Research*, vol. 23, no. 1-2, pp. 215–238, 2016.
- [5] A. G. Ramos, J. F. Oliveira, and J. F. Goncalves, "A container loading algorithm with static mechanical[J]," *Transportation Research Part B*, vol. 91, pp. 565–581, 2016.
- [6] A. Bortfeldt and H. Gehring, "A hybrid genetic algorithm for the container loading problem," *European Journal of Operational Research*, vol. 131, no. 1, pp. 143–161, 2001.
- [7] X. Zhu, L. E. I. Ding-you, and Y. Zhang, "Research on optimization of loading of multi-piece concentrated-weight freight into one car[J]," *Journal of the China Railway Society*, vol. 35, no. 9, pp. 7–13, 2013.
- [8] X. Zhu and Y. Xiang, "An optimization of multi-freights loading into multi-cars with balancing constraints[J]," *Industrial Engineering Journal*, vol. 23, no. 03, pp. 123–131, 2020.
- [9] T. Fanslau and A. Bortfeldt, "A tree search algorithm for solving the container loading problem," *INFORMS Journal on Computing*, vol. 22, no. 2, pp. 222–235, 2010.
- [10] Y. Jiang, C. H. A. Jian-zhong, and D. He, "Research on the packing of loading rectangular freight into a container[J]," *Journal of the China Railway Society*, vol. 22, no. 6, pp. 13–18, 2000.
- [11] A. Moura and J. F. Oliveira, "A GRASP approach to the container-loading problem," *IEEE Intelligent Systems*, vol. 20, no. 4, pp. 50–57, 2005.
- [12] I. Araya, K. Guerrero, and E. Nuñez, "VCS: a new heuristic function for selecting boxes in the single container loading problem," *Computers & Operations Research*, vol. 82, pp. 27–35, 2017.
- [13] Z. Wang and J. Liu, "Optimization model and algorithm of the one-kind-goods loading problem on the railway[J]," *Journal of Railway Science and Engineering*, vol. 8, no. 05, pp. 107–112, 2011.
- [14] L. I. U. Xiao-qun and M. A. Shi-hua, "Optimization algorithm of multi-truck multi-category goods loading based on benchmark methods[J]," *Journal of Traffic and Transportation Engineering*, vol. 7, no. 01, pp. 99–105, 2007.
- [15] E. E. Bischoff, "Three-dimensional packing of items with limited load bearing strength," *European Journal of Operational Research*, vol. 168, no. 3, pp. 952–966, 2006.
- [16] L. E. I. Ding-you, Y. A. N. Hong-ying, H. O. N. G. Shu-hua et al., "Research on mixed and balanced loading of light and

- heavy cargo in railway container[J],” *Journal of the China Railway Society*, vol. 43, no. 01, pp. 1–9, 2021.
- [17] Y. Huang, L. Lai, W. Li, and H Wang, “A differential evolution algorithm with ternary search tree for solving the three-dimensional packing problem,” *Information Sciences*, vol. 606, pp. 440–452, 2022.
- [18] Y Zhang, Y Yao, and Q. Gao, “Optimization model and algorithm of balanced loading layout of railway container mixed goods[J],” *Journal of Transportation Systems Engineering and Information Technology*, vol. 22, no. 02, pp. 214–222, 2022.
- [19] Y Wang, “A heuristic approach to container loading problem based on cargo combination[J],” *Logistics Engineering and Management*, vol. 40, no. 12, pp. 73–75, 2018.
- [20] *China Railway Corporation, Rules for railway cargo loading and reinforcement[S]*, Vol. 296, China Railway Corporation, Beijing, China, 2015.
- [21] W. Zhu, W. C. Oon, A. Lim, and Y Weng, “The six elements to block-building approaches for the single container loading problem,” *Applied Intelligence*, vol. 37, no. 3, pp. 431–445, 2012.
- [22] I. Araya and M. C. Riff, “A beam search approach to the container loading problem,” *Computers & Operations Research*, vol. 43, pp. 100–107, 2014.
- [23] E. E. Bischoff and M. S. W. Ratcliff, “Issues in the development of approaches to container loading,” *Omega*, vol. 23, no. 4, pp. 377–390, 1995.
- [24] A. G. Ramos, E. Silva, and J. F. Oliveira, “A new load balance methodology for container loading problem in road transportation,” *European Journal of Operational Research*, vol. 266, no. 3, pp. 1140–1152, 2018.

Research Article

Concrete Surface Crack Recognition Based on Coordinate Attention Neural Networks

Yuhao Zhang and Zhongwei Wang 

School of Logistics and Transportation, Central South University of Forestry and Technology, Changsha 410004, China

Correspondence should be addressed to Zhongwei Wang; 20191200203@csuft.edu.cn

Received 1 July 2022; Revised 16 July 2022; Accepted 19 July 2022; Published 11 August 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yuhao Zhang and Zhongwei 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.

In highway transportation infrastructure such as highways and tunnels, the proportion of concrete consumption is the highest, and concrete cracks are common concrete problems. Concrete cracks will greatly affect the bearing capacity and safety of the structure, easily leading to the interruption of transportation lines, causing great economic losses, and endangering personnel safety. Therefore, the effective identification and timely reporting of concrete cracks is of great significance for the maintenance of infrastructure such as roads and tunnels. In this paper, the CaNet, a deep learning network for identifying concrete cracks, is proposed, which takes ResNet50 as the backbone network. In order to capture the area with a small proportion of cracks, we added coordinate attention to the residual unit of ResNet50 to capture the cross-channel information, direction-aware information, and position-sensitive information from many vertical and horizontal directions so that the network can more accurately locate the narrow crack area. In experiments 3.2 and 3.3, the CaNet has an accuracy rate of 89.6%, which is higher than that of the compared network. In addition, the recall, F1 score, and precision of the CaNet network are 86%, 85%, and 87%, respectively. Therefore, the CaNet model is effective for identifying concrete cracks.

1. Introduction

The rapid take-off of the world economy benefits from the stable and sustainable development of the transportation industry to a certain extent. Highways and tunnels are the carriers of the land transportation industry [1], and their structure directly affects whether the transportation lines can continue to operate. However, with the increase in operation time of highway and tunnel facilities and the influence of external environmental factors, certain damage will inevitably occur to highway and tunnel engineering structures [2, 3]. To maintain the normal operation of the transport lines, the transportation department needs to continue the health inspection of highways and tunnels regularly. In highway and tunnel engineering structures, the proportion of concrete consumption is the highest, and shrinkage cracks, broken plates, and ash will cause concrete quality defects, which will not only affect the overall beauty but also affect the quality of the highway and tunnel [4]. Among

them, shrinkage concrete cracks are common concrete problems, which are manifested in the process of condensation. If the water evaporates quickly, it will indicate that it is easy to shrink, thus forming irregular or penetrating cracks [5]. The occurrence of cracks will not only reduce the impermeability of buildings but also affect the use function of buildings, causing corrosion of reinforcement, carbonation of concrete, and reducing the durability of materials. Finally, it will affect the bearing capacity of the building. Therefore, timely and effective detection of highway and tunnel surface cracks, detection of possible safety problems, and repair can effectively prevent accidents [6].

In real life, manual patrol inspection is still used for the patrol inspection and maintenance of cracks on the surface of highways and tunnels, relying on close observation with the help of width measuring instruments, tape, and other measuring instruments to complete the collection of crack information and mark and record the crack information in time [7]. However, manual patrol inspection has many

disadvantages: (1) manual patrol inspection is easily disturbed by human subjective factors and is subject to the experience of patrol inspectors. (2) The safety of personnel cannot be effectively guaranteed when the crack at the high position is inspected, or the detection position is in a relatively steep place. (3) With the development of the world economy, the number of roads, tunnels, and other infrastructure is increasing day by day. If the manual inspection method is adopted, it will cause a lot of waste of human, material, and financial resources. Moreover, the recorded data information is manually recorded, making it difficult to form a complete database, cannot be preserved for a long time, and cannot be comprehensively analyzed with big data technology. As a result, the transportation department must still figure out how to complete concrete surface detection quickly and accurately [8].

To solve the problems encountered in manual inspection, the traditional digital image processing technology for concrete cracks is widely used [9]. For example, Talab et al. [10] proposed a method to identify cracks in concrete images by taking advantage of the characteristics that cracks are dark areas in the image that have a large gap with the background. First, Sobel is used to extract the edges of grayscale images. The foreground and background of all pixels are separated by the threshold method; then, the Sobel or median filter is used to eliminate residual noise. Finally, the Otsu method is used for crack detection. However, the method based on edge detection is easily disturbed by the background information, and it is difficult to recognize the crack features in the image under a complex background. In order to reduce the influence of background on recognition, Shi et al. [11] used the random structured forest method to realize the image detection of complex road cracks, which to some extent solved the problem of difficult recognition under complex background, but the effect was not significant enough. Although the above processing methods have achieved good recognition results, the feature extractor is extremely dependent on the professional knowledge and design experience of designers, and the designed algorithm is only effective in the characteristic scene with low generalization.

Compared with the traditional digital image processing technology, the extensive promotion of deep learning has greatly improved the ability of concrete crack feature extraction and feature expression [12]. At the same time, by using a large amount of data for end-to-end self-learning, they can solve well the problem of insufficient generalization. Cha et al. [13] combined the SWT with CNN to detect concrete cracks, which can scan any crack greater than a 256×256 resolution image. Experiments show that the method combined with SWT and CNN has better performance in concrete crack identification than the Canny and Sobel detection methods, and the recording accuracy is about 98%; pan et al. [14] developed a layered network called SCHNet, which is composed of the VGG19 backbone network and a self-attention mechanism, and can support pixel-level automatic and reliable concrete crack segmentation. Experiments show that the average IOU reaches 85.31%. Wang et al. [15] used the improved inception

RESNET V2 model to identify bridge cracks. The experimental results show that each performance index has been improved to varying degrees: the accuracy is 99.24%; the recall rate is 99.03%; fF-measure is 98.79%; FPS is 196. Nayyeri and Zhou [16] based the MR-CrackNet model to identify cracks of different sizes in roads and bridge decks, and trained and tested it on a new crack dataset containing 2,532 concrete crack images. The experimental results show that MR-CrackNet is superior to the baseline model and can achieve high-precision crack identification. In this paper, the structural features of concrete cracks are extracted by designing the deep learning network structure to distinguish whether the concrete in the image has cracks.

Based on ResNet50 in the ResNet series [17] as the backbone network and adding coordinate attention to the residual unit, this paper proposes CaNet. Coordinate Attention (CA) [18] considers the location information of concrete cracks through the vertical and horizontal characteristic directions and models the long-term dependence so as to enhance the weight of the crack area and distinguish it from useless information such as background.

The main contributions of this paper are as follows:

- (1) The CaNet is proposed: coordinate attention is introduced, which enhances the extraction of micro features of concrete cracks and improves the recognition accuracy of crack images by 6.2% to ResNet50.
- (2) Compared with several neural networks set up in the experiment, the CaNet has a higher recognition rate in the identification of concrete cracks.

In the second section, we introduced the dataset (2.1) used in this paper and the constituent elements of the CaNet (2.2); the third chapter introduces the experiment performed in this paper, in which 3.1 describes the hardware and software configured for this experiment; 3.2 tests the practicality of the model; 3.3 tests the effectiveness of coordinated attention; 3.4 tests the comparison with other models. The fourth chapter introduces the conclusions and shortcomings of this paper.

2. Materials and Methods

2.1. Data Acquisition. The quality of the concrete crack dataset can greatly affect the designed network [19, 20]. This time, the concrete crack data set is from Kaggle's surface crack detection public data set [21]. The public dataset has 40000 images, all of which are RGB channels, 227×227 pixels, including 20000 images without cracks and 20000 images with cracks. Some images are shown in Figure 1. At the same time, when examining the dataset, it is found that the high-resolution images have great differences in surface finish and lighting conditions, as shown in Figure 2.

2.2. Concrete Crack Identification Based on CaNet Model. In CaNet, the main structure of ResNet50 is used. On each floor, 1×1 and 3×3 , coordinate attention is added between convolution blocks. The network structure of CaNet is shown in Figure 3.

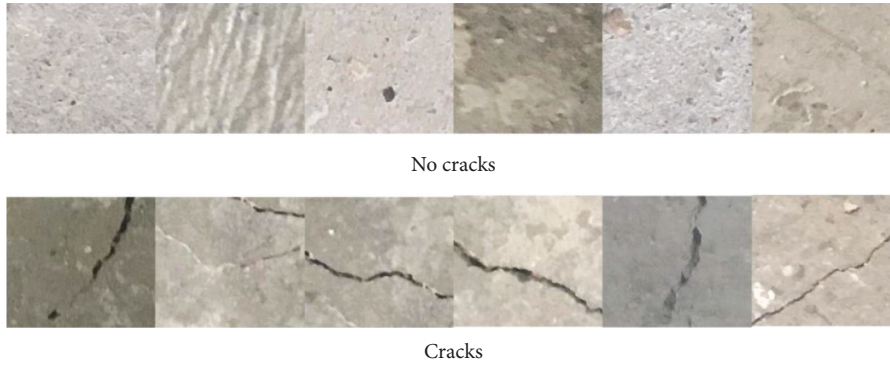


FIGURE 1: Partial image display.



FIGURE 2: Partial image difference.

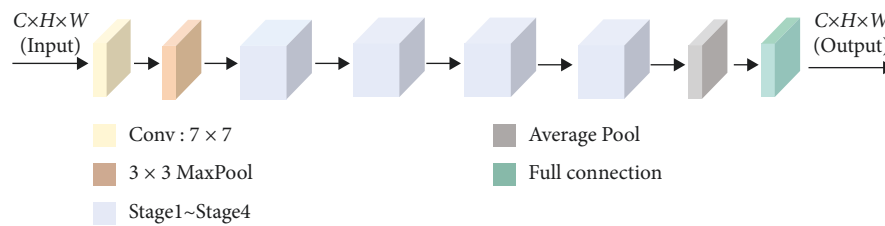


FIGURE 3: The overall architecture of CaNet.

2.2.1. *ResNet50 Model.* In the field of recognising concrete crack images, the depth convolution neural network has produced ideal results recently [22]. The amount of image information that the deep convolution network can learn increases with the number of network layers [23]. However, as the number of network layers rises, over-fitting and network model degradation issues will also arise, making it more difficult to train models and preventing the system from convergence. He K and others [15] created a residual block by incorporating shortcut connections into the original DCNN [24] network structure (RB). Figure 4 depicts the RB structure.

ResNet50 is a network structure commonly used in various fields, which is formed by stacking different residual blocks. Residual block (RB) is shown in formula :

$$H = F(x, \{W\}) + x, \tag{1}$$

x represents the input characteristic diagram of RB; H represents the output characteristic diagram of RB; $F(x, \{W\})$ represents the residual mapping of the network. The dimension of input x must be the same. If the dimension is different, W_q linear projection should not be added to the shortcut connection, as shown in the following formula:

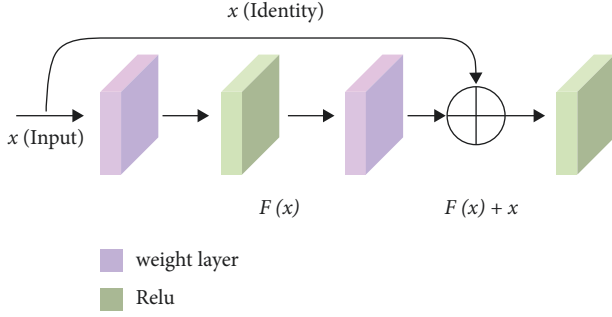


FIGURE 4: Residual block.

$$H = F(x, \{W\}) + Q(x, \{W_q\}). \quad (2)$$

2.2.2. Coordinate Attention. The 2D global pool is used to calculate channel attention in the most widely used squeeze and stimulate (SE) attention mechanism. However, SE only takes into account the coding of information between crack channels for concrete crack images, ignoring the significance of concrete crack location data, which is crucial for capturing the object structure in concrete crack recognition. Using the convolution layer to calculate the spatial attention information of concrete cracks, BAM [25] and CBAM [26] attempt to build the crack location information based on the SE attention mechanism's weakness. The long-term dependency required in computer vision tasks cannot be established by convolution though.

In order to address this issue, Hou et al. [18] suggested coordinate attention, which combines the vertical and horizontal inputs into two distinct directional graphs. The location data for cracks are then encoded into two crack attention maps from two-directional maps with different directions of cracks. This allows the generated crack attention map to store the location data for cracks. In order to emphasise the expression of interest, the two crack attention maps are finally applied to the input through multiplication. Figure 5 illustrates the precise organisation of coordinate attention.

Specific implementation process is

Step 1. Each channel of input X is encoded using a pooling kernel of size $(H, 1)$ or $(1, W)$, respectively.

The crack features output of c channel with h height can be expressed as

$$z_C^h(h) = \frac{1}{W} \sum_{0 \leq i < w} x_C(h, i). \quad (3)$$

The crack features output of c channel with w width can be expressed as

$$z_C^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_C(j, w). \quad (4)$$

Step 2. After the transformation in information embedding, this part concatenates the feature map after step 1 transformation and then uses the convolution to transform it.

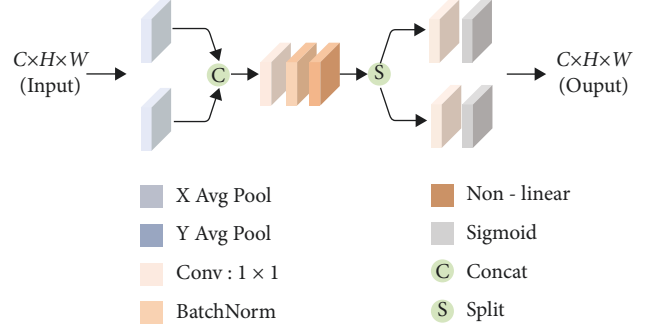


FIGURE 5: Coordinate attention.

$$f = \delta(F_1([z^h, z^w])), \quad (5)$$

where F_1 stands for 1×1 convolution kernel; δ indicates nonlinear activation function; $f \in \mathbb{R}^{C/r \times (H+W)}$ represents the intermediate feature map for coding the spatial information of cracks, and r represents the reduction rate of control block size.

Step 3. Divide $f \in \mathbb{R}^{C/r \times (H+W)}$ in step 2 into $f^h \in \mathbb{R}^{C/r \times H}$ and $f^w \in \mathbb{R}^{C/r \times W}$. Then, use 1×1 convolution kernels to transform f^h and f^w .

$$\begin{aligned} g^h &= \sigma(F_h(f^h)), \\ g^w &= \sigma(F_w(f^w)). \end{aligned} \quad (6)$$

σ stands for the sigmoid function. Both F_h and F_w represent 1×1 convolution. g^h and g^w represent weight characteristics.

Step 4. Use multiplication to fuse weights g^h and g^w with input X and output coordinate attention block Y , as shown in the following equation.

$$y_C(i, j) = x_C(i, j) \times g_C^h(i) \times g_C^w(j). \quad (7)$$

3. Results and Analysis

3.1. Experimental Environment and Setting. In this experiment, the network training and testing environments were carried out on a platform. The experimental parameters of training and testing are shown in Table 1.

The concrete crack dataset used this time is the Surface Crack Detection public dataset with 40000 images, which are divided into two types, crack free images and crack images. In all experiments of this paper, according to the ratio of 7 : 2 : 1, the dataset will be divided into the training set, test set, and verification set.

3.2. Application Scenarios. Concrete is indispensable in today's construction projects and is the most widely used building material. It is widely used in the construction of infrastructure such as roads, bridges, housing construction, tunnels, and dams. Concrete cracks are the most common quality problem in concrete subprojects. Due to the low

TABLE 1: Experimental parameters of CaNet.

Parameter type	Parameter name	Parameter setting
<i>Hardware environment</i>	Operating system	Windows 10 (64 bit)
	Version number	20H2
	RAM	64 GB(63.8 available)
	Processor	Intel(R) core(TM) i7-9700u
	Graphics card	2080ti GPU
<i>Software environment</i>	CaNet	Python 3.8.12 Pytorch 1.8.2 CUDA 10.2
	Batch size(training)	32
<i>Training parameter setting</i>	Batch size(testing)	8
	Learning rate lr	10^{-3}
	Epoches	140
	Optimizer	Adam optimizer
	Loss function	Cross entropy loss
	Training method	Incremental gradient descent

tensile strength of concrete and the combined effects of internal and external influences, such as shrinkage and creep, external temperature changes, and foundation deformation, concrete in construction and operation is used. Structures often have cracks to varying degrees and forms, and it is still impossible to completely prevent and eliminate them. Therefore, efficient inspection and monitoring of concrete cracks have become the key to ensuring the quality of concrete projects. Accurately identifying the length, direction, and width of cracks in concrete structures plays an important role in judging the disease degree and operating conditions of the structure.

Surface-based crack identification is a considerable task in health monitoring of structures. If cracks continue to expand, they will reduce the effective bearing surface area and can lead to structural failure over time. On the one hand, the detection process of manual cracks is time-consuming and labor-intensive, subject to the subjective judgment of inspectors, and for high-rise buildings and bridges, due to the large operational limitations, manual inspection may also be difficult to perform. On the other hand, although the image recognition technology of AI has been improved dramatically, and the technology of automatically detecting cracks through image recognition has been developed, most of them still require manual image acquisition work, and a lot of high-resolution image data are obtained, but it still takes a considerable amount of time and cost to record the crack detection and make the result file. Therefore, in this paper, we build a convenient and accurate crack identification model based on coordinate attention and test the model on real-world data (the recognition process is shown in Figure 6) and find that the model detects concrete and nonconcrete structure example roads that are accurate in terms of surface cracks. The results of experiment 3.3 show that the target recognition speed and accuracy of concrete surface crack detection technology based on coordinate attention can meet the requirements of real time. It has been widely used in the automatic identification of structural cracks in concrete structures of roads, bridges, and dams and has achieved good practical application results. However,

there are still some problems to be solved in the application of the actual scene of structural crack detection, especially for some situations where the on-site detection environment is more complicated, such as water stains on the concrete surface, climbing vines, artificial painting marks, expansion joints, and so on. The structural cracks cannot be accurately identified.

With the continuous improvement of Internet technology and the support of related hardware, artificial intelligence technology has developed rapidly, and it also provides a foundation for big data processing and application. At the same time, the popularity of smart phones, intelligent unmanned aircraft, and intelligent robots has also provided great convenience for big data collection. Based on this, this paper proposes to combine artificial intelligence with project engineering management; use artificial intelligence to collect concrete crack pictures; then use the artificial intelligence coordinate attention neural network to identify and locate cracks in pictures to achieve the purpose of crack detection and monitoring. Making full use of the advantages of artificial intelligence technology makes it possible to collect big data of concrete crack pictures and detect cracks. In addition, we deploy the trained CaNet model to the flask back-end framework and use the uni-app front-end writing framework to design the user interface, thus completing a landing project. At the same time, we used our smartphone to photograph cracks and crack-like noises from surrounding buildings and roads. A total of 100 cracked and uncracked images were taken with a resolution of 4000×1800 . The configuration of the mobile phone is shown in Table 2, and some pictures are shown in Figure 7.

We successively input the captured pictures into the CaNet model for testing and repeat the process to other networks for comparison. The recognition results are shown in Table 3. Table 4 shows the running smart phone configurations.

3.3. Effectiveness Experiment of Coordinate Attention. To test the effectiveness of coordinate attention in the CaNet and

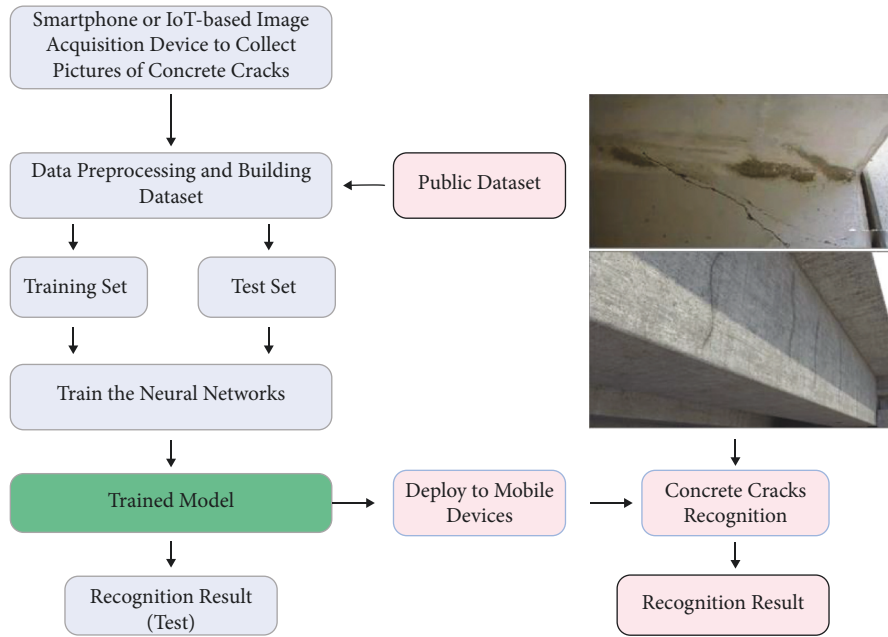


FIGURE 6: Training and recognition process of the concrete crack recognition model.

TABLE 2: Camera phone configuration.

Mobile phone configuration	Value
Processor	Qualcomm snapdragon 870 eight core
Running memory	8.00 GB
Operating system	Android 12
GPU	Adreno 650
Rear camera	71million pixels



FIGURE 7: Take some pictures.

determine its improvement on mAP and accuracy, we used SE, CBAM, and CA attention mechanisms as single variables in the experiment, and the invariant was the basic network ResNet50. The results of attention mechanism networks comparison are shown in Table 5.

It can be seen from Table 1 that the three attention mechanisms can improve the map and accuracy. Among

them, CA can improve the mAP and accuracy of the network.

3.4. Comparison with Other Networks. We compare CaNet with other networks in concrete crack identification and prove that its performance in concrete crack identification

TABLE 3: Actual test results.

Network model	Recall (%)	F1-score (%)	Precision (%)	mAP (%)	Accuracy (%)
CaNet	82	83	84	83	87
CNN	71	69	75	73	76
ResNet50	79	75	77	75	79

TABLE 4: Operating environment.

Mobile phone configuration	Value
Processor	Qualcomm snapdragon 870 eight core
Running memory	8.00 GB
Operating system	Android 12
GPU	Adreno 650

TABLE 5: Comparison of attention mechanism networks.

Network model	mAP (%)	Accuracy (%)
ResNet50	77.1	83.4
ResNet50-SE	79.7	85.3
ResNet50-CBAM	83.5	87.7
CaNet	85.8	89.6

TABLE 6: Results of comparative experiments.

Network model	Recall (%)	F1-score (%)	Precision (%)
DCNN [22]	76	73	77
ResNet50 [15]	79	76	79
SENet [23]	81	79	82
ResNet50-BAM [24]	84	82	83
ResNet50-CBAM [25]	84	80	85
SCHNet [14]	85	84	84
CaNet	86	85	87

is better than other models. In this comparative experiment, we have compared 7 kinds of deep learning networks. The results of network comparison are shown in Table 6.

4. Conclusion

In this paper, we construct CaNet to identify concrete cracks. In experiment 3.2, we have tested the recognition effects of CNN, ResNet50, and CaNet on concrete cracks in practical applications. The performance indexes of CaNet we proposed are better than the other two models. In experiment 3.3, we compared the effectiveness of the three attention mechanisms and concluded that coordinated attention improved the network the most in mAP and accuracy, by 8.7% and 6.2%, respectively. In experiment 3.4, we tested the performance of several groups of deep neural networks for concrete crack detection. The recall, F1-score, and precision of CaNet are 86%, 85%, and 87%, respectively, which are better than other networks. These experiments show that the CaNet model is effective for identifying concrete cracks, but further research is still needed. (1) At present, the focus of the research is to identify the concrete crack images of

highways and tunnels to complete the classification task. Later, it can be extended to the detection task of wall cracks and metal cracks. (2) The data in this network training are all from the public dataset. In the future, cameras, smartphones, and other shooting equipment should be used to collect concrete crack images in different scenes, expand the dataset, and improve the generalization of the model.

Data Availability

The dataset can be found here: <https://www.kaggle.com/datasets/arunrk7/surface-crack-detection?resource=download>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] B. Y. Kara and V. Verter, "Designing a road network for hazardous materials transportation," *Transportation Science*, vol. 38, no. 2, pp. 188–196, 2004.
- [2] J. A. Richards, "Inspection, maintenance and repair of tunnels: international lessons and practice," *Tunnelling and Underground Space Technology*, vol. 13, no. 4, pp. 369–375, 1998.
- [3] F. Ye, N. Qin, X. Liang, A. Ouyang, Z. Qin, and E. Su, "Analyses of the defects in highway tunnels in China," *Tunnelling and Underground Space Technology*, vol. 107, Article ID 103658, 2021.
- [4] J. Liu, R. An, Z. Jiang et al., "Effects of w/b ratio, fly ash, limestone calcined clay, seawater and sea-sand on workability, mechanical properties, drying shrinkage behavior and micro-structural characteristics of concrete," *Construction and Building Materials*, vol. 321, Article ID 126333, 2022.
- [5] H. J. Vogel, H. Hoffmann, A. Leopold, and K. Roth, "Studies of crack dynamics in clay soil: II. A physically based model for crack formation," *Geoderma*, vol. 125, no. 3-4, pp. 213–223, 2005.
- [6] W. Zhang, Z. Zhang, D. Qi, and Y. Liu, "Automatic crack detection and classification method for subway tunnel safety monitoring," *Sensors*, vol. 14, no. 10, pp. 19307–19328, 2014.
- [7] Q. Lu, M. Gunaratne, Z. Yu, L. Guo, S. Uddin, and M. Hoque, *Improving Safety in Pavement Field Testing*, Dept. of Transportation, Florida, 2017.
- [8] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 196–210, 2015.
- [9] L. Attard, C. J. Debono, G. Valentino, and M. Di Castro, "Tunnel inspection using photogrammetric techniques and image processing: a review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 144, pp. 180–188, 2018.

- [10] A. M. A. Talab, Z. Huang, F. Xi, and L. HaiMing, "Detection crack in image using Otsu method and multiple filtering in image processing techniques," *Optik*, vol. 127, no. 3, pp. 1030–1033, 2016.
- [11] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.
- [12] Y. Ren, J. Huang, Z. Hong et al., "Image-based concrete crack detection in tunnels using deep fully convolutional networks," *Construction and Building Materials*, vol. 234, Article ID 117367, 2020.
- [13] Y. J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.
- [14] Y. Pan, G. Zhang, and L. Zhang, "A spatial-channel hierarchical deep learning network for pixel-level automated crack detection," *Automation in Construction*, vol. 119, Article ID 103357, 2020.
- [15] J. Wang, X. He, S. Faming, G. Lu, H. Cong, and Q. Jiang, "A real-time bridge crack detection method based on an improved inception-resnet-v2 structure," *IEEE Access*, vol. 9, pp. 93209–93223, 2021.
- [16] F. Nayyeri and J. Zhou, "Multi-resolution ResNet for road and bridge crack detection," in *Proceedings of the 2021 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–8, Gold Coast, Australia, December 2021.
- [17] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3147–3155, Honolulu, HI, USA, July 2017.
- [18] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13713–13722, Nashville, TN, USA, June 2021.
- [19] M. H. Tekieh and B. Raahemi, "Importance of data mining in healthcare: a survey," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 1057–1062, France, August 2015.
- [20] M. S. Amin, Y. K. Chiam, and K. D. Varathan, "Identification of significant features and data mining techniques in predicting heart disease," *Telematics and Informatics*, vol. 36, pp. 82–93, 2019.
- [21] Kaggle, "Surface crack detection," <https://www.kaggle.com/datasets/arunrk7/surface-crack-detection?resource=download>.
- [22] B. B. Traore, B. Kamsu-Foguem, and F. Tangara, "Deep convolution neural network for image recognition," *Ecological Informatics*, vol. 48, pp. 257–268, 2018.
- [23] F. Cao, K. Yao, and J. Liang, "Deconvolutional neural network for image super-resolution," *Neural Networks*, vol. 132, pp. 394–404, 2020.
- [24] Z. Dong, X. Shen, H. Li, and X. Tian, "Photo quality assessment with dcnn that understands image well," *International Conference on Multimedia Modeling*, pp. 524–535, Springer, Cham, 2015.
- [25] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "Bam: bottleneck attention module," 2018, <https://arxiv.org/abs/1807.06514>.
- [26] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, Munich, Germany, September 2018.

Research Article

Integrated Prediction Framework for Clinical Scores of Cognitive Functions in ESRD Patients

Yutao Zhang,¹ Quan Sheng,¹ Xidong Fu,² Haifeng Shi,³ and Zhuqing Jiao ^{1,2}

¹School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213164, China

²School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China

³Department of Radiology, Changzhou Second People's Hospital Affiliated to Nanjing Medical University, Changzhou 213003, China

Correspondence should be addressed to Zhuqing Jiao; jzq@cczu.edu.cn

Received 30 May 2022; Revised 27 June 2022; Accepted 5 July 2022; Published 9 August 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yutao 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 clinical scores are applied to determine the stage of cognitive function in patients with end-stage renal disease (ESRD). However, accurate clinical scores are hard to come by. This paper proposed an integrated prediction framework with GPLWLSV to predict clinical scores of cognitive functions in ESRD patients. GPLWLSV incorporated three parts, graph theoretic algorithm (GTA) and principal component analysis (PCA), whale optimization algorithm with Levy flight (LWOA), and least squares support vector regression machine (LSSVRM). GTA was adopted to extract features from the brain functional networks in ESRD patients, while PCA was used to select features. LSSVRM was built to explore the relationship between the selected features and the clinical scores of ESRD patients. Whale optimization algorithm (WOA) was introduced to select better parameters of the kernel function in LSSVRM; it aims to improve the exploration competence of LSSVRM. Levy flight was used to optimize the ability to jump out of local optima in WOA and improve the convergence of coefficient vectors in WOA, which lead to an increase in the generalization ability and convergence speed of WOA. The results validated that the prediction accuracy of GPLWLSV was higher than that of several comparable frameworks, such as GPSV, GPLSV, and GPWLSV. In particular, the average of root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) between the predicted scores and the actual scores of ESRD patients was 2.40, 2.06, and 9.83%, respectively. The proposed framework not only can predict the clinical scores more accurately but also can capture imaging markers associated with decline of cognitive function. It helps to understand the potential relationship between structural changes in the brain and cognitive function of ESRD patients.

1. Introduction

In recent years, the mortality rate of patients with end-stage renal disease (ESRD) has been increasing year by year. ESRD is the second highest increase of any disease, which is one of the top risks to human health. Although ESRD patients represent 0.1% of the global population, World Public Health Organization spends 2%-3% of its total expenditure on the treatment of ESRD. ESRD not only places a severe economic burden on society and families but also places a significant mental burden on patients [1]. Cognitive impairment is common in ESRD patients, especially in the aspects of orientation, attention, and executive ability [2]. Bugnicourt et al. [3] found that the incidence of cognitive

impairment in patients undergoing hemodialysis was as high as 30% ~ 60%. Cognitive impairment may affect late treatment of ESRD patients, such as dietary adjustment and medication compliance. Cognitive impairment in ESRD patients is often ignored in clinical practice, and its pathophysiological mechanism has not been fully elucidated [4]. In consequence, it is of great significance to study the cognitive impairment mode of patients with ESRD and clarify the exact pathophysiological mechanism.

As neuroimaging technology develops by leaps and bounds, the neuropathological mechanism of ESRD can be learned from the perspective of central nervous systems. For example, Liang [5] et al. observed the brain neural activity of ESRD patients with the help of resting-state fMRI

technology. Compared with normal controls, the functional activity of ESRD patients significantly decreased in bilateral frontal parietal temporal lobe, suggesting that the abnormal connection of brain functional networks in ESRD patients could result in cognitive disorders. In virtue of Diffusion Tensor Imaging (DTI) technology, Chou et al. [6] pointed out that long-term hemodialysis would aggravate cerebral interstitial edema in ESRD patients and cause demyelination of pontine axons, indicating that hemodialysis may lead to extensive white matter damage in the brain. Chai et al. [7] found that the volume of gray matter in the left lobe and bilateral putamen of ESRD patients was significantly lower than that of healthy subjects through voxel-based morphometry. This suggests that changes in the volume of gray matter in the left lobe bring about cognitive impairment probably. These researches relying on neuroimaging technology have greatly deepened our understanding of the cognitive function of ESRD patients, but it is still impossible to determine the stage of cognitive function of ESRD patients.

Clinically, neurologists often judge the stage of cognitive function of ESRD patients according to the scores of the Montreal Cognitive Assessment Scale (MoCA), also known as clinical scores. Accurate prediction of clinical scores is beneficial to estimating the stage of cognitive functions of ESRD patients. Jiang et al. [8] calculated the correlation coefficient between topological attribute parameters of brain functional networks and clinical scores of cognitive functions in ESRD patients by Pearson correlation analysis. They mainly focused on the imaging markers affecting the cognitive function of ESRD patients and could not judge the current status of cognitive function well. Lu et al. [9] proposed a prediction method of clinical scores based on the brain functional networks. By virtue of simple definition and small computation, clustering coefficients in the brain functional networks were extracted as features to predict clinical scores, which were referred to judge the current stage of cognitive function. However, this method is subjective in the process of extracting features; at the same time, the influence of other features on the cognitive function of patients is ignored. Yang et al. [10] predicted clinical scores by support vector regression machine (SVRM). Regrettably, SVRM has high volatility and low prediction accuracy in the process of predicting clinical scores.

On the above considerations, we propose an integrated framework to predict clinical scores of ESRD patients. The main work is as follows: firstly, fMRI data were preprocessed to construct the brain functional networks, and graph theoretic algorithm (GTA) was adopted to extract features. Secondly, principal component analysis (PCA) was used to select features. Then, the least squares support vector regression machine (LSSVRM) was built to explore the relationship between the selected features and the clinical scores of ESRD patients. Meanwhile, the whale optimization algorithm (WOA) was introduced to select better parameters of kernel function in LSSVRM so as to enhance the exploration competence of LSSVRM. Finally, Levy flight

replaced the traditional selection of WOA and in the meantime optimized the ability to jump out of local optimum in WOA. The framework called GPLWLSV was constructed to predict the clinical scores of ESRD patients and then determine their current stage of cognitive function.

2. Data and Methods

2.1. Research Framework. Figure 1 shows our prediction framework, which mainly includes the following steps: (1) The original resting-state fMRI images were preprocessed by Data Processing Assistant for Resting-State fMRI (DPARSF). (2) Time series from preprocessed images were abstracted to construct the brain functional networks. (3) The area under the curve (AUC) of topological attribute parameters in the brain functional networks was extracted as features through GTA. (4) PCA was used to filter redundant features and to retain important features. (5) LSSVRM was built to retain features. (6) WOA was improved by Levy flight to optimize the selection strategy of kernel function parameters in LSSVRM. (7) GPLWLSV was constructed to predict clinical scores of cognitive functions in ESRD patients.

2.2. Experimental Data and Pretreatment. A total of 50 patients with ESRD were admitted to Changzhou Second People's Hospital Affiliated to Nanjing Medical University from May 2021 to March 2022, including 27 male and 23 female individuals aged 49.12 ± 8.23 years. Synchronously, a total of 40 normal controls were also admitted to the same hospital, including 22 male and 18 female individuals aged 47.26 ± 7.01 years. There were no significant differences ($P > 0.05$) in the gender ratio, age, and education level between them. One hour before fMRI examination, the stage of cognitive function of all subjects was assessed by trained neurologists who did not know the data of subjects via clinical scores of cognitive functions. Table 1 gives the demographic information of these two groups of subjects.

All subjects were scanned with GE Discovery MR 750W 3.0T superconducting MR scanner with 32-channel head and neck joint coil. The head of subject was fixed with a cushion to reduce the artifacts produced by head movement. Gradient echo plane echo imaging (GRE-EPI) sequence was used to collect fMRI images. During the collection process, subjects were required to keep quiet and awake and try not to think. Machine scanning parameters are as follows: repetition time (TR) = 2000 ms, echo time (TE) = 40 ms, field of view (FOV) = 24 cm, flip angle (FA) = 90° , matrix size = 64×64 , and layer thickness = 6 mm.

After achieving the fMRI images of all subjects, DPARSF (available at <http://rfmri.org/dpabi>) participated in pre-processing. The specific steps are as follows: (a) Image format was transformed: the correct format can only be opened by DPARSF. (b) The first 10 time points were deleted: errors may occur when the instrument and subjects were not in a stable state. (c) Time points were corrected:

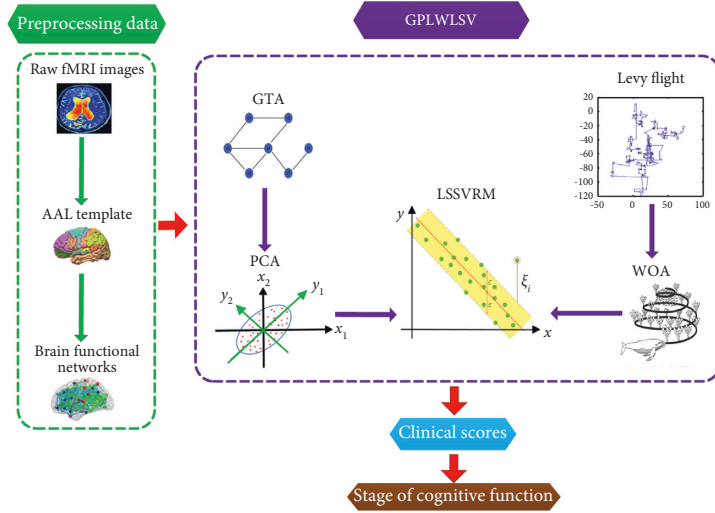


FIGURE 1: A flowchart of prediction framework.

TABLE 1: Demographic information of subjects.

	Gender (male/female)	Age (years, $\bar{x} \pm s$)	Education years (years, $\bar{x} \pm s$)	Clinical scores (points, $\bar{x} \pm s$)
ESRD patients ($n = 50$)	27/23	49.12 \pm 8.23	11.13 \pm 2.05	21.47 \pm 2.75
Normal controls ($n = 40$)	22/18	47.26 \pm 7.01	11.24 \pm 2.13	27.38 \pm 1.35
t/χ^2	0.009	1.133	0.387	-13.412
P	>0.05	0.260	0.778	0.000

data of time points at different levels were corrected to the same time point to maintain consistency. (d) Brain positions were corrected: the brain must remain in the same position throughout the examination. (e) Space was normalized: put the brain of subject into a standard space for the convenience of follow-up statistics and reports. (f) Spatial structure differences were reduced through smooth: better image results were beneficial to the validity of statistical tests. (g) Linear drift was removed: changes in time and temperature caused errors in the inspection instrument. (h) Filtering was performed: the frequency range is 0.01~0.08 Hz. The resting fMRI signal after low-frequency filtering has important physiological significance and may reflect spontaneous neural activity. (i) Final time series was obtained: the mean BOLD time series of head motion parameters, white matter, and cerebrospinal fluid were regression and removed.

2.3. Principle of GP

2.3.1. Features Were Extracted by GTA. The construction of brain functional networks is as follows. Firstly, the brain of each subject was divided into 90 brain regions on the basis of Automated Anatomical Labeling (AAL). Secondly, the Pearson correlation coefficient was calculated between the time series of two brain regions; \mathbf{K} -matrix was born, which is a 90×90 symmetric matrix with all 1s on the diagonal. Finally, the Fisher-Z transformation is performed on the elements in the \mathbf{K} -matrix divided by the diagonal, which is converted to the Z-value close to the normal distribution to generate the \mathbf{Z} -matrix:

$$\mathbf{K} = \begin{pmatrix} 1 & p_{(1,2)} & \cdots & p_{(1,90)} \\ p_{(2,1)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p_{(90,1)} & \cdots & \cdots & 1 \end{pmatrix}, \quad (1)$$

$$\mathbf{Z} = \begin{pmatrix} 1 & z_{(1,2)} & \cdots & z_{(1,90)} \\ z_{(2,1)} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ z_{(90,1)} & \cdots & \cdots & 1 \end{pmatrix}, \quad (2)$$

$$z_{ij} = \frac{1}{2} [\ln(1 + p_{ij}) - \ln(1 - p_{ij})] (i \neq j), \quad (3)$$

where p_{ij} represents the Pearson correlation coefficient between the time series of the i -th brain region and the j -th brain region.

Taking matrix sparsity as the threshold, the preprocessing for \mathbf{Z} -matrix was binary. The matrix sparsity was set to 0.1~0.4 with a span of 0.01. Within the threshold range of the matrix sparsity, the AUC of topological attribute parameters was calculated by GTA as features, which reflected the overall change of brain functional networks and reduced the influence of individual differences. These features include the AUC of global efficiency (Eglobal), local efficiency (Elocal), clustering coefficient (Cp), characteristic path length (Lp), standardized clustering coefficient (γ), standardized characteristic path length (λ), and small-world properties (σ).

2.3.2. Features Were Selected by PCA. In the process of machine learning, a large number of features need to be extracted and converted into various data that can be processed by computers [11]. Increasing the number of features can improve the effect of machine learning. The trouble is that as the number of features increases, the dimension of feature vectors will also add. This not only makes machine learning more difficult but can also lead to overfitting, affecting the final accuracy. In view of this situation, it is necessary to filter out some unimportant features or combine some related features [12]. Features were selected by PCA that keep the information contained in the original data as much as possible while reducing the dimension. The specific steps are as follows.

The first step is to standardize the matrix. The AUC of Eglobal, Elocal, Cp, Lp, γ , λ , and σ of ESRD patients was marked as $\alpha_1 \sim \alpha_7$. Constructing a matrix called A, $A = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7]$, and $\alpha_j = [x_{1j}, x_{2j}, x_{3j}, \dots, x_{nj}]^T$.

B-matrix is the normalization of A-matrix, $B = [\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7]$, and $\beta_j = [y_{1j}, y_{2j}, y_{3j}, \dots, y_{nj}]^T$ [13]:

$$y_{ij} = \frac{x_{ij} - \bar{x}}{\sqrt{\left(\frac{(x_{ij} - \bar{x})^2}{n-1}\right)}}, \quad (4)$$

where \bar{x} is the mean value of samples in α_j and n is the number of samples.

The second step is to construct the covariance matrix of B-matrix. Computing the covariance between β_i and β_j in $\beta_1 \sim \beta_7$,

$$r_{ij} = \text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}, \quad (5)$$

where X_i is the value of the i -th sample; \bar{X} represents the mean value of samples; n is the number of samples.

Constructing the covariance matrix called R,

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & \ddots & & r_{2m} \\ \vdots & & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix}, \quad (6)$$

where m represents the number of features and its value is 7.

The third step is to extract the original features. Computing the eigenvalues and eigenvectors of the R-matrix, the principal component is the eigenvector of the R-matrix, and the magnitude of the corresponding eigenvalue indicates the importance of the eigenvector. Calculating the weight of the features, $\lambda_1 \sim \lambda_7$ are the eigenvalues of the R-matrix, λ_j represents the dominance of α_j in the A-matrix, and the weight of α_j in the whole feature set called $\alpha_1 \sim \alpha_7$ can be expressed as [14]

$$w_i = \frac{\lambda_i}{\sum_{j=1}^7 \lambda_j}. \quad (7)$$

The weight of extracted features is higher than 0.6 set in this paper.

Finally, the extracted features are transformed to new features. Finding the eigenvector v_j in the R-matrix corresponding to α_j , the new feature c_j is transformed from v_j [15]:

$$c_j = B \cdot v_j. \quad (8)$$

2.4. Principle of LSSVRM. The principle of SVRM is to map data vectors from low-dimensional space to high-dimensional space. SVRM constructs decision functions with the aid of the principle of risk minimization. SVRM is suited to solve nonlinear problems of small samples and high dimension of feature vectors. LSSVRM is simplified on the basis of SVRM. There are two main points: (1) changing constraints in SVRM to improve computing efficiency; (2) selecting different decision functions to reduce the operation time. LSSVRM maps nonlinear vector $\Phi(x)$ to high-dimensional space and transforms it into a linear regression problem, as shown in formula (9) [16]:

$$y = \omega^T \Phi(x) + b, \quad (9)$$

where ω is the n -dimensional weight vector; b is the deviation.

Following the principle of risk minimization, linear regression can be transformed into an optimal problem [17]:

$$\min J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \frac{1}{2} J \sum_{i=1}^m \xi_i^2, \quad (10)$$

$$y_i = \omega^T \Phi(x_i) + b + \xi_i, \quad (11)$$

where J represents the normalization function; ξ_i represents the error; i represents the i -th dimension of the space vector; ω represents the weight vector. Since ω belongs to the high-dimensional space, it cannot be solved directly, so the kernel function is introduced [18]:

$$y = \sum_{i=1}^n a_i K(x, x_i) + b. \quad (12)$$

In order to ensure the efficiency of operation, RBF is chosen as the kernel function, and its function is expressed as follows [19]:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right). \quad (13)$$

It can be seen from the above process that J and σ^2 have the greatest influence on the LSSVRM. In searching for optimal J and σ^2 , WOA is replaced by LWOA to seek the optimal solution.

2.5. Principle of LWOA. WOA is a new optimization algorithm that simulates the predatory behavior of whales. Whales locate and surround prey through their scent. We can define a certain number of virtual whales as search agents, assuming that the location of the scent of prey is at or

near the current optimal location. We look for the optimal solution as the next location of the whale by comparing the feasible solutions of various search agents. Other search agents update their location to complete the strategy of finding the optimal solution.

Whales constantly adjust their position according to the position of prey during hunting. To describe this strategy of hunting, the following mathematical model is presented [20]:

$$\begin{aligned} \mathbf{M}_1 &= |\mathbf{C}\mathbf{X}^*(t) - \mathbf{X}(t)|, \\ \mathbf{X}(t+1) &= \mathbf{X}^*(t) - H\mathbf{M}_1, \end{aligned} \quad (14)$$

where t is the number of current iterations; $\mathbf{X}(t)$ is the coordinate vector of the current whale; $\mathbf{X}(t+1)$ is the target coordinate vector after the next iteration; $\mathbf{X}^*(t)$ is the coordinate vector of the best solution at present. If there is a better feasible solution, $\mathbf{X}^*(t)$ should be updated immediately. C and H are coefficients, which are acquired by formulas (22) and (23) [21]:

$$\begin{aligned} H &= 2hr_1 - h, \\ C &= 2r_2, \end{aligned} \quad (15)$$

where r_1 and r_2 are random numbers between 0 and 1; h is computed from formula (24) [21]:

$$h = 2 - \frac{2t}{T_{\max}}. \quad (16)$$

For purpose of finding a better target position to approach the prey, the whale will randomly use any whale coordinate vector to replace the whale coordinate vector of the next iteration so as to achieve the purpose of deviating from the prey. This avoids falling into local optimality. The following mathematical model is shown [22]:

$$\begin{aligned} \mathbf{X}(t+1) &= \begin{cases} \mathbf{X}^*(t) - H\mathbf{M}_1, & |H| < 1, p < 0.5, \\ \mathbf{X}_{\text{rand}}(t) - H\mathbf{M}_2, & |H| \geq 1, p < 0.5, \\ \mathbf{X}^*(t) + \mathbf{M}_3 e^{bl} \cos(2\pi l), & p \geq 0.5, \end{cases} \\ \mathbf{M}_1 &= |\mathbf{C}\mathbf{X}^*(t) - \mathbf{X}(t)|, \\ \mathbf{M}_2 &= |\mathbf{C}\mathbf{X}_{\text{rand}} - \mathbf{X}(t)|, \\ \mathbf{M}_3 &= |\mathbf{X}^*(t) - \mathbf{X}(t)|, \end{aligned} \quad (17)$$

where t is the current iteration times; $\mathbf{X}^*(t)$ is the optimal position vector so far; \mathbf{X}_{rand} is the random position vector of whale; $\mathbf{X}(t)$ is the current position vector of whale; b is the constant, and default is 1; the role of b is to control the hunting path shape; l is acquired by the following formulas:

$$\begin{aligned} l &= (h_2 - 1)r_3 + 1, \\ h_2 &= -1 - \frac{t}{T_{\max}}, \end{aligned} \quad (18)$$

where r_3 is a random number between 0 and 1; t represents the current iteration number; T_{\max} represents the maximum iteration number. When $t \geq 0.5T_{\max}$, H is always less than 1.

At this point, the whale enters the attack mode and no longer deviates from the prey through search agents.

Due to the few adjustment parameters of WOA, it has a fast rate of convergence and a certain ability to jump out of the local optimum. It is worth noting that WOA can be optimized further. The reasons are as follows: Firstly, WOA searches by means of random system. Excessive reliance on random system limits the convergence speed of WOA. Secondly, WOA is subject to coefficient vector. WOA will lose the ability to jump out of the local optimum when the number of iterations reaches half of the maximum number of iterations set earlier. Consequently, WOA is accompanied by a certain risk of falling into the local optimum, leading to inaccurate results of prediction [23].

The defects of WOA can be solved by an improved WOA with Levy flight named LWOA. Levy flight is a kind of random search that relies on Levy distribution, which has been applied many times in the optimization field in recent years. Levy flight is able to improve cuckoo and particle swarm optimization algorithms [23–25] and so on. LWOA owns a faster convergence speed and higher convergence accuracy; LWOA has a better ability to jump out of the local optimum. The specific steps of improvement are as follows.

WOA is improved by Levy flight, and formula (22) is replaced by the following formula [16, 26]:

$$H = 2h\text{Levy}(\lambda) - h, \quad (19)$$

where $\text{Levy}(\lambda)$ means that it obeys the Levy distribution with parameter λ [16, 27]:

$$\text{Levy} \sim \mu = t^{-\lambda}. \quad (20)$$

Due to the complexity of Levy flight, the Mantegna algorithm is adopted to simulate it, and its mathematical expression is as follows [16, 28]:

$$s = \frac{\mu}{|\nu|^{(1/\beta)}}, \quad (21)$$

where μ and ν obey the normal distribution with parameters σ_μ and σ_ν [16, 29]:

$$\begin{aligned} \mu &\sim N(0, \sigma_\mu^2), \\ \nu &\sim N(0, \sigma_\nu^2), \end{aligned} \quad (22)$$

$$\sigma_\mu = \begin{cases} \frac{\Gamma(1 + \beta)\sin(\pi\beta/2)}{\Gamma(1 + \beta/2)\beta 2^{(\beta-1/2)}}, \end{cases} \quad (23)$$

$$\sigma_\nu = 1. \quad (24)$$

For higher operation efficiency of the algorithm, β is a constant 1.5, and σ_μ is a constant 0.7.

The coefficient vector \mathbf{H} in WOA converges linearly with certain limitations. WOA should be promoted to jump out of local optimum; formula (24) changed by the following formula [16]:

$$h = 2e^{0.15(-\log(10t/T_{\max}))^4}, \quad (25)$$

where t represents the current iteration number and T_{\max} represents the maximum iteration number.

In the early stage of iteration, the value of h will decrease slowly as the increase of iteration times, which is conducive to the global search out of local optimum. At the end of iteration, the value of h will decrease exponentially to improve the ability of rapid local search.

3. Results

3.1. Experimental Settings. Table 2 shows the AUC of topology attribute parameters of the brain functional networks of ESRD patients and normal controls calculated by GTA. Within the whole matrix sparsity threshold range, the AUC of γ , σ , and Elocal in ESRD patients was significantly lower than that in normal controls, with statistical significance ($P < 0.05$). Nevertheless, there were no significant differences ($P > 0.05$) in the AUC of λ , Cp, Lp, and Eglobal.

Table 3 shows the corresponding weight of each feature in the feature set. The proportion of AUC of Elocal was the highest, up to 65.31%. The proportion of AUC of Eglobal was next to Elocal, accounting for 30.89%. The proportion of AUC of γ , λ , σ , Cp, and Lp was less than 5%. The proportion of γ was the lowest, only 0.03%.

The experimental results can be seen from Tables 2 and 3. The AUC of Elocal in ESRD patients was significantly lower than that in the normal controls, with statistically significant differences ($P < 0.05$). Meanwhile, the AUC of Elocal accounted for the highest proportion in the feature set. Obviously, the AUC of Elocal was selected as the feature to construct GPSV [30], GPLSV [31], GPWLSV [32], and GPLWLSV for predicting the clinical scores of cognitive functions in ESRD patients [33].

The prediction accuracy of various frameworks was evaluated by a tenfold cross-validation method. Firstly, the AUC of Elocal of 50 ESRD patients and the corresponding clinical scores were collected as data set called D . Secondly, D was divided into 10 mutually exclusive subsets of the same size, $D = D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5 \cup D_6 \cup D_7 \cup D_8 \cup D_9 \cup D_{10}$, and $D_i \cap D_j = \emptyset$. Each subset D_i was separated from D through stratified sampling, which ensures consistency of data distribution. Each subset D_i contains five samples. Taking the union of 9 subsets as the training set and the remaining subset as the test set, 10 groups of training sets and test sets are formed. Different frameworks were trained and tested for 10 times, and the average of 10 groups of test results was calculated.

3.2. Experimental Results. The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were selected as the testing standards of prediction accuracies. Table 4 shows the prediction accuracies of some representative frameworks. All the comparable predictive frameworks were evaluated by tenfold cross-validation method. The test results show that the prediction accuracy of GPLWLSV is higher than that of

GPSV, GPLSV, and GPWLSV. The average of RMSE of GPLWLSV was 2.40, 1.04, 0.93, and 0.46 points lower than that of GPSV, GPLSV, and GPWLSV, respectively. The average of MAE of GPLWLSV was 2.06, which was 0.82, 0.74, and 0.42 points lower than that of GPSV, GPLSV, and GPWLSV, respectively. MAPE can reflect the relative errors of frameworks better than MAE. The average of MAPE of GPLWLSV was 9.83%, lower than 10%, which was 4.10%, 3.75%, and 1.93% lower than that of GPSV, GPLSV, and GPWLSV, respectively. The bar chart in Figure 2 shows that the prediction accuracy of GPLWLSV is better than that of GPSV, GPLSV, and GPWLSV intuitively.

Figure 3 shows the comparison between the predicted scores of various frameworks and the actual scores. The solid blue line represents the actual scores, and the solid purple line represents the predicted scores. As can be seen from the figure, different from GPSV and GPLSV, GPWLSV and GPLWLSV can fit most of the scores of tests set well, and the predicted scores are closer to the actual scores in the case of large fluctuation. The strong fluctuation of actual scores will lead to a large error between the actual scores and the predicted scores of GPSV and GPLSV, while the predicted scores of GPWLSV and GPLWLSV are stable relatively. This is due to the powerful optimization ability of WOA.

Specifically, WOA optimizes the penalty factor and kernel parameters of LSSVRM, and we only need to adjust two parameters to improve the generalization ability and prediction accuracy of LSSVRM. The prediction accuracy of GPLWLSV is higher than that of GPWLSV for two reasons. Firstly, WOA adjusts parameters through a random system, and excessive reliance on a random system limits GPWLSV. Secondly, WOA is subject to coefficient vector. When the number of iterations reaches half of the maximum set earlier, WOA will lose its ability to jump out of the local optimum. At this time, WOA may fall into local optimum to some extent, which will lead to inaccurate prediction results. LWOA replaces the random adjustment of WOA with Levy flight; LWOA improves the convergence mode of coefficient vector of WOA; LWOA optimizes the ability of WOA to jump out of local optimum and thus improves the generalization ability and convergence speed of WOA. As a result, the prediction accuracy of GPLWLSV is higher than GPWLSV.

3.3. Discriminative Brain Regions. Node efficiency is one of the measures of node centrality. The higher the node efficiency is, the stronger the capacity of information transmission of the node is and the more important its position in the network is. The higher the efficiency of a node is, the stronger the information transmission capacity of the node is and the more important the node is in the network [34]. For the sake of finding out the key brain regions which affect the cognitive function of patients with ESRD [35], the multiple regression method was taken to analyze the relationship between the node efficiency and clinical scores of ESRD patients [36]. The results showed that the node efficiency of the left amygdala (AMYG.L) was negatively correlated with the clinical scores ($r = -0.562$, $P = 0.014$)

TABLE 2: Comparison of AUC of global topological parameters between two groups (mean \pm SD).

Parameter	ESRD patients ($n = 50$)	Normal controls ($n = 40$)	t	P
γ	0.646 ± 0.071	0.669 ± 0.056	-1.714	0.004
λ	0.326 ± 0.123	0.329 ± 0.024	0.710	0.480
σ	0.589 ± 0.065	0.607 ± 0.065	-1.384	0.008
Cp	0.175 ± 0.014	0.176 ± 0.016	-0.296	0.768
Lp	0.543 ± 0.027	0.553 ± 0.069	-0.966	0.337
Eglobal	0.171 ± 0.006	0.170 ± 0.011	-0.477	0.635
Elocal	0.230 ± 0.007	0.241 ± 0.006	-0.253	0.001

TABLE 3: Weight of AUC of global topological parameters of ESRD patients in feature set.

Parameter	γ (%)	λ (%)	σ (%)	Cp (%)	Lp (%)	Eglobal (%)	Elocal (%)
$w_i \times 100\%$	0.03	0.07	0.17	0.28	3.25	30.89	65.31

significantly, as shown in Figure 4(a), while the node efficiency of the right parahippocampal gyrus (PHG.R) was positively correlated with the clinical scores ($r=0.551$, $P = 0.035$) significantly, as shown in Figure 4(b). In addition, no correlation was found between node efficiency and clinical scores in other brain regions ($P > 0.05$). The amygdala is mainly involved in emotional processing [37], and previous studies have also shown that depression of varying degrees is common in patients with ESRD [38].

Given the above, it can be speculated that the amygdala node efficiency may be related to the depressed mood of ESRD patients, and the emotional instability and intense reaction of patients may lead to the increase of amygdala node efficiency. Qin et al. [39] found that patients with type 2 diabetes had multiple brain regions with increased node efficiency. This phenomenon is explained as the compensatory mechanism of the network. The right parahippocampal gyrus is associated with learning and memory functions [40], and its reduction of node efficiency may be connected with cognitive impairment in ESRD patients.

To find out the relationship between each brain region and the cognitive function of ESRD patients, the node efficiency of 90 brain regions was calculated and contrasted d between the patients and the normal controls. The brain regions with significant differences were found; the result is shown in Figure 5. Blue nodes represent the brain regions with reduced node efficiency in ESRD patients compared to the normal controls. These included left insula (INS.L), right insula (INS.R), left median cingulate and paracingulate gyrus (DCG.L), right median and paracingulate gyrus (DCG.R), left hippocampus (HIP.L), right hippocampus (HIP.R), left parahippocampal gyrus (PHG.L), right parahippocampal gyrus (PHG.R), right transverse temporal gyrus (HES.R), left superior temporal gyrus (STG.L), left temporal pole: superior temporal gyrus (TPOsup.L), right temporal pole: superior temporal gyrus (TPOsup.R), left temporal pole: middle temporal gyrus (TPOmid.L), and right temporal pole: middle temporal gyrus (TPOmid.R). The red nodes represent the brain regions with increased node efficiency in ESRD patients compared to the normal controls. These include left amygdala (AMYG.L), right amygdala (AMYR.R), right calcarine fissure and surrounding cortex (CAL.R), left cuneus (CUN.L), right cuneus

(CUN.R), left superior occipital gyrus (SOG.L), and left middle occipital gyrus (MOG.L).

The brain is a highly optimized complex system capable of integrating information to deal with changes in cognitive needs. According to research findings, the brain regions with reduced node efficiency in ESRD patients are mainly located in the paralimbic network. The hippocampus (HIP) and parahippocampal gyrus (PHG) are related to learning and memory functions [40]; meanwhile, the median and paracingulate gyrus (DCG) are involved in cognitive control function [41]. Chou et al. [42] applied DTI technology to view the changes in brain functional networks in ESRD patients, and the results also found that the node efficiency in the paralimbic network was significantly decreased compared with the normal controls. Notably, ESRD patients had higher node efficiency in brain regions associated with visual networks, which are responsible for processing visual information. These regions are about attention, visual memory, and other neurocognitive functions [43]. Accordingly, it can be speculated that the increased node efficiency in these brain regions may also be a compensatory mechanism [39]. In summary, node efficiency can effectively distinguish ESRD patients from the normal controls, thus achieving accurate classification.

4. Discussion

The purpose of this study was to look for imaging markers related to the decline of cognitive function in ESRD patients, which could be modeled to predict the clinical scores of the patients, and the clinical scores could determine the current stage of cognitive function. As a consequence, the GPLWLSV is constructed. The experimental results show that the GPLWLSV achieves the optimal effect of prediction than comparable frameworks, indicating its effectiveness. The GPLWLSV has three main advantages.

First of all, the imaging markers are accurate. By right of the principle of GP, GPLWLSV found the imaging markers related to the decline of cognitive function in ESRD patients. Compared with the normal controls, ESRD patients showed a significant decrease in the AUC of γ , σ , and Elocal at multiple sparsity. This result is consistent with a previous study based on resting-state fMRI networks [44]. γ and

TABLE 4: Prediction accuracies of comparable frameworks.

Prediction framework	The test group	RMSE	MAE	MAPE
GPSV	1	3.5064	2.3835	0.1364
GPSV	2	3.482	3.0951	0.1454
GPSV	3	3.2648	2.8499	0.1267
GPSV	4	3.6154	3.1792	0.1542
GPSV	5	2.5508	2.2965	0.1109
GPSV	6	2.2724	1.7919	0.079
GPSV	7	4.115	3.6825	0.1788
GPSV	8	2.8204	2.659	0.1277
GPSV	9	5.7247	4.0523	0.1999
GPSV	10	3.0707	2.7852	0.1334
Average		3.4423	2.8775	0.1392
GPLSV	1	5.2634	4.197	0.2149
GPLSV	2	4.6628	3.1899	0.1305
GPLSV	3	3.4819	2.8121	0.1491
GPLSV	4	3.1611	2.7837	0.1385
GPLSV	5	2.1046	1.8884	0.0898
GPLSV	6	2.162	2.1246	0.0994
GPLSV	7	3.3545	3.0199	0.1524
GPLSV	8	3.9891	3.725	0.18
GPLSV	9	2.2805	1.7733	0.0812
GPLSV	10	2.8879	2.4838	0.1214
Average		3.3348	2.7998	0.1357
GPWLSV	1	4.2408	3.4782	0.1695
GPWLSV	2	3.3293	2.8178	0.124
GPWLSV	3	3.0657	2.4178	0.1188
GPWLSV	4	3.3027	2.9839	0.1466
GPWLSV	5	1.3125	1.1388	0.0529
GPWLSV	6	1.9272	1.4114	0.0618
GPWLSV	7	2.9412	2.7196	0.1311
GPWLSV	8	2.7899	2.6865	0.1284
GPWLSV	9	2.6758	2.3282	0.1092
GPWLSV	10	3.0313	2.8014	0.1335
Average		2.8616	2.4784	0.1176
GPLWLSV	1	3.9132	3.1026	0.1561
GPLWLSV	2	3.2506	2.8334	0.1258
GPLWLSV	3	2.7307	2.4334	0.1075
GPLWLSV	4	2.9976	2.3738	0.1216
GPLWLSV	5	1.2192	1.1111	0.053
GPLWLSV	6	0.718	0.6497	0.0313
GPLWLSV	7	2.5167	2.1156	0.1036
GPLWLSV	8	2.0709	1.9636	0.0914
GPLWLSV	9	2.1757	1.9467	0.0916
GPLWLSV	10	2.4163	2.0302	0.1007
Average		2.4009	2.056	0.0983

Elocal mainly affect specific information processing or fault tolerance rate of the network [45]. The decrease of γ and Elocal indicates that information transmission efficiency of different brain regions is reduced in ESRD patients, which impairs the ability to manage the brain potentially [46]. This provides a new perspective and potential imaging markers for understanding the underlying pathophysiological mechanisms of cognitive impairment in ESRD patients.

Secondly, the operation speed is faster. GPLSV has a shorter operation time than GPSV, although the prediction accuracy is similar. This is because LSSVRM is an improvement on SVRM. It changes inequality constraints into equality constraints in SVRM and transforms solving quadratic programming problems into solving linear equations,

speeding up the operation speed greatly. In a word, the LSSVRM enables GPLWLSV to run more efficiently.

Finally, the prediction accuracy is higher. The prediction accuracy of GPWLSV and GPLWLSV is higher than that of GPSV and GPLSV significantly. This is because WOA possesses powerful optimization ability; it optimized the strategy of selecting two parameters in the kernel function of SVRM and LSSVRM, thus improving the prediction accuracy of GPWLSV and GPLWLSV. However, WOA still has deficiencies. Firstly, WOA adopts a random system to adjust parameters and relies on random system excessively. Secondly, WOA is limited by the coefficient vector. When the number of iterations reaches half of the maximum iterations set earlier, WOA will lose its ability to jump out of local

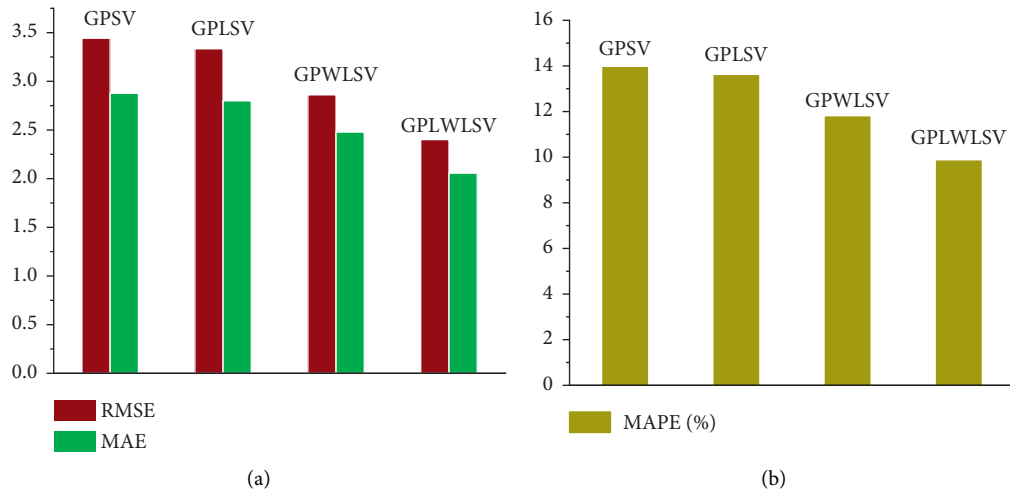


FIGURE 2: Average of prediction accuracies of various frameworks. (a) The RMSE and MAE of various frameworks. (b) The MAPE of various frameworks.

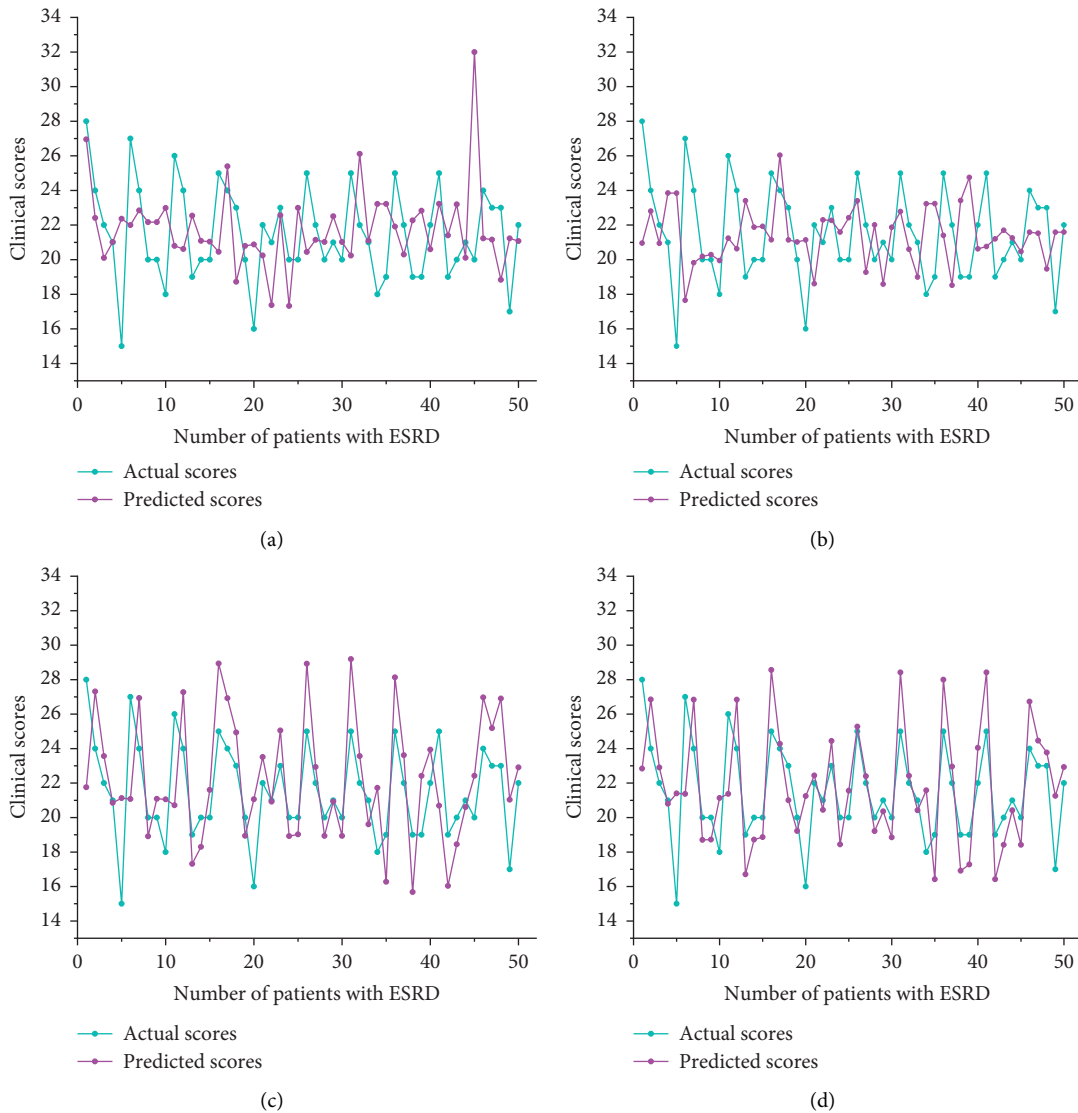


FIGURE 3: Actual scores and predicted scores of various frameworks. (a) GPSV, (b) GPLSV, (c) GPWLSV, and (d) GPLWLSV.

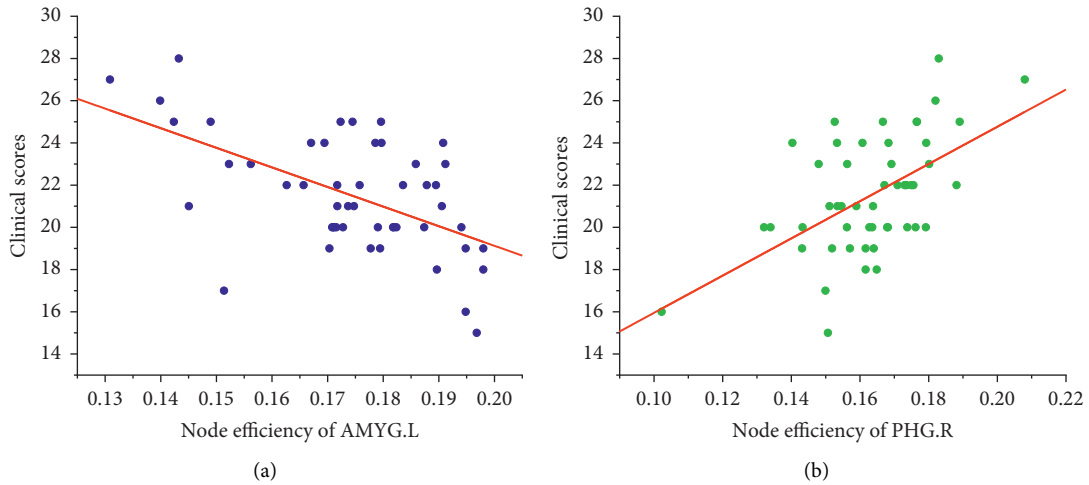


FIGURE 4: Relationship between node efficiency and clinical scores.

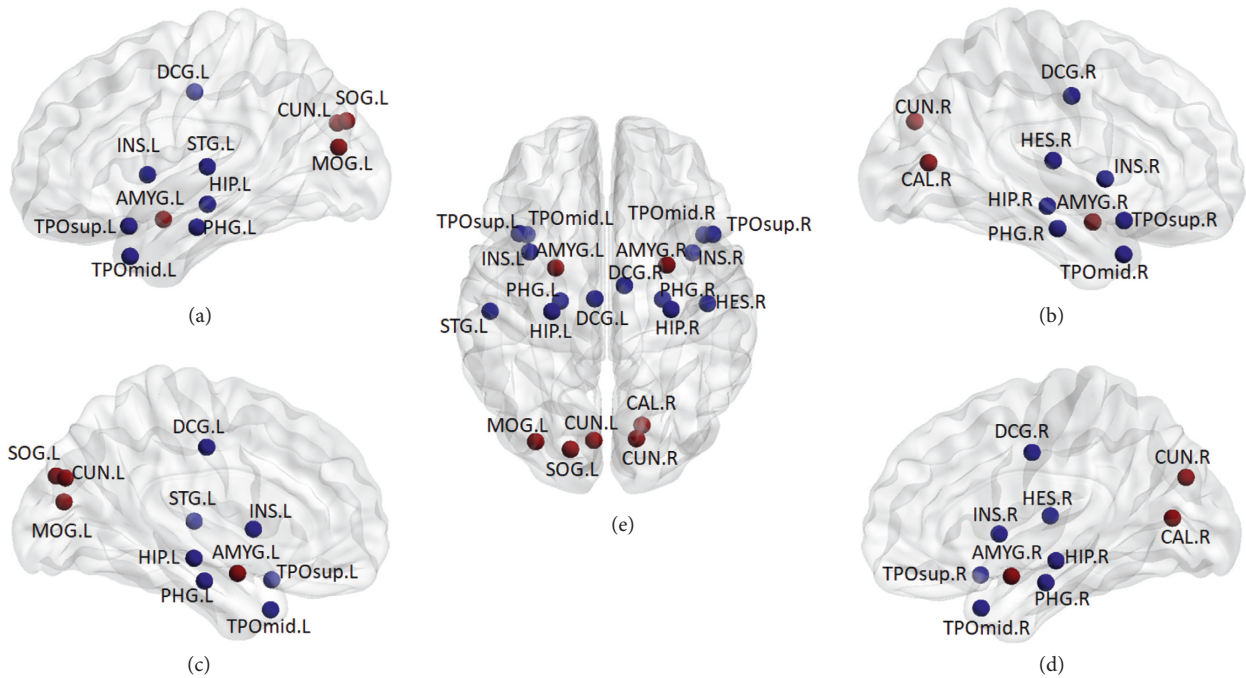


FIGURE 5: Brain regions with statistically significant differences in node efficiency between the two groups. (a) Left lateral view, (b) right lateral view, (c) left lateral view, (d) right lateral view, and (e) dorsal view of the whole brain.

optimum. In brief, WOA may fall into local optimum to some extent, leading to low prediction accuracy. GPLWLSV chose Levy flight instead of WOA to adjust parameters. LWOA improved the convergence mode of coefficient vector of WOA and hoisted the capacity of WOA to jump out of local optimum, and the convergence speed of WOA was improved. To sum up, the prediction accuracy of GPLWLSV was higher than comparable frameworks.

The prediction framework still has room for further improvement. The main reason is that the size of the data set in this paper is small, which affects the training effectiveness of GPLWLSV to a certain extent. Researches show that DTI, DWI, DKI, and other biomarkers of different modes are

complementary to each other to some extent. Compared with single-mode analysis, multimode analysis can often achieve better results of prediction by integrating complementary information of different modes [47]. In future studies, more clinical data of different modes will be collected, leading to a more in-depth and comprehensive study on a larger data set.

5. Conclusion

In this paper, an integrated prediction framework with GPLWLSV was constructed to predict the clinical scores of cognitive functions in ESRD patients. According to the

principle of GP, the framework found the imaging markers related to the decline of cognitive function in ESRD patients; meanwhile, GP determined that the AUC of Elocal accounted for the highest proportion of all the features. Elocal had a great influence on the cognitive function of ESRD patients. The inequality constraints in SVRM are simplified, and the operation speed is improved. LWOA was introduced to optimize the parameter selection strategy in LSSVRM, and the prediction accuracy was improved. In clinical diagnosis, it is often necessary to analyze the clinical scores of a large number of patients with ESRD in order to judge the stage of cognitive function. In general, our framework can obtain imaging markers related to the decline of cognitive function accurately and give consideration to work efficiency and accuracy simultaneously.

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 to report regarding the present study.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51877013), the Jiangsu Provincial Key Research and Development Program (BE2021636), and the Science and Technology Project of Changzhou City (CE20205056). This work was also sponsored by the Qing Lan Project of Jiangsu Province.

References

- [1] N. Saji, T. Sato, K. Sakuta et al., "Chronic kidney disease is an independent predictor of adverse clinical outcomes in patients with recent small subcortical infarcts," *Cerebrovascular Diseases Extra*, vol. 4, no. 2, pp. 174–181, 2014.
- [2] E. O'Lone, M. Connors, P. Masson et al., "Cognition in people with end-stage kidney disease treated with hemodialysis: a systematic review and meta-analysis," *American Journal of Kidney Diseases*, vol. 67, no. 6, pp. 925–935, 2016.
- [3] J. M. Bugnicourt, O. Godefroy, J. M. Chillon, G. Choukroun, and Z. A. Massy, "Cognitive disorders and dementia in CKD: the neglected kidney-brain Axis," *Journal of the American Society of Nephrology*, vol. 24, no. 3, pp. 353–363, 2013.
- [4] Y. L. Zhao, Y. H. Zhang, Z. K. Yang et al., "Sleep disorders and cognitive impairment in peritoneal dialysis: a multicenter prospective cohort study," *Kidney & Blood Pressure Research*, vol. 44, no. 5, pp. 1115–1127, 2019.
- [5] X. Liang, J. Q. Wen, L. Ni et al., "Altered Pattern of Spontaneous brain activity in the patients with end-stage renal disease: a resting-state functional MRI study with regional homogeneity analysis," *PLoS One*, vol. 8, no. 8, Article ID e71507, 2013.
- [6] M. C. Chou, T. J. Hsieh, Y. L. Lin et al., "Widespread white matter alterations in patients with end-stage renal disease: a voxelwise diffusion tensor imaging study," *American Journal of Neuroradiology*, vol. 34, no. 10, pp. 1945–1951, 2013.
- [7] C. Chai, M. J. Zhang, M. M. Long et al., "Increased brain iron deposition is a risk factor for brain atrophy in patients with haemodialysis: a combined study of quantitative susceptibility mapping and whole brain volume analysis," *Metabolic Brain Disease*, vol. 30, no. 4, pp. 1009–1016, 2015.
- [8] Z. J. Jiang, Y. J. Zhang, Z. N. Cheng et al., "Evaluation of cognitive impairment by voxel incoherent motor imaging in patients with end-stage renal disease," *Chinese Journal of Behavioral Medicine and Brain Science*, vol. 30, no. 5, pp. 415–419, 2021.
- [9] Z. X. Lu, L. Y. Tu, C. Zu, and D. Q. Zhang, "Prediction of clinical variable values for Alzheimer's disease based on brain connectivity networks," *CAAI Transactions on Intelligent Systems*, vol. 12, no. 03, pp. 355–361, 2017.
- [10] M. Y. Yang, W. Hou, P. Yang, W. B. Zou, T. F. Wang, and B. Y. Lei, "Prediction of Alzheimer's disease clinical score based on longitudinal incomplete data combined with deep integrated regression," *Chinese Journal of Biomedical Engineering*, vol. 38, no. 02, pp. 166–175, 2019.
- [11] Y. D. Zhang and S. H. Wang, "Detection of Alzheimer's disease by displacement field and machine learning," *PeerJ*, vol. 3, p. e1251, 2015.
- [12] S. H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y. D. Zhang, "Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2021.
- [13] W. Wen, Y. H. Wan, X. H. Zhang, and Z. Y. Wen, "Text feature selection based on improved CHI and PCA," *Computer Engineering & Science*, vol. 43, no. 9, pp. 1645–1652, 2021.
- [14] S. T. Yang, "Research on CET 4 score prediction model based on SVR," *Computer Knowledge and Technology*, vol. 17, no. 18, pp. 26–28, 2021.
- [15] S. F. Wang, X. Z. Yang, Z. Y. Dong, and C. C. Shi, "Remote sensing image change detection based on relief-PCA feature selection," *Journal of Graphics*, vol. 40, no. 01, pp. 117–123, 2019.
- [16] W. D. Zheng, Z. G. Li, H. Z. Jia, and C. Gao, "Prediction model of steelmaking end point based on improved Whale optimization algorithm and least square support vector machine," *Acta Electronica Sinica*, vol. 47, no. 3, pp. 700–706, 2019.
- [17] L. Shen, Q. T. Wang, and J. Shi, "Single-modal neuroimaging computer aided diagnosis for schizophrenia based on ensemble learning using privileged information," *Journal of Biomedical Engineering*, vol. 37, no. 3, pp. 405–411, 2020.
- [18] X. P. Zhang and X. Z. Zhang, "Research on photovoltaic fault diagnosis based on time domain characteristics," *Renewable Energy Resources*, vol. 39, no. 6, pp. 760–765, 2021.
- [19] J. G. Wang, K. Chen, W. X. Zhang, and B. Qin, "Fault diagnosis of rolling bearing based on whale optimized multi-core support vector machine," *Electronic Design Engineering*, vol. 29, no. 3, pp. 31–35, 2021.
- [20] L. M. Liu, P. Li, M. X. Chu, and C. Gao, "End-point prediction of 260 tons basic oxygen furnace (BOF) steelmaking based on WNPSVR and WOA," *Journal of Intelligent and Fuzzy Systems*, vol. 41, no. 2, pp. 2923–2937, 2021.
- [21] W. Zhe, "Optimizing BP neural network prediction model based on WOA," *International Core Journal of Engineering*, vol. 7, no. 9, pp. 342–348, 2021.
- [22] Z. Y. Lian, L. J. Duan, Y. H. Qiao, J. C. Chen, J. Miao, and M. G. Li, "The improved ELM algorithms optimized by bionic

- WOA for EEG classification of brain computer interface,” *IEEE Access*, vol. 9, Article ID 67405, 2021.
- [23] Z. K. Zhu and Y. Sun, “A minimum cross-entropy multi-thresholds segmentation algorithm based on improved WOA,” *MATEC Web of Conferences*, vol. 336, Article ID 07003, 2021.
- [24] C. Charin, D. Ishak, M. A. A. Mohd Zainuri, B. Ismail, and M. K. Mohd Jamil, “A hybrid of bio-inspired algorithm based on Levy flight and particle swarm optimizations for photovoltaic system under partial shading conditions,” *Solar Energy*, vol. 217, pp. 1–14, 2021.
- [25] Y. D. Zhang, S. H. Wang, Y. X. Sui et al., “Multivariate approach for Alzheimer’s disease detection using stationary Wavelet entropy and predator-prey particle swarm optimization,” *Journal of Alzheimer’s Disease*, vol. 65, no. 3, pp. 855–869, 2018.
- [26] D. Nasri, D. Mokeddem, B. Bourouba, and J. Bosche, “A novel Levy flight trajectory-based salp swarm algorithm for photovoltaic parameters estimation,” *Journal of Information and Optimization Sciences*, vol. 42, no. 8, pp. 1841–1867, 2021.
- [27] K. Balakrishnan, R. Dhanalakshmi, and U. M. Khair, “Improved salp swarm algorithm based on the Levy flight for feature selection,” *The Journal of Supercomputing*, vol. 77, no. 11, Article ID 12399, 2021.
- [28] A. Isazadeh, O. Tarkhaneh, and H. J. Khamnei, “A new hybrid strategy for data clustering using cuckoo search based on Mantegna Levy distribution, PSO and k-means,” *International Journal of Computer Applications in Technology*, vol. 58, no. 2, p. 137, 2018.
- [29] P. Sulewski, “DS normal distribution: properties and applications,” *Lobachevskii Journal of Mathematics*, vol. 42, no. 12, pp. 2980–2999, 2021.
- [30] A. Loechte, I. Rojas Ruiz, and P. Gloesekoetter, “Battery state estimation with ANN and SVR evaluating electrochemical impedance spectra generalizing DC currents,” *Applied Sciences*, vol. 12, no. 1, p. 274, 2021.
- [31] C. N. Li, Y. H. Shao, D. Zhao, Y. R. Guo, and X. Y. Hua, “Feature selection for high-dimensional regression via sparse LSSVR based on Lp-norm,” *International Journal of Intelligent Systems*, vol. 36, no. 2, pp. 1108–1130, 2020.
- [32] Y. T. Zhang, Z. T. Xi, J. H. Zheng, H. F. Shi, and Z. Q. Jiao, “GWLS: a novel model for predicting cognitive function scores in patients with end-stage renal disease,” *Frontiers in Aging Neuroscience*, vol. 14, Article ID 834331, 2022.
- [33] S. H. Wang, S. Du, Y. Zhang et al., “Alzheimer’s disease detection by Pseudo Zernike moment and linear regression classification,” *CNS & Neurological Disorders - Drug Targets*, vol. 16, no. 1, pp. 11–15, 2017.
- [34] J. H. Wang, X. Zuo, and Y. He, “Graph-based network analysis of resting-state functional MRI,” *Frontiers in Systems Neuroscience*, vol. 4, p. 16, 2010.
- [35] X. A. Bi, Y. Xie, H. Wu, and L. Y. Xu, “Identification of differential brain regions in MCI progression via clustering-evolutionary weighted SVM ensemble algorithm,” *Frontiers of Computer Science*, vol. 15, no. 6, Article ID 156903, 2021.
- [36] S. H. Wang, D. R. Nayak, D. S. Guttery, X. Zhang, and Y. D. Zhang, “COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis,” *Information Fusion*, vol. 68, pp. 131–148, 2021.
- [37] P. H. Janak and K. M. Tye, “From circuits to behaviour in the amygdala,” *Nature*, vol. 517, no. 7534, pp. 284–292, 2015.
- [38] J. F. Farragher, H. J. Polatajko, and S. V. Jassal, “The relationship between fatigue and depression in adults with end-stage renal disease on chronic in-hospital hemodialysis: a scoping review,” *Journal of Pain and Symptom Management*, vol. 53, no. 4, pp. 783–803, 2017.
- [39] C. Qin, Y. Liang, X. Tan et al., “Altered whole-brain functional topological organization and cognitive function in type 2 diabetes mellitus patients,” *Frontiers in Neurology*, vol. 10, p. 599, 2019.
- [40] L. R. Squire, J. T. Wixted, and R. E. Clark, “Recognition memory and the medial temporal lobe: a new perspective,” *Nature Reviews Neuroscience*, vol. 8, no. 11, pp. 872–883, 2007.
- [41] A. J. Shackman, T. V. Salomons, H. A. Slagter, A. S. Fox, J. J. Winter, and R. J. Davidson, “The integration of negative affect, pain and cognitive control in the cingulate cortex,” *Nature Reviews Neuroscience*, vol. 12, no. 3, pp. 154–167, 2011.
- [42] M. C. Chou, C. H. Ko, J. M. Chang, and T. J. Hsieh, “Disruptions of brain structural network in end-stage renal disease patients with long-term hemodialysis and normal-appearing brain tissues,” *Journal of Neuroradiology*, vol. 46, no. 4, pp. 256–262, 2019.
- [43] S. Luo, R. F. Qi, J. Q. Wen et al., “Abnormal intrinsic brain activity patterns in patients with end-stage renal disease undergoing peritoneal dialysis: a resting-state functional MR imaging study,” *Radiology*, vol. 278, no. 1, pp. 181–189, 2016.
- [44] S. M. Li, X. F. Ma, R. W. Huang et al., “Abnormal degree centrality in neurologically asymptomatic patients with end-stage renal disease: a resting-state fMRI study,” *Clinical Neurophysiology*, vol. 127, no. 1, pp. 602–609, 2016.
- [45] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical Review Letters*, vol. 87, no. 19, Article ID 198701, 2001.
- [46] A. Kunz and C. Iadecola, “Cerebral vascular dysregulation in the ischemic brain,” *Handbook of Clinical Neurology*, vol. 92, pp. 283–305, 2009.
- [47] X. A. Bi, X. Hu, H. Wu, and Y. Wang, “Multimodal data analysis of Alzheimer’s disease based on clustering evolutionary random forest,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2973–2983, 2020.

Research Article

Multimodal Sentiment Analysis Based on Cross-Modal Attention and Gated Cyclic Hierarchical Fusion Networks

Zhibang Quan , Tao Sun , Mengli Su, and Jishu Wei

School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

Correspondence should be addressed to Tao Sun; suntao0906@163.com

Received 21 April 2022; Accepted 15 July 2022; Published 9 August 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Zhibang Quan 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.

Multimodal sentiment analysis has been an active subfield in natural language processing. This makes multimodal sentiment tasks challenging due to the use of different sources for predicting a speaker's sentiment. Previous research has focused on extracting single contextual information within a modality and trying different modality fusion stages to improve prediction accuracy. However, a factor that may lead to poor model performance is that this does not consider the variability between modalities. Furthermore, existing fusion methods tend to extract the representational information of individual modalities before fusion. This ignores the critical role of intermodal interaction information for model prediction. This paper proposes a multimodal sentiment analysis method based on cross-modal attention and gated cyclic hierarchical fusion network MGHF. MGHF is based on the idea of distribution matching, which enables modalities to obtain representational information with a synergistic effect on the overall sentiment orientation in the temporal interaction phase. After that, we designed a gated cyclic hierarchical fusion network that takes text-based acoustic representation, text-based visual representation, and text representation as inputs and eliminates redundant information through a gating mechanism to achieve effective multimodal representation interaction fusion. Our extensive experiments on two publicly available and popular multimodal datasets show that MGHF has significant advantages over previous complex and robust baselines.

1. Introduction

Every day, a large and meaningful amount of information is generated around us. Most of this information is generated on the web, and social media is a centralized area of information on the web. It covers many topics, opinions, sentiments, and emotions closely related to our lives. Multimodal sentiment analysis (MSA) has been an active subfield in natural language processing [1, 2]. This is mainly due to its wide range of applications, such as government elections [3], intelligent healthcare [4], and chatbot recommendation systems for human-computer interaction [5]. Compared to traditional sentiment analysis, MSA uses multiple sources (excerpted raw text, acoustic, and visual information) to make predictions about the sentiment expressed by a specific object in a specific period. One of the multimodal sentiment analysis challenges is to model the interactions between different modalities because they

contain supplementary and complementary information [6]. Another factor that limits the performance of multimodal sentiment analysis tasks is data fusion. This is because there are multiple recurring problems, such as missing values and misalignment in visual and auditory modalities [7].

In recent years, researchers have designed sophisticated fusion models. Zadeh et al. [8] designed the tensor fusion network, which uses a Cartesian product to fuse the feature vectors of three modalities; this provided a new idea for multimodal data processing. Tsai et al. [9] designed a multimodal transformer that processed all modalities together to obtain the predicted sentiment scores. Although these methods have achieved good results, a problem that may affect the final prediction effect is that these models ignore the differences between different modalities, which may lead to the loss of crucial prediction information during the modal representation acquisition stage. Hazarika et al. [10] designed a modality-specific and modality-invariant

feature space, combining two types of representations with similarity loss, reconstruction loss, and dissimilarity loss to evaluate the model effect. Yu et al. [11] used a multitask format and introduced an automatic modal label generation module in the training phase to assist the main task channel, saving manual labelling time, and thus improving efficiency. Although these studies also achieved encouraging results, they lacked intermodal information interaction during the modal fusion phase. Doing so may result in the redundant information present in the upper stage being retained in the final prediction stage, making the model performance poor. As shown in Figure 1, there are two opposite prediction results after the same text interacts with different modalities. For example, an ordinary language with ordinary acoustic features is predicted as a negative sentiment. In contrast, the same type of language with positive visual features is predicted as a positive sentiment. This indicates that different modal combinations have a fundamental impact on sentiment prediction. It should be noted that, in Figure 1, “?” indicates that sentiment cannot be accurately identified, “-” represents negative sentiment, and “+” represents positive sentiment. The number of these symbols signifies the intensity of the sentiment.

To address the mentioned issues, inspired by cross-modal matching and interaction modelling, we propose a novel multimodal sentiment analysis framework, MGHF. It includes mid-term interactions performed in the modal representation phase and post-term interactions in the modal fusion phase. This approach allows the model to fully perceive various modalities’ potential representational sentiment information, which helps us improve the fusion and prediction results. Although previous studies have shown that text modality is the most critical [9, 12], we still believe that the information implied by any modality should be considered in the MSA task. Specifically, MGHF employs a flexible strategy for modality variability by using appropriate neural networks for different modalities. In the medium-term interaction learning phase, MGHF performs cross-modal attention interactions for acoustic modality, visual modality, and text modality, respectively, to obtain text-based acoustic representation and text-based visual representation. Several past studies [13] have pointed out that task-related information is not evenly distributed across modalities, with the text modality contributing much more than other modalities. There are also studies [8, 9] that would fuse the text-video and audio modalities as a ternary symmetric structure, which does not take into account the variability of the various modalities and thus fails to fuse them correctly. According to previous experience, in order to make the text modality occupy a higher weight than other modalities in the later fusion stage. We combined the text-based acoustic representation with the text representation, the text-based visual representation with the text representation, and the text-based acoustic representation with the text-based visual representation in a two-by-two combination. We also design gated recurrent hierarchical fusion networks that dynamically interact with learning information representations between modal combinations to complement the information between combinations. Our

extensive experiments on the publicly available and popular datasets CMU-MOSI [14] and CMU-MOSEI [15] show that MGHF shows strong competitiveness over previous complex interaction and fusion baselines.

The contributions of this paper are summarized as follows:

- (i) A gated cyclic hierarchical fusion network for multimodal sentiment analysis is proposed. It dynamically interacts with information representations between 3 different modal pairs. The gated cyclic hierarchical fusion network enables sufficient interaction between each modal pair, eliminates redundant information between modal pairs, and maximizes the retention of valid representations for modal prediction.
- (ii) Inspired by distribution matching, we consider the interactions within different modalities. In the modal representation acquisition stage, we make the nonverbal sequences to cross-modal attention with text sequences, which can capture potential representations within different modalities while making the modal representations closer to the real sentiment expressions.
- (iii) Experiments conducted on two publicly available multimodal datasets show that our model has significant advantages over previous advanced complex baselines.

2. Related Work

This section introduces multimodal sentiment analysis, as well as related work on multimodal representation learning and data fusion.

2.1. Multimodal Sentiment Analysis. Unlike traditional sentiment analysis, multimodal sentiment analysis often uses multiple sources (excerpted text, audio, video, and other information) to fully and accurately predict the speaker’s sentiment orientation. Researchers have various ways to deal with MSA tasks, one of which is representative of the extraction of intramodal temporal information and the other is the extraction of intermodal interaction information. The former mainly uses neural networks such as the Long Short-Term Memory (LSTM) Network [16] for the extraction of modal contextual information [10, 17]. The latter can be further divided into early, late, and hybrid, depending on the fusion stage. Early fusion is the fusion approach used in the pre-extraction phase of the data. Rozgic et al. [18] used early fusion to connect multimodal representations as input to an inference model, which provides a novel idea for modal fusion. Zadeh et al. [19] designed a memory fusion network (MFN) using multiview sequential learning, which explicitly illustrates two interactions in the neural architecture. The post-fusion approach performs a series of necessary processing within the modality and intermodal data fusion in the final stage. Liu et al. [20] proposed a low-rank multimodal fusion approach to reduce the computational

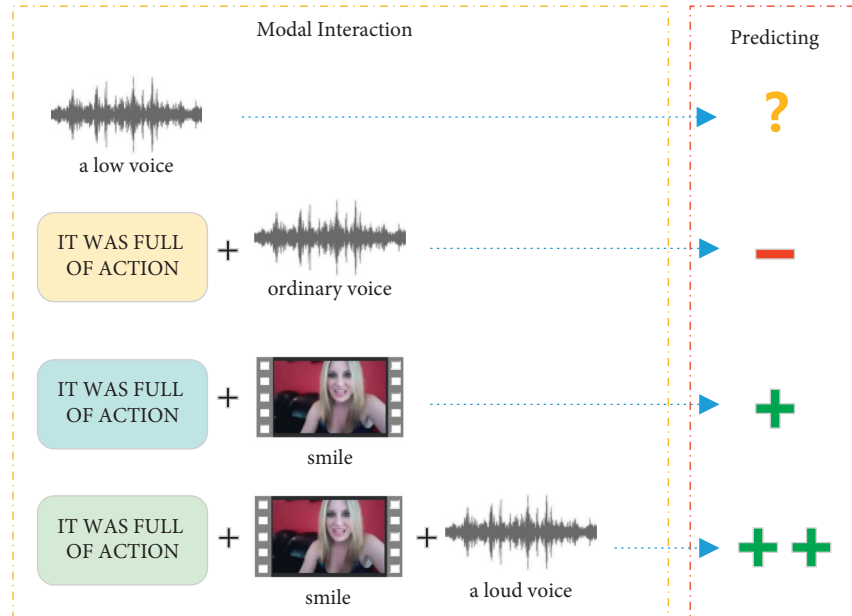


FIGURE 1: The combination of different modal pairs and sentiment prediction results.

complexity by using low-rank tensor fusion to improve efficiency. Other researchers have used hybrid fusion to improve the performance of MSA tasks. Dai et al. [21] used a simple but very effective hybrid modal fusion approach using weakly supervised multitask learning to improve the generalization performance of the dataset.

We differ fundamentally from previous work in that. First, there is a modal divide between different modalities, and using only the same neural network does not seem to yield useful information. Instead of considering a piece of single contextual information, we use the most appropriate strategy based on the modal sequence characteristics. After obtaining the initial representations, unlike in previous work, our interaction fusion does not only occur in the final stage. Useful potential information can be induced from the companion representations through the intermediate interaction stage. Similarly, the post-interaction stage of the modality is used to better retain information useful for prediction and eliminate redundant information. It is worth noting that instead of the traditional approach of treating text, audio, and video equally, we flexibly utilize the information useful to the task for each modality based on the contribution of the modality.

2.2. Representation Learning and Data Fusion. Representation learning methods can also be applied to multimodal sentiment analysis and have achieved significant results. Wang et al. [22] proposed a recursive attentional change embedding network to generate multimodal shifts. Hazarika et al. [10] proposed a way to learn multimodal invariant and specific representations while combining four different losses to evaluate the performance of the model. Yu et al. [11] proposed self-supervised multitask learning to learn modality-specific representations and introduced a single-peak annotation generation module to assist the main

task channel. In the context of sentiment analysis, multimodal fusion is essential because sentiment cues are usually distributed over different modalities [23]. Xiangbo et al. [24] proposed an extended-squeezed-excitation fusion network (ESE-FN) that fuses multimodal features in the modal and channel directions. The network learns extended-squeezed-excitation (ESE) caveats in the modal and channel directions to effectively solve the elderly activity recognition problem. Shu et al. [25] proposed a new weakly shared deep transport network (DTN) for converting cross-domain information from text to images. This provides ideas for interconversion across modalities. Based on this, Tang et al. [26] proposed a new generalized deep transmission network (DTN) for the transmission of information across heterogeneous, textual, and visual domains by establishing parameter sharing and representation sharing layers.

In view of this, our model is based on the late fusion of representation learning. Unlike previous studies, we learn representations across intramodal interactions while employing different combinations of modal interactions to obtain intermodal representations.

3. Materials and Methods

In this section, we will detail the main components of our model and their specific roles.

3.1. Task Setup. Multimodal data sequences in sentiment analysis consist of three main modalities which are the text modality (t), acoustic modality (a), and visual modality (v), respectively. The goal of multimodal sentiment analysis (MSA) is to predict the speaker's emotional polarity from a segment of discourse, which is also the input to the model in this paper. First, given the input discourse $U_{s \in \{t, a, v\}}$, this paper uses U_v to denote visual modal information, U_a to

denote acoustic modal information, and U_t to denote textual modal information. Here, $a \in R^{T_a \times d_a}$, $t \in R^{T_t \times d_t}$, $v \in R^{T_v \times d_v}$, and $T_{s \in \{t, a, v\}}$ denote the sequence length of a discourse, and $d_{s \in \{t, a, v\}}$ denote the dimensionality of the respective features.

3.2. Overall Architecture. In this paper, our multimodal sentiment analysis architecture consists of three primary and flexible modules as shown in Figure 2. They are the feature extraction module for each modality, the (acoustic-text/visual-text) cross-attention module, and the gated recurrent hierarchical fusion network module. For the text channel, we use pretrained BERT for its high-dimensional semantic extraction. For the acoustic and visual channels, we first feed the initial sequence into a 1D temporal convolution to obtain enough perceptual and temporal information. The obtained (acoustic/visual) representations are then learned cross-modally with textual representations, which can induce potential representational information for both acoustic and visual modalities, synergistic to the overall effective orientation. Notably, this cross-modal matching has been prominent in recent cross-modal learning approaches [27, 28]. Afterward, we feed the output of the two cross-modal attention (text-based acoustic representation and text-based visual representation) and the extracted textual modal representation into a gated recurrent hierarchical fusion network, which eliminates redundant modal information to obtain the final information for prediction. Of course, some of the modules in our model are flexible and can be reconfigured with any suitable baseline to accomplish different types of tasks.

3.3. Modality Representation. The acquisition of representation for our model is divided into three channels, namely, text channel, video channel, and audio channel. In the following, we describe the essential details of the model acquisition of representations.

3.3.1. Text Channel. For the text channel, we fine-tuned the pretrained model BERT [29] used as an extractor of text features, consisting of a 12-layer stacked transformer. The input text is preprocessed and fed to BERT for embedding by adding two special tags CLS and SEP. Consistent with recent work, the first word vector of the last layer is chosen in this paper as the average representation of the representation in the final 768-dimensional implicit state [30].

$$\begin{aligned} t_i &= \{[CLS], w_1, w_2, \dots, w_n, [SEP]\}, \\ f_t &= \text{BERT}(t_i, \theta_t^{\text{bert}}) \in R^{d_t}, i \in [1, n]. \end{aligned} \quad (1)$$

Here, t represents the initial sequence of text and θ_t^{bert} represents the hyperparameters of the BERT pretrained model.

3.3.2. Audio and Video Channels. For the audio and video channels, we designed two independent modal characterization modules for the nonverbal sequences, and they function before fusion. We followed previous work [11] and processed the raw data using a pretrained toolkit to obtain the initial vector features.

Temporal Convolutions. First, to make our modalities sufficiently perceptible, we pass the input sequence through a one-dimensional temporal convolution layer.

$$U_m^* = \text{Conv1D}(U_m, k_m) \in R^{T_m \times d}, \quad (2)$$

where $\text{Conv1D}(\bullet)$ is the one-dimensional temporal convolution function, k_m is the size of the convolution kernel used by the modality m , U_m is the input sequence of modality m , d is the common dimension, and T_m denotes the discourse length of modality m ; here, $m \in \{a, v\}$.

Positional Embedding. To equip the sequences with temporal information, following Vaswani et al. [31], the position embedding (PE) is bracketed to U_m^* as follows:

$$U_m^{*'} = U_m^* + \text{PE}(T_m, d), \quad (3)$$

where $\text{PE}(T_m, d) \in R^{T_m \times d}$, the purpose is to compute the embedding for each position index. $\text{PE}(\bullet)$ represents the position embedding function, $m \in \{a, v\}$.

Cross-Attention Transformers. We then perform cross-modal cross-attention on the resulting sequences, which induces potential representational information for both acoustic and visual modalities that are synergistic to the overall practical orientation. It is worth noting that our cross-modal attention occurs only between text and acoustic modalities and between text and visual modalities, which allows the text modality that contributes most to the task to be weighted higher than the other modalities and ensures the relative independence of the visual and acoustic channels. We justify this approach in Section 5.2.1.

$$\text{C_Attention}_{a-t}(Q, K, V) = \text{softmax}\left(\frac{Q_t K_a^T}{\sqrt{d_h}}\right) V_a, \quad (4)$$

$$\text{C_Attention}_{v-t}(Q, K, V) = \text{softmax}\left(\frac{Q_t K_v^T}{\sqrt{d_h}}\right) V_v,$$

where Q_t represents the query vector for the text modality and $K_a, V_a, K_v,$ and V_v denote the key vectors and value vectors of the acoustic and visual modalities. $\text{softmax}(\bullet)$ represents the softmax function, d_h represents the dimensionality of the modality, and T represents transpose.

Transformer computes multiple parallel attentions, and the output of each attention is called a head. The i^{th} head is computed as

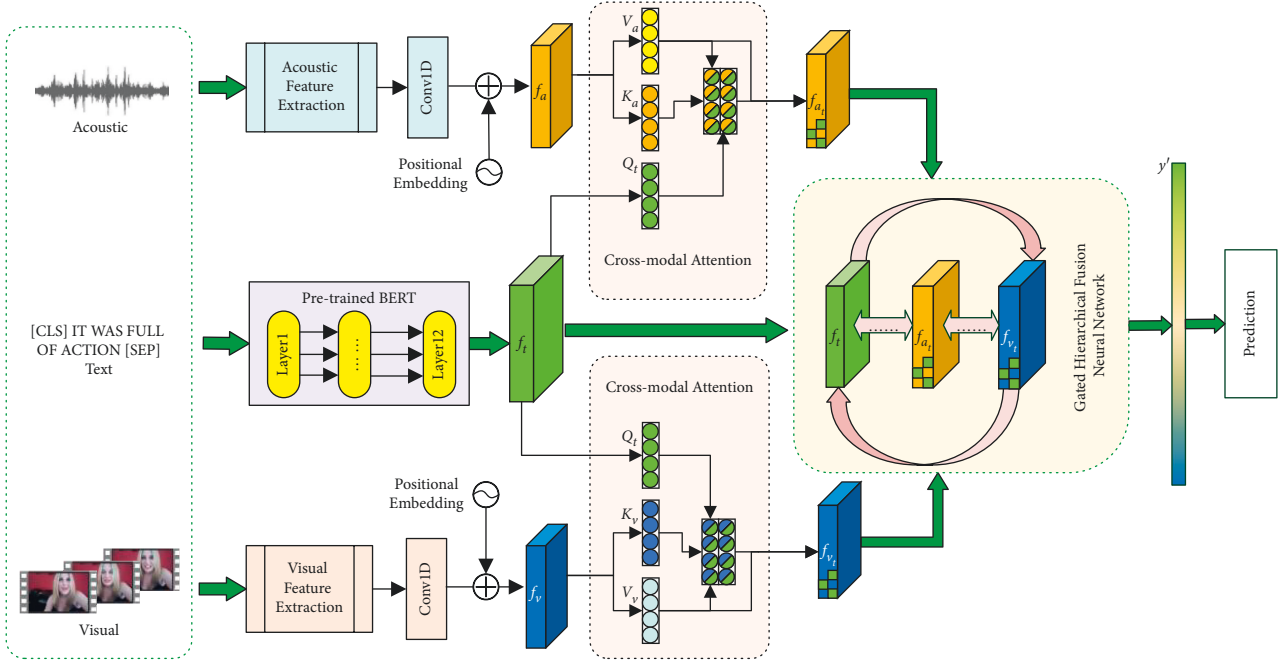


FIGURE 2: MGHF: cross-modal attention with hierarchical recurrent fusion network.

$$\text{head}_m^i = \text{Attention}_m(Q_t W_i^{Q_t}, K_m W_i^{K_m}, V_m W_i^{V_m}), \quad (5)$$

where $W_i^{Q_t} \in R^{d_t \times d_q}$ is the weight matrix of Q_t when computing the head of the i^{th} text modality; $W_i^{K_m} \in R^{d_m \times d_k}$ is the weight matrix of K_m when computing the head of the i^{th} modality; and $W_i^{V_m} \in R^{d_m \times d_v}$ is the weight matrix of V_m when computing the head of the i^{th} modality, where $m \in \{a, v\}$.

After that, we connect all heads of m modalities, which is denoted as Y_m^* as follows:

$$\begin{aligned} Y_m^* &= \text{MultiHead}(Q_t, K_m, V_m) \\ &= \text{Concat}(\text{head}_m^1, \text{head}_m^2, \dots, \text{head}_m^n) W_m^o, \end{aligned} \quad (6)$$

where W_m^o is the weight matrix multiplied after the splicing the head of m modalities and n denotes the number of self-attention heads we use. Here, we have $n=10$, $\text{Concat}(\bullet)$ is the splicing operation, $m \in \{a, v\}$.

Thus, the text-based acoustic representation $f_{a,t}$ and the text-based visual representation $f_{v,t}$ can be obtained.

$$\begin{aligned} f_{a,t} &= \text{MultiHead}(Y_a^*; \theta_a^{-att}), \\ f_{v,t} &= \text{MultiHead}(Y_v^*; \theta_v^{-att}), \end{aligned} \quad (7)$$

where $\theta_a^{att} = \{W_a^Q, W_a^K, W_a^V, W_a^O\}$ and $\theta_v^{att} = \{W_v^Q, W_v^K, W_v^V, W_v^O\}$ represent the main hyperparameters required for the cross-attention module.

3.4. Gated Cyclic Hierarchical Fusion Networks. In previous studies [10, 11], after obtaining valid representations, most of the modal representations are simply spliced directly for final prediction. This can inadvertently add redundant information to them. To allow the redundant information in

the representations to be effectively removed, we designed a gated recurrent fusion network (see Figure 3). This module is flexible and can be paired with other benchmarks to enhance the effect. Of course, we also verified the effectiveness of the hierarchical fusion network.

We used the text-based acoustic representation $f_{a,t}$ and text-based visual representation $f_{v,t}$ as well as text representation f_t as inputs to the gated recurrent hierarchical network. Previous experience [9, 12] has shown that the text modality contributes much more to the task than the other modalities. Given this, we combined the text-based visual representation, the text-based acoustic representation, and the text representation in two combinations to ensure that the text modality accounts for a high weight, which would result in three combinations of representations.

$$\begin{aligned} f_{a,t\oplus t} &= \text{Concat}(f_t, f_{a,t}), \\ f_{v,t\oplus t} &= \text{Concat}(f_t, f_{v,t}), \\ f_{a,t\oplus v,t} &= \text{Concat}(f_{a,t}, f_{v,t}). \end{aligned} \quad (8)$$

where $\text{Concat}(\bullet)$ denotes the combination operation, $f_{a,t\oplus t}$ denotes the combination of text-based acoustic representation with text, $f_{v,t\oplus t}$ denotes the combination of text-based visual representation with text, and $f_{a,t\oplus v,t}$ denotes the combination of text-based acoustic representation with text-based visual representation.

After obtaining the specified three combinations, we fed them into a bi-directional gated recurrent network (Bi-GRU). The purpose of doing so is to allow the information between different modalities to be fully perceived and to effectively remove redundant and irrelevant information from the representations through the gating mechanism. We also employ a bi-directional long and short memory

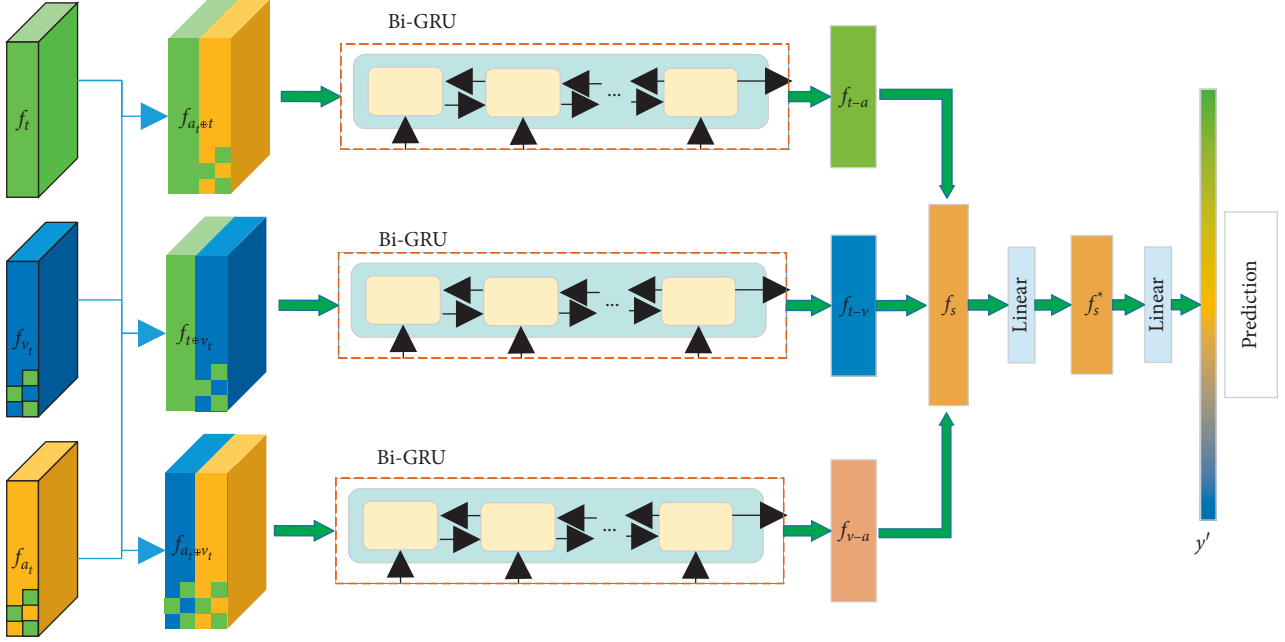


FIGURE 3: Gated cyclic hierarchical fusion network.

(Bi-LSTM) network. By comparison, we found that the former has more straightforward parameters and faster training speed, and its results are comparable.

$$\begin{aligned} f_{t-a} &= Bi_GRU(f_{a_t \oplus t}, \theta^{gru}), \\ f_{t-v} &= Bi_GRU(f_{v_t \oplus t}, \theta^{gru}), \\ f_{a-v} &= Bi_GRU(f_{a_t \oplus v_t}, \theta^{gru}), \end{aligned} \quad (9)$$

where $Bi_GRU(\bullet)$ represents the bi-directional gated recurrent cell network and θ^{gru} represents the hyperparameters of the gated recurrent cell network.

After that, we combine the outputs of the gated cyclic hierarchical fusion networks and feed them into the fully connected layer for the final prediction.

$$\begin{aligned} f_s &= \text{concat}(f_{t-a}, f_{t-v}, f_{a-v}), \\ f_s^* &= \text{ReLU}(W_{l1}^{sT} \otimes f_s + b_{l1}^s), \end{aligned} \quad (10)$$

where $W_{l1}^s \in R^{(d_t+d_a+d_v) \times d_s}$ and ReLU are the relu activation functions and \otimes represents the elemental product.

Finally, f_s^* is used as the final representation and for the prediction task.

$$y' = \text{ReLU}(W_{l2}^{sT} \otimes f_s^* + b_{l2}^s), \quad (11)$$

where $W_{l2}^s \in R^{d_s \times 1}$.

4. Experiment

In this section, we will detail the specifics of our experiments.

4.1. Datasets. *CMU-MOSI* [14]. The Multimodal Sentiment Intensity Corpus dataset is a collection of 2199 viewpoint video clips. This dataset is a popular benchmark for

multimodal sentiment analysis. Each opinion video is annotated with sentiment in the range of $[-3, 3]$. The dataset is strictly labelled using tags for subjectivity, emotional intensity, per-frame, per-viewpoint annotated visual features, and per-millisecond annotated audio features.

CMU-MOSEI [15]. The multimodal Opinion Sentiment and Sentiment Intensity dataset is the largest multimodal sentiment analysis and recognition dataset. MOSEI contains more than 23,500 sentence expression videos from more than 1,000 online YouTube speakers. The dataset is gender-balanced. All sentences were randomly selected from different videos of topics and monologues. Videos were transcribed and correctly punctuated. We give the detailed dataset settings in the experiments (see Table 1).

4.2. Modality Processing. To ensure fair competition with other baselines, we follow previous work [11] and treat the three modalities as a typical tensor described as follows:

Text Modality. Most previous studies have used glove [32] as a source of word embedding and achieved good results. Considering the strong performance of pre-trained models, we prefer to use the pretrained language model BERT [29]. For a fair and objective comparison, we adopted the latter as the processing tool for our text modality.

Audio Modality. For audio data, the acoustic analysis framework COVAREP [33] was used to extract up to 12 Mel-frequency cepstral coefficients, pitch, turbid/apparent segmentation features, and so on. All features are related to mood and intonation. It is worth noting that acoustic features are processed to align with the text features.

TABLE 1: MOSI and MOSEI dataset size settings.

Dataset	MOSI	MOSEI
Train	1284	16326
Valid	229	4659
Test	686	1871
All	2199	22856

Video Modality. Video modality raw features are used to extract facial expression features using Facet (<https://imotions.com/platform/>), which includes facial action units and facial poses based on the Facial Action Coding System (FACS) [34]. The process is repeated for each sampled frame within the vocalized video sequence.

Eventually, we align the initial modalities with the text for the alignment operation. This will allow our experiments to proceed appropriately and ensure fair experimental comparison results.

4.3. Evaluation Metrics. Again, to be fair, we split the MSA task into a regression task and a classification task. This paper will have five valuation metrics, which are: secondary precision (ACC-2) and F1-score. Mean Absolute Error (MAE): it directly calculates the error between the prediction and the authentic number labels. Level 7 Precision (ACC-7) and Pearson Correlation (Corr) measure the standard deviation from the human-annotated actual value. It is worth noting that the secondary precision and F1 scores were divided into two groups: negative and non-negative feelings (including neutral feelings), and negative and positive feelings, respectively. In addition to the value of MAE, higher scores imply better results.

4.4. Baseline. We compared the performance of MGHF with several multimodal fusion frameworks, including state-of-the-art models, as follows.

4.4.1. Previous Models

- (i) *TFN.* Tensor fusion network [8] is based on Cartesian product to calculate the tensor of each modality for capturing the interaction information of unimodal, bimodal, and three modalities.
- (ii) *LMF.* Low-order multimodal fusion [20] is an improvement of the tensor fusion network (TFN) to reduce the computational complexity and improve the efficiency by using low-order tensor fusion.
- (iii) *MFM.* Multimodal Factorization Model [35] demonstrates flexible generation capability by adjusting independent factors and reconstructs missing modes.
- (iv) *MULT.* Multimodal Transformer (MULT) [9] extends the multimodal converter architecture using directed pairwise cross-attention, which converts

one modality to another using directed pairwise cross-attention.

- (v) *ICCN.* Interaction Canonical Correlation Network (ICCN) [13] learns correlations between text, audio, and video through Deep Typical Correlation Analysis (DCCA).
- (vi) *MISA.* Learning Modality-Invariant and Modality-Specific Representations (MISA) [10] combines a combination of distribution similarity, orthogonal loss, reconstruction loss, and task prediction loss for learning the representation of different modalities and the representation of fused modalities.
- (vii) *MAG-BERT* [36]. A multimodal adaptation gate was designed for the BERT alignment gate and inserted into the general BERT model to optimize the fusion process.

4.4.2. State-of-the-Art. For sentiment analysis tasks, the results of Self-MM [11], a self-supervised multitask learning framework, on both MOSI and MOSEI datasets represent state-of-the-art (SOTA) models. Self-MM assigns a single-peaked training task with automatically generated labels to each modality, allowing multimodal sentiment analysis tasks to be performed in a multitask context.

5. Results and Discussion

In this section, the experimental results of the model are analysed and discussed in detail.

5.1. Quantitative Results. We compared the MGHF with currently popular benchmarks, including the state-of-the-art (SOTA) model (see Tables 2 and 3). For a fair comparison, we divided the models into two categories depending on the data setup, aligned and unaligned. In our experiments, first, compared with the aligned advanced models, our models all achieved similar or even surpassed results. In addition, our models achieve significant gains on all indicators of the regression as well as on some of the categorical indicators compared to the unaligned models. In addition, we reproduce two strong baselines, MISA and self-mm, under the same conditions. We find that MGHF outperforms them on most indicators. On the MOSI dataset, MGHF achieves competitive scores on both classification tasks. On the regression task, MGHF also improves the SOTA model by various degrees. Our model also outperforms some complex fusion mechanisms, such as TFN and LFN. The above results show that our model can be applied to different data scenarios and achieve significant improvements. We visualized some of the metrics, which can help us visualize how the model is performing (see Figure 4).

5.2. Ablation Study. We set up ablation experiments to verify the performance of our model, which is divided into the following main parts.

TABLE 2: Results on MOSI. Note: (B) Means the language features are based on BERT; model with * represents the best results for recurrence under the same conditions. \circ is from [10], and \diamond is from [11]. In indicators Acc-2 and F1-score, the left side of “/” is calculated for negative and non-negative sentiment, while the right side of “/” is calculated for negative and positive sentiment.

Models	MOSI					Data setting
	MAE (\downarrow)	Corr (\uparrow)	Acc-7 (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)	
TFN (B) \circ	0.901	0.698	34.9	-/80.8	-/80.7	Unaligned
LMF (B) \circ	0.917	0.695	33.2	-/82.5	-/82.4	Unaligned
MFM (B) \circ	0.877	0.706	35.4	-/81.7	-/81.6	Aligned
MULT*	0.918	0.680	36.47	77.93/79.3	77.91/79.34	Aligned
ICCN (B) \diamond	0.860	0.710	39.0	-/83.0	-/83.0	Unaligned
MISA (B) \diamond	0.783	0.761	42.3	81.8/83.4	81.7/83.6	Aligned
MAG-BERT (B) \diamond	0.731	0.789	—	82.54/84.3	82.59/84.3	Aligned
Self-MM (B) \diamond	0.713	0.798	—	84.42/85.95	84.42/85.95	Unaligned
MISA (B)*	0.759	0.787	42.57	81.05/82.93	81.03/82.97	Aligned
Self-MM (B)*	0.718	0.796	45.77	83.09/84.09	83.10/84.96	Aligned
MGHF (B)	0.709	0.802	45.19	83.38/85.21	83.32/85.21	Aligned

TABLE 3: Results on MOSEI. Note: (B) Means the language features are based on BERT; model with * represents the best results for recurrence under the same conditions. \circ is from [10], and \diamond is from [11]. In indicators Acc-2 and F1-score, the left side of “/” is calculated for negative and non-negative sentiment, while the right side of “/” is calculated for negative and positive sentiment.

Models	MOSEI					Data setting
	MAE (\downarrow)	Corr (\uparrow)	Acc-7 (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)	
TFN (B) \circ	0.593	0.700	50.2	-/82.5	-/82.1	Unaligned
LMF (B) \circ	0.623	0.677	48.0	-/82.0	-/82.1	Unaligned
MFM (B) \circ	0.568	0.717	51.3	-/84.4	-/84.3	Aligned
MULT \circ	0.580	0.703	51.8	-/82.5	-/82.3	Aligned
ICCN (B) \circ	0.565	0.713	51.6	-/84.2	-/84.2	Unaligned
MISA (B) \diamond	0.555	0.756	52.2	83.6/85.5	83.8/85.3	Aligned
MAG-BERT (B) \diamond	0.539	0.753	—	83.79/85.23	83.74/85.08	Aligned
Self-MM (B) \diamond	0.530	0.765	—	82.81/85.17	82.53/85.30	Unaligned
MISA (B)*	0.558	0.748	51.45	82.14/85.09	82.44/84.94	Aligned
Self-MM (B)*	0.534	0.764	53.32	84.37/85.28	84.42/85.06	Aligned
MGHF (B)	0.528	0.767	53.70	85.25/85.30	85.09/84.86	Aligned

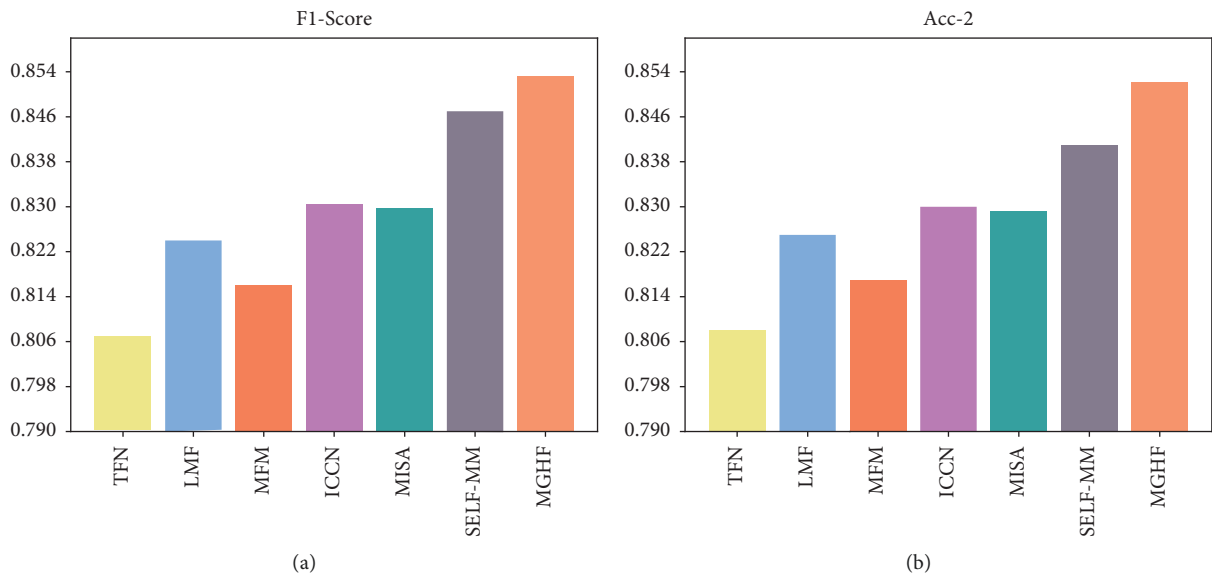


FIGURE 4: Comparison of the performance of each baseline model. (a) F1-score. (b) Acc-2.

5.2.1. *Representational Interaction.* First, for cross-modal attention interactions, we conducted the following experiments. The first group was performed for the interaction

between two modalities, and we did not consider acoustic-based text features and visual-based text features because this would make the text modality so heavily dominated that

TABLE 4: Performance tables for different cross-modal notes on MOSI and MOSEI datasets.

Task	MOSI				MOSEI			
	MAE (\downarrow)	Corr (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)	MAE (\downarrow)	Corr (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)
f_{a_v}	1.442	0.210	53.81	46.48/	1.315	0.197	60.13/61.25	60.48/59.38
f_{v_a}	1.321	0.233	62.33	57.85/58.94	1.244	0.182	58.48/59.47	61.63/61.48
f_{a_t}	0.896	0.393	68.52	64.44/65.38	0.815	0.213	64.15/63.18	64.85/64.25
f_{v_t}	0.901	0.384	71.49	69.12/67.20	0.843	0.241	63.48/63.14	63.54/63.89
$f_{v_t} + f_{a_v}$	0.976	0.223	73.62/71.40	71.04/64.31	0.821	0.213	61.84/62.37	61.23/61.66
$f_{a_t} + f_{v_a}$	0.957	0.381	74.22/72.67	71.04/65.86	0.784	0.230	63.24/62.56	61.05/60.72
$f_{a_t} + f_{v_t}$	0.819	0.486	76.80/76.01	75.72/74.84	0.763	0.361	72.18/72.56	74.37/74.03

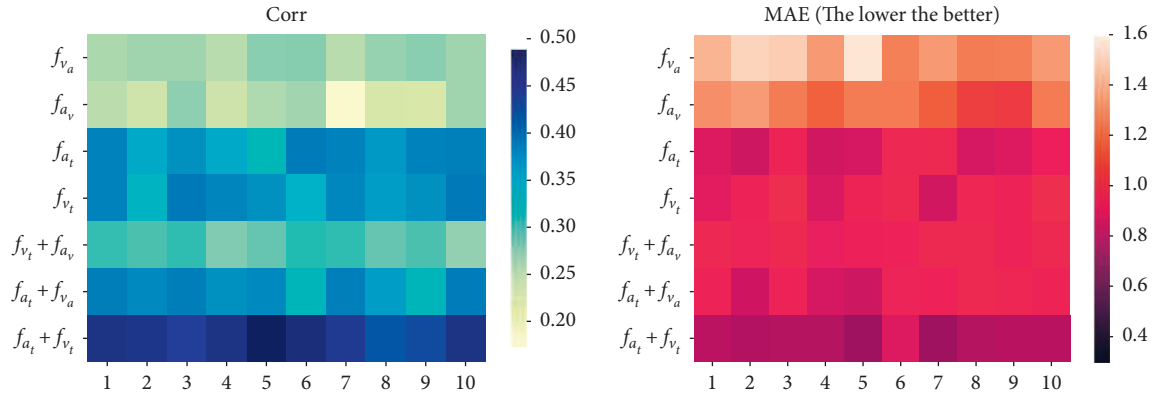


FIGURE 5: Visualization of different cross-modal interactions and their combined performance. (a) Corr. (b) MAE (the lower the better).

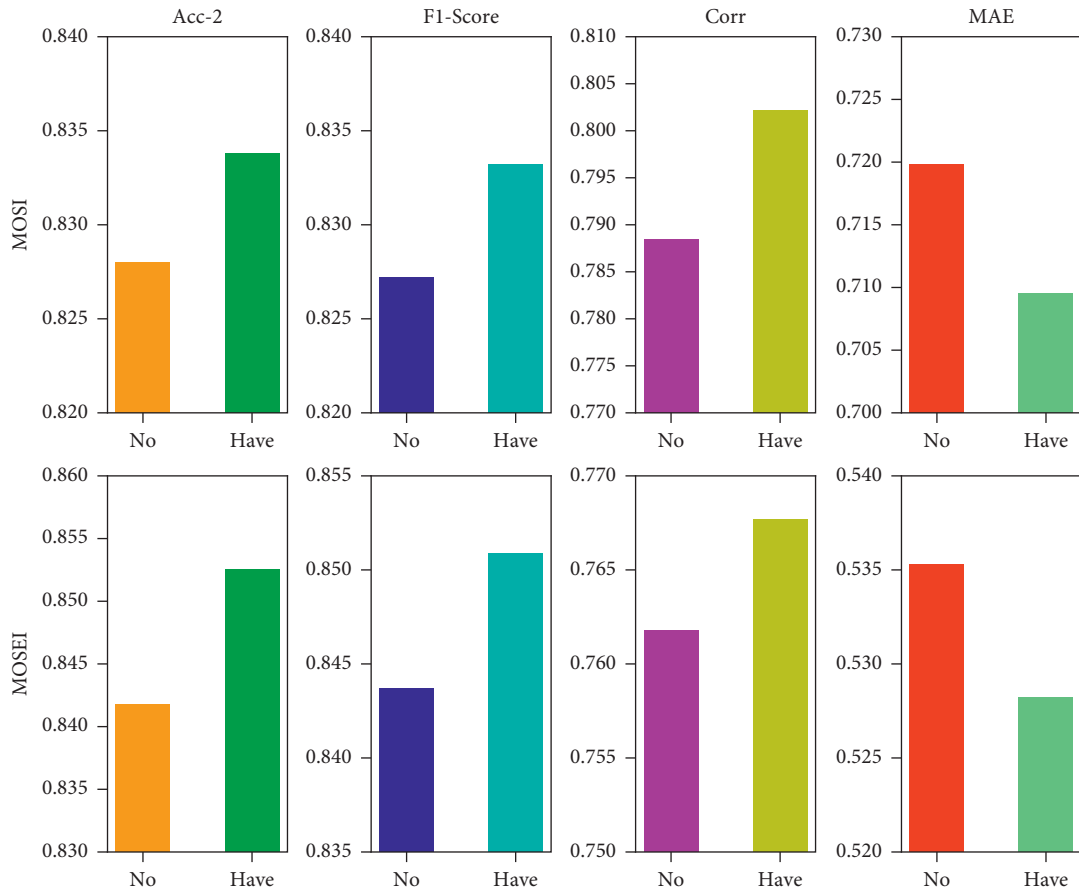


FIGURE 6: Gated cyclic hierarchical fusion network performance visualization.

TABLE 5: Ablation study results of fusion strategies on MOSI and MOSEI datasets.

Task	MOSI				MOSEI			
	MAE (\downarrow)	Corr (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)	MAE (\downarrow)	Corr (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)
MGHF_w/o(pc)	0.714	0.793	82.88	82.63	0.538	0.758	84.27	84.48
MGHF_LSTM	0.712	0.800	83.27	82.94	0.530	0.767	85.22	84.89
MGHF_original	0.709	0.802	83.38	83.32	0.528	0.767	85.25	85.09

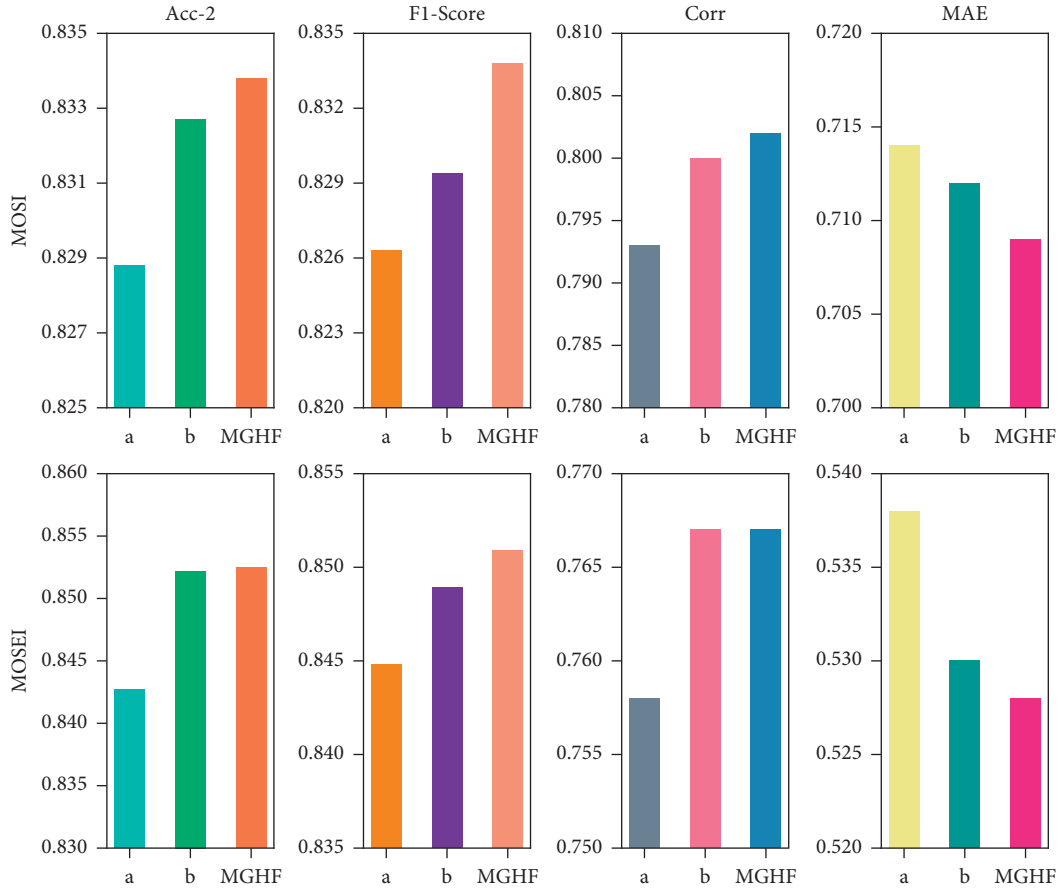


FIGURE 7: Ablation research in fusion strategies.

modal independence would be reduced or even disappear. The cross-modal attention between nonverbal sequences is hardly satisfactory, probably due to the characteristics of nonverbal sequence data. Acoustic, visual, and textual cross-modal attention seems to play an important role, which is consistent with previous studies [9, 12]. The second set of experiments was conducted after combining the cross-modal interaction representations obtained in the first set, which could help us elucidate whose combination of cross-modal interactions is more beneficial for the MSA task. In Table 4, it seems apparent that the combination of text-based acoustic and text-based visual representations performs the best. We believe this is partly because the text modality enhances the complementary acoustic and visual information, providing additional cues for semantic and affective disambiguation [37], and partly because it preserves the independence of the acoustic and visual modalities. We visualized this part of the experimental index scores for ten

randomly selected samples from the MOSI test set (see Figure 5), and similar results were observed on MOSEI.

5.2.2. Gated Recurrent Hierarchical Fusion Network Effectiveness. To verify the reliability of our proposed gated cyclic hierarchical fusion network, we will perform the multimodal sentiment analysis task under the same conditions without this fusion strategy. For visual comparison, two representative metrics from the classification and regression tasks are selected for evaluation, while the evaluation results are visualized. It is worth noting that among these metrics, higher scores imply better performance, except for the MAE metric. The results are shown in (Figure 6). Specifically, the gating mechanism effectively removes the redundant information contained in the previous stage. This not only implies that the representations obtained by the model in the prediction stage are inclusive of the potential

TABLE 6: Effect of inputs on prediction results for different gated cyclic hierarchical fusion networks on MOSI and MOSEI datasets.

Input	MOSI				MOSEI			
	MAE (\downarrow)	Corr (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)	MAE (\downarrow)	Corr (\uparrow)	Acc-2 (\uparrow)	F1-score (\uparrow)
$f_a + f_{a_t} + f_{v_t}$	0.887	0.641	78.56	78.42	0.699	0.548	80.92	80.33
$f_v + f_{a_t} + f_{v_t}$	0.852	0.728	80.74	80.57	0.644	0.615	82.14	82.06
$f_t + f_{a_t} + f_{v_t}$	0.709	0.802	83.38	83.32	0.528	0.767	85.25	85.09

representations of each modality but also helps us clarify the need for representation interaction learning at a later stage.

In addition, we also conduct ablation experiments of the fusion strategy (as shown in Table 5). In this experiment, we do not combine the resulting text-based visual modality, text-based acoustic modality, and text modality. The settings are marked as “a” in Figure 7 and $MGHF_{w/o(pc)}$ in Table 5. At the same time, we replace Bi-GRU in the fusion network with Bi-LSTM neural network. This setting is marked as “b” in Figure 7 and $MGHF_{LSTM}$ in Table 5. As mentioned before (Section 3.4), only in this experiment does Bi-GRU achieves comparable or even better performance on some metrics.

As shown in the previous section (Section 5.2.1) the combined contribution of text-based sound representations f_{a_t} and text-based visual representations f_{v_t} is the highest. We used these two representations combined with the initial representations f_a , f_v , and f_t to evaluate which set performs best for the hierarchical fusion network (see Table 6). It is easy to see that the combination of f_{a_t} and f_{v_t} with the textual representation f_t , which is the input to our gated recurrent hierarchical fusion network, performs best.

6. Conclusions

In this paper, we propose a complete solution for multimodal sentiment analysis, MGHF, which differs in two main parts: modal representation and modal fusion. By using distribution matching in the representation learning phase, the neighbouring modalities are made to contain potential representations of the companion modalities to achieve modal information interaction in time series. Meanwhile, we design a gated recurrent hierarchical fusion network in the fusion phase through the intermodal representation interactions performed in the later fusion phase. It eliminates redundant modal representations and retains those valid for prediction in the final stage, making the prediction results closer to the actual scores. We show that our model is intensely competitive with previous complex baselines through extensive experiments on two publicly available datasets.

Data Availability

The data used to support the findings of this study are available on the following address. Dataset Address: https://immortal.multicomp.cs.cmu.edu/raw_datasets/processed_data/.

Conflicts of Interest

The author(s) declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Key Research and Development Program of China (Grant No. 2019YFB1404700).

References

- [1] S. Mai, H. Hu, and S. Xing, “Modality to modality translation: an adversarial representation learning and graph fusion network for multimodal fusion,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 164–172, Washington, DC, USA, April 2020.
- [2] W. Yu, X. Hua, F. Meng, and Y. Zhu, “Ch-sims: a Chinese multimodal sentiment analysis dataset with fine-grained annotation of modality,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.
- [3] D. K. Nugroho, “US presidential election 2020 prediction based on Twitter data using lexicon-based sentiment analysis,” in *Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 136–141, IEEE, Noida, India, January 2021.
- [4] S. Garg, “Drug recommendation system based on sentiment analysis of drug reviews using machine learning,” in *Proceedings of the 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 175–181, IEEE, Noida, India, January 2021.
- [5] D. S. Chauhan, M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, “Context-aware interactive attention for multi-modal sentiment and emotion analysis,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5647–5657, Hong Kong, China, November 2019.
- [6] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: a survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019.
- [7] S. Verma, C. Wang, L. Zhu, and W. Liu, “Deepcu: integrating both common and unique latent information for multimodal sentiment analysis,” in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 3627–3634, Vienna, Austria, August 2019.
- [8] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L. P. Morency, “Tensor Fusion Network for Multimodal Sentiment Analysis,” 2017, <https://arxiv.org/abs/1707.07250>.
- [9] Y. H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L. P. Morency, and R. Salakhutdinov, “Multimodal transformer for

- unaligned multimodal language sequences,” *Proceedings of the conference. Association for Computational Linguistics. Meeting*, vol. 2019, p. 6558, 2019.
- [10] D. Hazarika, R. Zimmermann, and S. Poria, “Misa: modality-invariant and-specific representations for multimodal sentiment analysis,” *Proceedings of the 28th ACM International Conference on Multimedia*, New York, NY, USA, 2020.
- [11] W. Yu, H. Xu, Z. Yuan, and J. Wu, “Learning modality-specific representations with self-supervised multi-task learning for multimodal sentiment analysis,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, Article ID 10790, 2021, May.
- [12] N. Q. Pham, J. Niehues, T. L. Ha, and A. Waibel, “Improving zero-shot translation with language-independent constraints,” *WMT*, vol. 13, 2019.
- [13] Z. Sun, P. Sarma, W. Sethares, and Y. Liang, “Learning relationships between text, audio, and video via deep canonical correlation for multimodal language analysis,” *AAAI Conference on Artificial Intelligence*, vol. 34, no. 5, pp. 8992–8999, 2020, April.
- [14] A. Zadeh, R. Zellers, E. Pincus, and L. P. Morency, “Multimodal sentiment intensity analysis in videos: facial gestures and verbal messages,” *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 82–88, 2016.
- [15] A. A. B. Zadeh, P. L. Paul, P. Soujanya, E. Cambria, and M. Louis-Philippe, “Multimodal language analysis in the wild: cmu-mosei dataset and interpretable dynamic fusion graph,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, vol. 1, Vancouver, Canada, 2018.
- [16] S. Schmidhuber and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [17] E. C. SoujanyaPoria, D. Hazarika, N. Majumder, Amir Zadeh, and L.-P. Morency, “Context-dependent sentiment analysis in user-generated videos,” vol. 1, pp. 873–883, in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, Association for Computational Linguistics, Vancouver, Canada, 2017.
- [18] V. Rozgić, S. Ananthakrishnan, S. Saleem, R. Kumar, A. N. Vembu, and R. Prasad, “Emotion recognition using acoustic and lexical features,” in *Proceedings of the Thirteenth Annual Conference of the International Speech Communication Association*, South Brisbane, QLD, Australia, April 2012.
- [19] A. Zadeh, P. P. Liang, N. Mazumder, S. Poria, E. Cambria, and L. P. Morency, “Memory fusion network for multi-view sequential learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018, April.
- [20] Z. Liu, Y. Shen, V. B. Lakshminarasimhan, P. P. Liang, A. Zadeh, and L. P. Morency, “Efficient low-rank multimodal fusion with modality-specific factors,” 2018, <https://arxiv.org/abs/1806.00064>.
- [21] W. Dai, S. Cahyawijaya, Y. Bang, and P. Fung, “Weakly-supervised Multi-Task Learning for Multimodal Affect Recognition,” 2021, <https://arxiv.org/abs/2104.11560>.
- [22] Y. Wang, Y. Shen, Z. Liu, P. P. Liang, A. Zadeh, and L.-P. Morency, “Words can shift: dynamically adjusting word representations using nonverbal behaviors,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 7216–7223, 2019.
- [23] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal fusion for multimedia analysis: a survey,” *Multimedia Systems*, vol. 16, no. 6, pp. 345–379, 2010.
- [24] X. Shu, J. Yang, R. Yan, and Y. Song, “Expansion-squeeze-excitation fusion network for elderly activity recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.
- [25] X. Shu, Q. Guo-Jun, J. Tang, and J. Wang, “Weekly-shared deep transfer networks for heterogeneous-domain knowledge propagation,” *ACM International Conference on Multimedia (ACM MM)*, Brisbane, Australia, 2015.
- [26] “Generalized deep transfer networks for heterogeneous-domain knowledge propagation,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 4s, 2016.
- [27] D. Kiela, S. Bhooshan, H. Firooz, E. Perez, and D. Testuggine, “Supervised Multimodal Bitransformers for Classifying Images and Text,” 2019, <https://arxiv.org/abs/1909.02950>.
- [28] C. Xi, G. Lu, and J. Yan, “Multimodal sentiment analysis based on multi-head attention mechanism,” in *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, pp. 34–39, Haiphong City, Viet Nam, January 2020.
- [29] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2018, <https://arxiv.org/abs/1810.04805>.
- [30] R. Aharoni and Y. Goldberg, “Unsupervised domain clusters in pretrained language models,” 2020, <https://arxiv.org/abs/2004.02105>.
- [31] A. Vaswani, N. Shazeer, N. Parmar et al., “Attention is all you need,” in *Advances in Neural Information Processing Systems* Long Beach, CA, USA, 2017.
- [32] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Doha, Qatar, 2014.
- [33] G. Degottex, J. Kane, T. Drugman, TuomoRaitio, and S. Scherer, “COVAREP—a collaborative voice analysis repository for speech technologies,” in *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (Icassp)*, pp. 960–964, IEEE, Florence, Italy, May 2014.
- [34] E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, Oxford University Press, Oxford, UK, 1997.
- [35] Y. H. H. Tsai, P. P. Liang, A. Zadeh, L. P. Morency, and R. Salakhutdinov, “Learning factorized multimodal representations,” in *Proceedings of the International Conference on Representation Learning*, New Orleans, LA, USA, February 2019.
- [36] W. Rahman, M. K. Hasan, S. Lee et al., “Integrating multimodal information in large pretrained transformers,” in *Proceedings of the conference. Association for Computational Linguistics*, vol. 2020, p. 2359p. 2359, July 2020.
- [37] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal deep learning,” pp. 689–696, 2011, https://icml.cc/2011/papers/399_icmlpaper.pdf.

Research Article

A Survival Status Classification Model for Osteosarcoma Patients Based on E-CNN-SVM and Multisource Data Fusion

Qiang Zhang , Peng Peng , and Yi Gu 

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 5 June 2022; Revised 27 June 2022; Accepted 28 June 2022; Published 9 July 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Qiang 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.

Traditional algorithms have the following drawbacks: (1) they only focus on a certain aspect of genetic data or local feature data of osteosarcoma patients, and the extracted feature information is not considered as a whole; (2) they do not equalize the sample data between categories; (3) the generalization ability of the model is weak, and it is difficult to perform the task of classifying the survival status of osteosarcoma patients better. In this context, this paper designs a survival status prediction model for osteosarcoma patients based on E-CNN-SVM and multisource data fusion, taking into full consideration the characteristics of the small number of samples, high dimensionality, and interclass imbalance of osteosarcoma patients' genetic data. The model fuses four gene sequencing data highly correlated with bone tumors using the random forest algorithm in a dimensionality reduction and then equalizes the data using a hybrid sampling method combining the SMOTE algorithm and the TomekLink algorithm; secondly, the CNN model with the incentive module is used to further extract features from the data for more accurate extraction of characteristic information; finally, the data are passed to the SVM model to further improve the stability and classification performance of the model. The model has been demonstrated to be more effective in improving the accuracy of the classification of patients with osteosarcoma.

1. Introduction

Cancer [1], also known as malignant tumors by the medical profession, is the result of abnormal cell growth. There are more than 100 types of cancer. Although the types and clinical characteristics of these cancers vary, they have the following commonalities: (1) they are detected late, and patients who are in the middle or late stages when diagnosed account for the majority of the overall cancer diagnosis; (2) the current cancer treatment methods include multiple excision of the malignant tumor, radiotherapy, and chemotherapy. This process will bring great pain to the patient. Osteosarcoma [2], as a type of cancer, has a low incidence rate of only 3 per million, or 2 per 1,000 of the total number of all cancers. However, the degree of malignancy is quite high and the patients are mostly children and adolescents [3]. The cell turnover and growth rate of these people are faster than others, so the cancer cells will grow faster too and the possibility of cure is lower. Early symptoms of

osteosarcoma are skeletal pain [4] and unexplained fractures, so they are misdiagnosed as growing pains caused by a body that is growing too fast, delaying the best chance of treatment, which in later stages can develop into severe pain and abnormalities in most organs of the body.

With the development of modern biomedicine, a very close link between osteosarcoma and genes has been discovered through a large number of investigations and studies [5], which makes it possible to treat patients with osteosarcoma precisely based on genetic data. Copy number variation [6] data refer to the reordering of the genome from its original base and can occur in small gene fragments ranging from 1 kb to large gene fragments of several MB. Single nucleotide polymorphisms (SNPs), deletions, insertions and duplications of gene fragments, and variation at multiple loci are the reasons why copy number variation occurs. Its mechanism is the mispairing of two highly homologous DNA sequences on the genome during meiosis or mitosis, resulting in the appearance of a sequence deletion or

duplication. The article “The impact of copy number variants on clinical staging and drug resistance mechanisms in osteosarcoma multiforme” suggests that copy number variation is one of several genetic abnormalities that manifest in osteosarcoma. DNA methylation [7] refers to the process of making gene expression different from the original result without modifying the original sequence order of the DNA. Under the influence of DNA methyltransferase, the base pair located in the DNA sequence changes the connection mode of S-adenosylmethionine from a double covalent bond to a single methyl group. The article titled “Research Progress on the Relationship between DNA Methylation and Osteosarcoma” shows that the methylation of various genes is closely related to the staging, diagnosis, metastasis, and prognosis of osteosarcoma. Compared to DNA sequencing data, RNA sequencing [8] data are more sensitive and accurate in detecting fusion genes, which are one of the driving genes for osteosarcoma. In summary, copy number variation data, DNA methylation data, RNA expression data, etc., are all very closely related to osteosarcoma. Their combined analysis is of great importance for the early diagnosis of osteosarcoma, the analysis of the severity of the disease, the establishment of prognostic models, and the assistance of doctors in accurate treatment.

With the development of next-generation gene-sequencing technology [9], the accuracy of the sequencing results has been greatly improved and the types of sequenced genes have become more comprehensive. The use of machine learning methods to establish a survival model and a prognosis model for patients with osteosarcoma can be used as a means of adjuvant therapy to help doctors formulate a more effective and targeted diagnosis and treatment plans and ultimately achieve the purpose of precise treatment [10]. However, the development of gene-sequencing technology also brings some problems to model training. The data dimension is much higher than before. However, there are not many genes that are highly related to diseases. This problem can be described as a dimensional disaster. How to select the genetic data highly related to this disease from the high-dimensional genetic data are a challenge in the current biomedical field. Manual operations and processing based on statistical methods will miss some key genes, and the final result will lead to an inaccurate analysis of the survival status of cancer patients.

With the development of artificial intelligence techniques [11] in the past decades, machine learning [12] as a branch of artificial intelligence has been extremely widely used in many fields, such as image processing, natural language processing, assisted medical diagnosis, and web recommendation systems. The use of machine learning techniques to analyze gene expression data has become one of the most important tools to achieve the goal of precision medicine. However, machine learning faces two challenges in processing gene expression data from osteosarcoma patients: (1) the high dimensionality of the data, which reaches over twenty-four thousand dimensions, is much larger than the number of samples. The redundant and unimportant gene features are removed and suppressed to motivate the important features, which are of considerable importance

for predicting the survival status of patients. (2) The amount of data is small, with less than eighty samples of data. In this context, if the model is too complex it will produce an overfitting phenomenon, showing a high classification accuracy in the training set but a low accuracy in the test set. (3) The amount of data between samples of different categories is extremely unbalanced [13], which ultimately leads to falsely high accuracy, low recall, and serious errors in identifying categories with a small number of samples.

Data dimensionality reduction [14] means eliminating unimportant features and retaining important features to improve data quality and enhance model classification. It can be divided into two main categories: feature selection and feature extraction. The former is based on prior knowledge, calculation algorithm results and statistical algorithm results, etc., to filter from original features, and select features that have a great impact on the results. For example, IL1B, IL1RN, IL8, IL10, IL17, and other genes have been confirmed to be highly correlated with gastric cancer [15]. When reducing the dimensionality of gastric cancer gene data, these gene features can be directly selected as the features after dimensionality reduction. The latter method of data dimensionality reduction is different from the former method. Feature extraction needs to change the feature space of the original data and combine different features according to the relationship between the features to obtain the features after dimensionality reduction. The main methods are principal component analysis [16] (hereinafter referred to as PCA) and so on. The PCA algorithm is a linear, global, unsupervised dimensionality reduction algorithm that is implemented by projecting the original data into a new coordinate system, followed by a projection into a k -dimensional space, with k being the number of features that ultimately need to be retained. The variance in the projection direction is used as a criterion, the larger the variance, the higher the dispersion range. Due to the characteristics of its algorithm, the number of features will be strictly smaller than the number of samples after the dimensionality reduction of the samples using the PCA algorithm. The number of samples is too small, which means that the PCA algorithm cannot obtain an effective projection space, which affects the dimensionality reduction effect and finally makes it difficult for the trained model to show strong classification ability. It can be seen that the rare disease data set represented by osteosarcoma is not suitable for dimensionality reduction using the spatial projection algorithm represented by the PCA algorithm. In this paper, the random forest algorithm is used to first reduce the dimensionality of copy number abnormal data, DNA methylation data, RNA gene sequencing data, and RNA homolog sequencing data individually and then fuse the data according to equal weights.

In recent years, with the development of convolutional neural network [17] (referred to as CNN), many problems that cannot be solved by human computing power alone can be solved with the help of CNN. CNN is a network inspired by the structure of the human neural network. Convolutional neural networks have a wide range of applications including but not limited to computer vision, natural language processing, medical diagnosis, and other fields.

Compared with the traditional neural network, the difference is that the convolutional neural network has a convolution layer and a downsampling layer, which can automatically perform feature extraction and transfer the feature-extracted data to the corresponding classifier for classification processing. However, CNN has certain limitations. It uses the convolution kernel to perform convolution calculation on the data, which can only obtain the local receptive field and does not consider the relationship between features from a global perspective. Inspired by SENet [18], the algorithm used in this paper adds an excitation module to the convolutional neural network. Recalculating the weights of different channel features, suppresses the influence of unimportant features and enhances the role of important features. By fully comparing the characteristics of different algorithms and the characteristics of the osteosarcoma gene dataset, this paper will use E-CNN to further extract the features after dimension reduction.

As a kind of machine learning, a support vector machine [19] (SVM for short) has better performance in small sample binary classification and stronger generalization ability compared with other machine learning algorithms. Different from the working principle of the CNN algorithm, SVM is aimed to find an optimal hyperplane between different categories through geometric calculation and uses the hyperplane as the boundary between the categories to complete the classification work. In addition, SVM can also map the data to a higher-dimensional space through kernel functions such as radial basis functions (RBF) and then solve the classification problem of linear inseparable data. As a classifier, SVM has a strong ability to perform binary classification of datasets with small data volumes and low dimensionality. The osteosarcoma gene sequencing data has a small amount of data, but its dimensionality is relatively high. The output data after initial feature selection by random forest [20], followed by deeper feature extraction using a pretrained E-CNN model, satisfies the characteristics of small data volume and low dimensionality described in the previous section. Therefore, this paper uses SVM to replace the Sigmoid activation function of the convolutional neural network as the final classifier, which can further improve the generalization ability and stability of the model.

Ho et al. proposed a model combining principal component analysis (PCA) and convolutional neural network (CNN) to predict the survival status of patients with gastric cancer by analyzing their RNA-seq data [21]; Liu et al. used a combination of a stacked autoencoder and a convolutional neural network model to expand the data using a stacked autoencoder, followed by the analysis of the patient's RNA-seq data using a CNN to classify the type of cancer [22]; Rukhsar et al. [23] used to convert the one-dimensional RNA-seq data into two-dimensional image data, perform zero-completion operations on the edge data and then expand the data using flip, rotate, etc. Then, commonly used networks for processing images, such as ResNet, were used to process the converted image data and thus achieve the function of classifying the survival status of cancer patients.

Copy number variation, DNA methylation data, RNA gene sequencing data, and RNA homologue sequencing data

all have varying degrees of impact on the prognosis of cancer patients, but the algorithms described previously only use RNA-seq data for classification studies and do not combine these data for analysis; secondly, these algorithms only use CNN or ResNet, which can only obtain local receptive field and do not consider the relationship between different channels and are prone to overfitting; then secondly, due to the computing mechanism of the PCA algorithm they use, the dimensionality of the data after dimensionality reduction is required to be strictly smaller than that of the sample data, which will inevitably lead to some key genetic data being eliminated. In addition to this, there is an imbalance of data between the different categories of the medical genetic dataset, and these algorithms do not equalize the dataset, which can lead to inflated and unreliable final results.

In this context, this paper proposes a survival status classification model for osteosarcoma patients based on E-CNN-SVM and multisource data fusion. In this paper, an extensive literature review was conducted to fully investigate the impact of copy number variation data, DNA methylation data, RNA gene sequencing data, and RNA homologue sequencing data on osteosarcoma patients, and the four types of data were each reduced in dimension using a feature selection algorithm called a random forest, and then the data were fused with equal weights in the form of data concatenation. Embedding an excitation module in the CNN model to calculate weights based on the importance of features before they are passed to the fully connected layer, to enhance the role of important and weaken the role of unimportant features. In terms of binary classification, SVM shows extremely strong classification performance for samples with a small sample size and low dimensionality. In this paper, we combine both E-CNN and SVM. SVM is trained with data after E-CNN feature extraction to further improve the classification ability and generalization ability of the model.

2. Data Acquisition

In this article, 65 patients with osteosarcoma were downloaded from the TargetOS database, a globally recognized and authoritative cancer database. These 65 cases include (1) clinical data on osteosarcoma patients over 3 years, whether the patient died, whether the disease recurred, and whether the cancer cells metastasized. (2) Copy number variation data for osteosarcoma patients, the number of its dimensions per patient is 60,447. (3) DNA methylation data for osteosarcoma patients. The number of its dimensions per patient is 385,292. (4) RNA gene sequencing data for osteosarcoma patients, the number of its dimensions per patient is 59,956. (5) RNA homozygous sequencing data for osteosarcoma patients, the number of its dimensions per patient is 201759. The clinical characteristics of the 65 patients are shown in Table 1.

3. Related Work

The algorithmic model devised in this paper fully considers the characteristics of the high-dimensional, unbalanced and

TABLE 1: Clinical characteristics of these patients.

Clinical features	
Number of cases	65
Age at illness (years)	15.09 ± 4.89
Gender	
Male	36
Women	29
Race	
White people	39
Black people	10
Asian	7
Unknown	9
Primary site	
Lower limbs	59
Pelvis	2
Upper limb	4
Survival time (days)	1339.31 ± 982.08

small number of gene sequence data sets of osteosarcoma patients, and fully investigates copy number variation (CNV for short), DNA methylation, RNA gene sequence (RNA-Seq-Gene for short), RNA isoform sequence (RNA-Seq-Ios for short) data characteristics, and their respective impact on the survival status of patients with osteosarcoma.

The flowchart of the algorithm is shown in Figure 1 and Table 2. Firstly, the random forest algorithm was used to reduce the dimensionality of the four different aspects of the gene sequencing dataset described in the previous section, followed by fusion of the data with equal weights and then normalization. In addition, the data are balanced by the combination of SMOTE algorithm and the TomekLink algorithm. The SMOTE algorithm uses the upsampling method to expand the samples of the category with a smaller number of samples so that the number of samples of the two categories is consistent; secondly, the TomekLink algorithm is used to clean overlapping samples and edge samples between classes to prevent overfitting during model training. The preprocessed data were divided into a training set and a test set in the ratio of 8:2. Due to the small amount of data, the validation set is not divided separately. The five-fold cross-validation method is used to divide the training data into five equal parts, and one of the data is used as the validation set in turn, and the remaining 4 data are used as the training set.

After initial dimensionality reduction, normalization, and partitioning of the data into test and validation sets, the data are secondly subjected to further feature extraction using a convolutional neural network incorporating an excitation module to further improve the quality of the data. Inspired by the SENet neural network, weights are calculated for each feature using the excitation module before the data are fed into the fully connected layer, multiplying the weights with the features and secondly adding them to the unweighted features. Feature weighting fusion through the stimulus module suppresses the effect of unimportant features and enhances the effect of important features. In this paper, we first use the processed data to train the E-CNN model for multiple rounds to optimize the model parameters

and save the E-CNN model with the best classification results. When training the E-CNN model, the model is prone to overfitting due to the small number of osteosarcoma samples. In this paper, the early stop algorithm is used to avoid this problem during training, and the model parameters with the smallest rounds of the validation set error function value are saved as the final model parameters when parameter iterations are performed. After the E-CNN model is trained, deeper feature extraction is performed on the data using its input layer to the fully connected layer. The data that has been feature extracted by the E-CNN model is passed into the SVM to train the SVM model, followed by classification using the trained SVM model.

3.1. Random Forest. Random Forest is one of many high-performance machine learning algorithms. Because of its high classification accuracy, not easy to overfit, strong antinoise ability, fast training speed, and high tolerance to outliers in data sequences, it has a wide range of applications in economics, geography, and medicine. The Random Forest algorithm is an integrated learning algorithm consisting of several decision trees, based on the Bootstrap sampling method, which selects M batches of data from unprocessed source data multiple times, randomly and with putbacks, and uses these M batches to build M decision trees. A single decision tree is built as an embodiment of the greedy algorithm. First, the impurity of each feature is calculated using the Gini coefficient (as shown in (1)) or the information entropy (as shown in (2)) as an evaluation metric, the lower the impurity, the more suitable the feature is as a branch node in the decision tree. Each time, the node with the lowest impurity is selected from the unselected features, and the above-given operation is repeated several times in the remaining nodes until the Gini coefficient is less than a threshold or there are no features to continue building the decision tree.

$$\text{Gini} = 1 - \sum_{k=1}^N P_k^2 \sqrt{b^2 - 4ac}. \quad (1)$$

Here, P_k is the proportion of the k -th category of samples to the total number of samples N .

$$H(D) = - \sum_{k=1}^c p_k \log_2 p_k. \quad (2)$$

Here, D is the data set, c is the total number of categories, p_k is the number of samples in a category, and k is a proportion of the total number of samples.

For classification, the random forest uses each batch of data to train a classification decision tree and takes the output of the M classification trees with the largest weight as the classification result of the random forest. For the regression task, the random forest uses each batch of data to train a regression tree, and the output of these M regression trees is summed and divided by M as the final output of the entire random forest. Compared with a single model or one of the parametric models, the classification and prediction

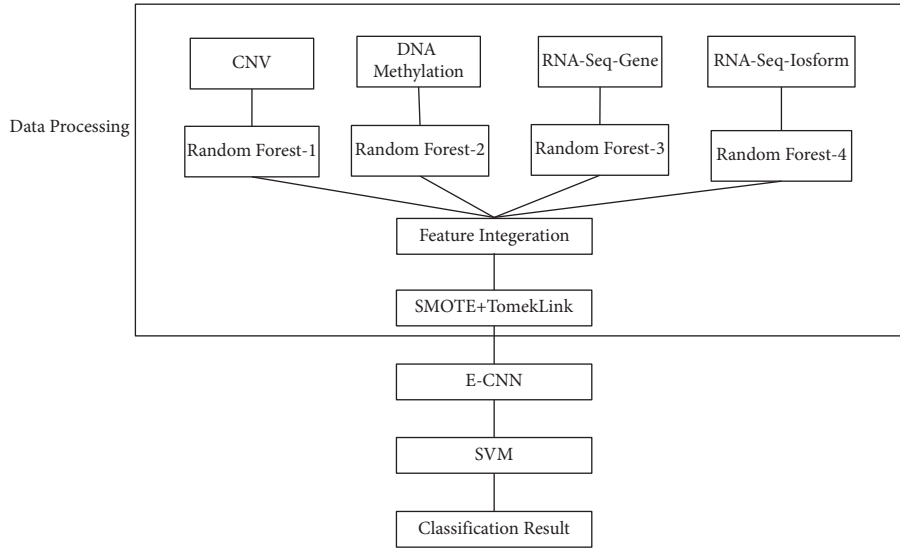


FIGURE 1: The flow chart of the E-CNN-SVM and multisource algorithm.

TABLE 2: The steps of the E-CNN-SVM and multisource algorithm.

Step 1	The copy number variation data, DNA methylation data, RNA gene sequencing data, and RNA homologue sequencing data are each reduced in dimension using the random forest algorithm
Step 2	Equal weighted fusion of these four types of data
Step 3	Combine the SMOTE algorithm with the TomekLink algorithm to clean and equalize the data
Step 4	Pretrain the E-CNN model and save the optimal model
Step 5	Feature extraction of data using the input layer to the fully connected layer of the E-CNN model
Step 6	Use the processed data to train the SVM model and use the trained model for classification

capabilities of the random forest are significantly improved, overcoming the problems of overfitting and low accuracy of a single decision tree, as shown in Figure 2.

Currently, the most widely used algorithm for dimensionality reduction is the PCA algorithm, but due to the limitations of the algorithm, the number of features retained after dimensionality reduction must be strictly smaller than the number of samples. Therefore, when the PCA algorithm reduces the dimensionality of the data with a small number of samples, although some noises are removed, some important features are also removed at the same time. It can be seen that the small sample data set represented by osteosarcoma is not suitable for using this algorithm as a dimensionality reduction algorithm. When performing classification and regression tasks, the random forest will determine the importance of each feature based on its contribution to the building of these decision trees. The features with the highest contribution are selected and retained after feature extraction. Because of the properties, it exhibits in dimensionality reduction, the random forest algorithm is used as the dimensionality reduction algorithm in this paper.

3.2. SMOTE+ TomekLink. Machine learning is a data-based discipline. When performing classification tasks, where the amount of sample data varies significantly between categories, the classification performance of the trained model

can be severely affected, resulting in its false accuracy and low recall. There are two methods for resolving data imbalance: oversampling and undersampling, with the SMOTE algorithm [24] belonging to the former and the TomekLink [25] algorithm to the latter. The algorithm flow of the SMOTE algorithm is shown in Figure 3. The SMOTE algorithm is used to generate new samples for classes with relatively small sample sizes and put them back into the dataset. In this way, the classes with smaller sample sizes are expanded, and thus the data is balanced between the samples of different classes. The algorithmic process is as follows:

- (1) Using the Euclidean distance as a criterion, draw m samples from a relatively small number of categories, and denote these m samples as $\{X_1, X_2, X_3, X_4, X_5, \dots, X_{m-1}, X_m\}$
- (2) Based on the idea of the nearest neighbor node algorithm, for each node $X_i \in \{X_1, X_2, X_3, X_4, X_5, \dots, X_{m-1}, X_m\}$, compute the k nearest neighbor nodes of each sample X_i , denoted as $X_{i(pq)}$
- (3) Select several samples from X_{ik} , denote them as $X_{i(pq)}$ and generate a random number c , where $c > 0$ and $c < 1$, to synthesize a new sample according to the following equation:

$$X_{\text{new}} = X_i + \text{rand}(0, 1) * (X_{i(pq)} - X_i). \quad (3)$$

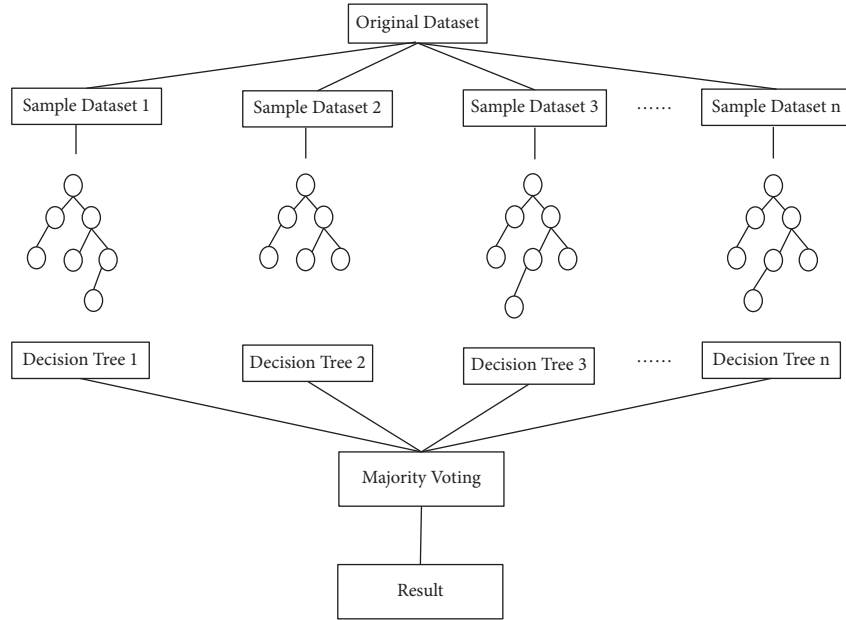


FIGURE 2: The random forest construction process.

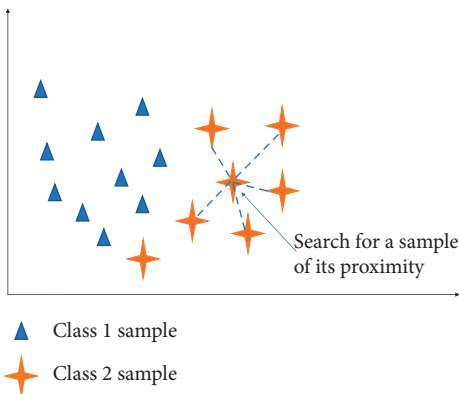


FIGURE 3: Schematic diagram of the SMOTE algorithm for constructing new nodes.

- (4) Place the newly synthesized new sample X_{new} into the original dataset

However, using SMOTE as an oversampling technique to equalize the data can lead to the following problems: (1) SMOTE only focuses on the category with a small number of samples but does not take into account the relationship between the two categories, which can lead to the new composite samples crossing the category boundaries. (2) The new samples are generated by merging the samples at the original category boundary as the center, which will blur the category boundary and reduce the quality of the data and eventually lead to the classification model not being able to calculate an appropriate hyperplane. In this paper, we use TomekLinks, an undersampling method, to solve these problems caused by the use of a single SMOTE. The idea of the TomekLink algorithm is as follows: select some sample pairs (X_i, X_j) , where X_i is the category with a larger number of samples and X_j is the category with a smaller number of

samples, and note that, the distance between the two samples is $d(X_i, X_j)$, if there is no X_k , satisfying $d(X_i, X_k) < d(X_i, X_j)$, then (X_i, X_j) is a pair of TomekLinks pair. When performing data cleaning, both pairs are removed from the sample data.

3.3. E-CNN

3.3.1. Introduction to CNN. A convolutional neural network (CNN) is a feed-forward neural network that simulates the way human nerves work and was designed by HUB et al. CNN evolved from multilayer perceptrons (MLP for short). Compared with MLPs, CNNs have features such as weight sharing and local sampling. In recent years, CNN has been used extensively in the fields of image, sound, and medicine. The core parts of the CNN network involved in this article are described as follows:

- (1) *Convolutional Layer.* The convolution layer, the most important part of the entire CNN, uses convolutional kernels (also known as filters) to perform convolutional calculations on the data, a process that involves the extraction of features from the input data. The weights of each kernel do not change during the convolutional computation, which is the weight-sharing property of CNN. The result of the convolutional computation is then output to the next layer through a nonlinear activation function. In this paper, the ReLu activation function is used.
- (2) *Dropout Layer.* During the training of a neural network, overfitting can occur due to the depth of the network or the small amount of data. The result is locally optimal, but not for the whole. The result is that the model has poor classification performance and does not perform well in the classification task. The Dropout layer randomly discards some weights

in each training batch according to preset parameters, which can effectively avoid overfitting and improve the robustness of the model.

- (3) *Fully Connected Layer*. Each nerve in the fully connected layer is connected to a neuron in the previous layer, and in this way, the features output from the previous layer are aggregated.

In this paper, a shallow convolutional neural network was designed, which mainly consisted of an input layer, two convolutional layers, two Dropout layers, and two fully connected layers, taking into full consideration the characteristics of the small number and high dimensionality of the osteosarcoma gene data after equal weight fusion. The CNN model designed in this paper is shown in Figure 4(a).

3.3.2. Introduction to Squeeze-and-Excitation Network. The CNN algorithm has the advantages of weight sharing and automatic feature extraction, but it uses convolutional kernels to obtain only local perceptual fields and does not take into account the correlation between the overall features. The core module of SENet is the Squeeze-and-Excitation (SE) module, which uses the degree of influence of a feature on the result as a criterion to recalculate the weights of different channel features to strengthen the effect of features with high influence and weaken the effect of features with low influence. This module is not a complete network structure, but a module for processing data that can be embedded in other network structures. The SE module algorithm flow is as follows:

- (1) *Squeeze (Sequence)*: aggregates data within the same channel to form a multidimensional statistic containing only interchannel correlations. This is achieved by global average pooling (GAP). For example, the original data U has dimension $H \times W$ and is compressed along its spatial dimension to give the statistic $Z \in R^c$, by which the i -th element of the compressed Z_i can be obtained.

$$Z_i = F_{sq}(u_i) = \frac{1}{H \times W} \sum_{p=1}^H \sum_{q=1}^W u_i(p, q), \quad (4)$$

where $u_i(p, q)$ are the p row and q column elements passed into the SE module.

- (2) *Excitation*: the excitation process is achieved through two fully connected layers. The role of the first fully connected layer is to compress the vector from X dimensional to X/C dimensional, followed by a call to the ReLU activation function; the role of the second fully connected layer is to restore the vector to X dimensions, followed by a call to the Sigmoid activation function, whose calculation process is shown in the following equation:

$$s = F_{ex}(z, W) = \partial(g(z, W)) = \partial(W_2 \delta(W_1 z)). \quad (5)$$

The sigmoid activation function is denoted as $\partial(\cdot)$, the ReLU activation function expression is denoted as $\delta(\cdot)$, the parameters of the two fully connected layers are denoted as $W_1 \in R^{(C/r) \times C}$ and $W_2 \in R^{C \times (C/r)}$, the value of r is determined by the model performance requirements and the computational complexity, and the generated s are the weights of the channels.

- (3) *Scale*: the weights S represent the importance of each feature through the excitation process. The final output X_{new} is obtained by multiplying the weights S and the unweighted data U after convolution, as shown in equations (6) and (7).

$$X_c = F_{\text{scale}}(u_c, s_c) = s_c u_c, \quad (6)$$

$$X_{\text{new}} = [X_{1\text{new}}, X_{2\text{new}}, \dots, X_{i\text{new}}, \dots, X_{c\text{new}}]. \quad (7)$$

After the SE module calculation, the weight relationship between the channels after the CNN convolution calculation can be obtained. By multiplying the unweighted data with the weights, the performance of the features with high importance is amplified and the performance of the unimportant features is suppressed. The result is a model that can be effectively enhanced without adding much computing power.

3.3.3. Introduction to the E-CNN Model. The main application of SENet is in the field of image classification, where the image data needs to be flattened before recalculating the individual feature weights, for example by compressing the original spatial dimension of $H \times W$ to $1 \times 1 \times C$. The sequencing data of the osteosarcoma gene after equal-weight fusion is itself flattened data, so there is no need to compress the original data before performing the stimulation operation. Inspired by the SE module, an E-CNN neural network model was designed to further improve the classification performance of the model by considering the relationship between each gene feature as a whole and efficiently extracting features from the sequencing data. The E-CNN neural network and CNN neural network are shown in Figure 4.

3.4. SVM. Support Vector Machine (SVM) is a high-performance supervised machine learning algorithm with stable classification performance, high interpretability, and generalization ability. The SVM works as follows: let the data set be T , the hyperplane be M , and the minimum geometric distance between T and M be r . When the value r is maximum, then this hyperplane is the optimal hyperplane, and the optimal hyperplane is used as the criterion for classification.

Because of its good results in small sample binary classification, the RBF kernel-based SVM is used as a classifier in this paper. The parameter setting of the SVM affects the heaviest classification performance, and the penalty factor C directly affects the choice of the

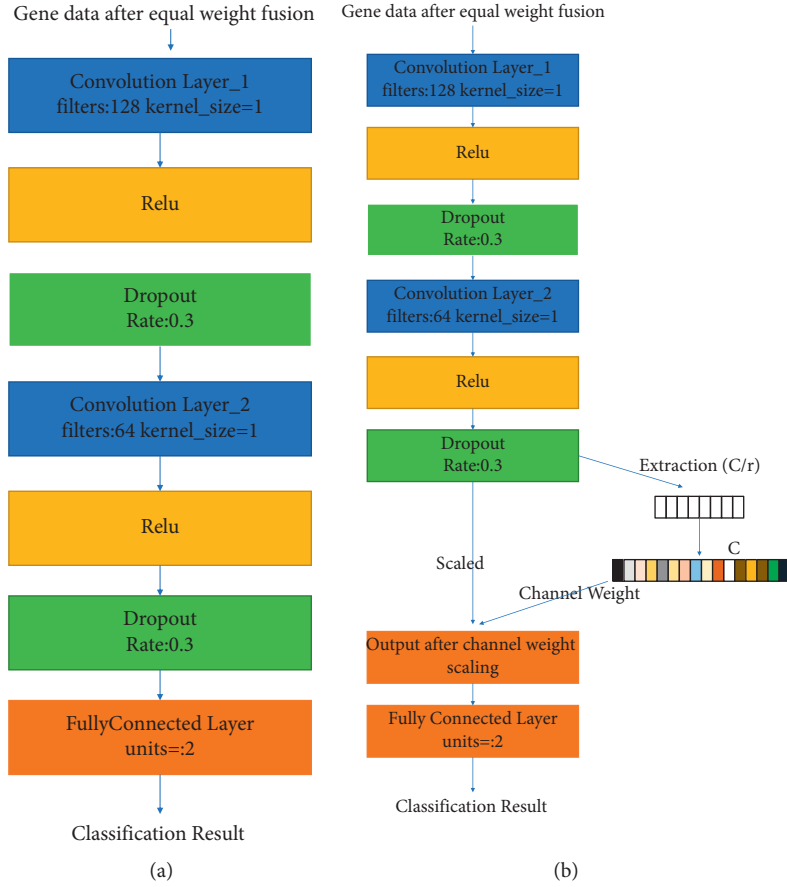


FIGURE 4: The schematic diagram of CNN & E-CNN. (a) The schematic diagram of the CNN model. (b) The schematic diagram of the E-CNN model.

classification boundary. The RBF kernel function is shown in the following equation:

$$K(x, x_i) = e^{-\|x - x_i\|^2 / 2\delta}. \quad (8)$$

In equation (8), δ is the width of the RBF; $\|x - x_i\|$ is the distance between the selected point and the centroid.

4. Analysis of Experimental Results

4.1. Experimental Environment. All experiments in this article are run on a computer with Intel Xeon(R) CPU E5-2640 v4@2.40 GHz * 40 and 64-bit Linux operating system. The compilation software was PyCharm (PyCharm Community Edition 2020.2.2 x64).

4.2. Model Parameter Settings. The parameters of the E-CNN-SVM model proposed in this paper are shown in Table 3.

4.3. Evaluation Indicators. In the binary classification problem, as shown in Table 4, if the original category of data is death, the number of data correctly predicted to be in the category of death is labeled TP and the number of

data incorrectly predicted to be in the category of survival is denoted FN; if the original category of data is survival, the number of data correctly predicted to be in the category of survival is denoted TN and the number of data incorrectly predicted to be in the category of death is denoted FP. To show the advantages of the algorithm in this paper, it will be compared with other algorithms in terms of accuracy (ACC), recall (Recall), $F1$ score($F1$), and variance (variance).

Accuracy is a measure of the classifier's classification performance; the higher the accuracy, the better the classification performance of the classifier, and is calculated as shown in the following equation:

$$ACC = \frac{TP + TN}{FP + FN + TP + TN}. \quad (9)$$

The recall is how many positive samples are correctly identified and shows the find-all rate of the classifier. The higher the recall, the more accurately the classifier identifies positive example samples, calculated as shown in the following equation:

$$Recall = \frac{TP}{TP + FN}. \quad (10)$$

TABLE 3: The E-CNN-SVM model parameters.

CNN			
First convolutional layer	Filters: 128		kernel_size: 1
First dropout layer		Rate: 0.3	
Second convolutional layer	Filters: 64		kernel_size: 1
Second dropout layer		Rate: 0.3	
Motivation module		Compression ratio: 4	
Fully connected layer		Units: 2	
SVM			
Penalty factor		0.9	
Cache_size		3000	
Kernel		rbf	

TABLE 4: Confusion matrix.

Patient survival status	Classification predicted as death	Category prediction for survival
Death	TP	FN
Survival	FP	TN

The precision rate is the number of samples correctly detected as dead as a proportion of all samples predicted as dead by the model. The $F1$ score takes into account both the precision and recall of the classifier, with the precision rate calculated as shown in equation (11) and the $F1$ score calculated as shown in equation (12).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (11)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (12)$$

The variance represents the stability of the model; a smaller value of variance means a more stable model; conversely, a less stable model. The variance calculation formula is shown in the following equation:

$$\text{Variance} = \frac{1}{n} \sum_{i=1}^n (x - x_i)^2. \quad (13)$$

In equation (13), n is the number of experiments, x is the average of the accuracy of the n rounds of experiments, and x_i is the accuracy of the i -th round of experiments.

4.4. Comparison of Evaluation Results. All experiments performed in this paper used a subset of the same osteosarcoma gene sequencing dataset, and the data were processed consistently. Firstly, the random forest algorithm is used for feature selection to reduce the dimensionality of the data, and secondly, the SMOTE algorithm is combined with the TomekLink algorithm to equalize the amount of data between the different categories and to clean the edge data. A random selection of 20% of the data was used as the test set, followed by a five-fold cross-validation approach using the remaining data, with four of the data rotated as the training set and one as the validation set, to avoid overfitting during the training process.

To validate the performance of the proposed E-CNN-SVM multisource data fusion-based survival status classification model for osteosarcoma patients in this paper, the experiments are designed from the following two perspectives as well as comparative experiments: (1) under the same algorithm, the classification results of the model trained with single-source data are compared with those of the model trained with multisource fused data. (2) The algorithm adopted in this paper is compared with the rest of the algorithms under the condition that the models are all trained using multisource fused data.

4.4.1. Data Source Comparison Experiments. E-CNN-SVM, E-CNN, and CNN were used as the experimental models for this subsection. First, the models were trained and tested using four single sources of data: copy number variation data, DNA methylation data, RNA gene sequencing data, and RNA homologous isoform sequencing data, respectively. Secondly, the data fused with equal weights from these 4 data were used for training and testing. The test results of the same model trained with different source data were compared. The experimental results of the E-CNN-SVM model are shown in Table 5, the experimental results of the E-CNN model are shown in Table 6, and the experimental results of the CNN model are shown in Table 7.

From the experimental results in Tables 5–7, it can be seen that the multi-source fusion data integrated copy number variation data, DNA methylation data, and RNA gene sequencing data related to osteosarcoma, resulting in an improvement of the individual models in all directions. The accuracy, recall, $F1$ score, and variance of the same model after training with multiple sources of data were all better than those after training with single-source data.

4.4.2. Model Comparison Experiments. The results of the data source comparison experiments in the previous section have shown that the ability of the models trained using multiple sources of fused data is better than that of the models trained using single-source data. Therefore, in the model comparison experiments, the models were trained using multisource fused data. To illustrate the effectiveness and feasibility of the E-CNN-SVM model used in this paper, it was chosen to be compared with the existing paper models E-CNN model, CNN model, SVM

TABLE 5: E-CNN-SVM experimental data.

Number	Model	Data type	Accuracy (%)	Recall (%)	F1 score (%)	Variance
1	E-CNN-SVM	Multisource	100	100	100	0
2	E-CNN-SVM	Copy number variation	100	100	100	0
3	E-CNN-SVM	DNA methylation	100	100	100	0
4	E-CNN-SVM	RNA-seq-gene	100	100	100	0
5	E-CNN-SVM	RNA-seq-Ios	92	75	86	0

TABLE 6: E-CNN experimental data.

Number	Model	Data type	Accuracy (%)	Recall rate (%)	F1 score (%)	Variance
1	E-CNN	Multisource	97	97	96	0.0015
2	E-CNN	Copy number variation	79	69	69	0.0036
3	E-CNN	DNA methylation	86	76	77	0.0031
4	E-CNN	RNA-seq-gene	82	71	78	0.0105
5	E-CNN	RNA-seq-Ios	88	88	84	0.0012

TABLE 7: CNN experimental data.

Number	Model	Data type	Accuracy (%)	Recall rate (%)	F1 score (%)	Variance
1	CNN	Multisource	86	97	81	0.0031
2	CNN	Copy number variation	75	69	66	0.0036
3	CNN	DNA methylation	80	74	74	0.0041
4	CNN	RNA-seq-gene	79	71	73	0.0036
5	CNN	RNA-seq-Ios	83	75	80	0.0054

TABLE 8: Model comparison experimental data.

Number	Model	Data type	Accuracy (%)	Recall rate (%)	F1 score (%)	Variance
1	E-CNN-SVM	Multisource	100	100	100	0.0000
2	E-CNN	Multisource	97	97	96	0.0015
3	CNN	Multisource	86	97	81	0.0031
4	SVM	Multisource	76	35	51	0.0081
5	XGBoost	Multisource	76	66	64	0.0065
6	CNN-LSTM	Multisource	59	0	0	0.0000

model, XGBoost model, and CNN-LSTM model. The experimental results are shown in Table 8.

From the experimental data in Table 8, we can learn that the E-CNN-SVM model proposed in this paper obtains a better result in terms of accuracy, recall, *F1* score, and variance. Its performance is much higher than the comparison model in this paper, showing the high performance, robustness, and stability of the E-CNN-SVM model rejected in this paper.

5. Conclusion

- (1) An E-CNN-SVM classification algorithm based on multisource feature homofusion is proposed. The copy number abnormal data, DNA methylation data, RNA gene sequencing data, and RNA homologue sequencing data of osteosarcoma patients were weighted and fused to predict the survival status of patients for classification. The accuracy, recall, and *F1* score of the predictive classification reached 100%, which is significantly more accurate than other methods for predicting the survival status of

osteosarcoma patients and is extremely important for achieving precise treatment of osteosarcoma patients.

- (2) The features of each channel are weighted by the calculation of the stimulus module before being fed into the fully connected layer of the CNN model, which can significantly improve the feature extraction capability of the CNN model.
- (3) The equal-weight fused data is again subjected to deeper feature extraction with the parameter-optimized E-CNN model and then fed into SVM for classification, which has higher accuracy than simply using E-CNN and SVM.
- (4) Although the algorithm in this paper exhibits strong properties, its generalization capability cannot be validated due to the relatively small dataset, which is a pain point in the whole field of medical analysis. The dataset will certainly be expanded from different aspects in the future to enhance the generalization ability of the model.

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 declare that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant no. 62171203 in part by the Natural Science Foundation of Jiangsu Province under Grant no. BK20210449, in part by the Science and Technology Demonstration Project of Social Development of Jiangsu Province under Grant BE2019631, and in part by the open project fund of Key Laboratory of Image Processing and Intelligent Control (Huazhong University of science and technology), Ministry of Education.

References

- [1] G. P. Dunn, L. J. Old, and R. D. Schreiber, "The immunobiology of cancer immunosurveillance and immunoediting," *Immunity*, vol. 21, no. 2, pp. 137–148, 2004.
- [2] G. Ottaviani and N. Jaffe, "The epidemiology of osteosarcoma," *Cancer Treatment and Research*, vol. 152, pp. 3–13, 2009.
- [3] M. Trivedi, C. S. Guruprasad, P. Thankamony, S. Sugath, J. Raj, and J. Kattoor, "Langerhans cell histiocytosis and osteosarcoma in children: a radiological mimic," *Indian Pediatrics*, vol. 59, no. 3, pp. 255–256, 2022.
- [4] X. Mao, Y. Jin, T. Feng et al., "Ginsenoside Rg3 inhibits the growth of osteosarcoma and attenuates metastasis through the wnt/ β -catenin and EMT signaling pathway," *Evidence-based Complementary and Alternative Medicine*, vol. 2020, no. 1, pp. 1–12, 2020.
- [5] M. Cortini, S. Avnet, and N. Baldini, "Mesenchymal stroma: role in osteosarcoma progression," *Cancer Letters*, vol. 405, pp. 90–99, 2017.
- [6] B. A. Walker, P. E. Leone, L. Chiecchio et al., "A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value," *Blood*, vol. 116, no. 15, pp. e56–e65, 2010.
- [7] R. Singal and G. D. Ginder, "DNA methylation," *Blood*, vol. 93, no. 12, pp. 4059–4070, 1999.
- [8] M. Li, H. Chen, Y. Zhao, S. Gao, and C. Cheng, "Retracted: H19 functions as a ceRNA in promoting metastasis through decreasing miR-200s activity in osteosarcoma," *DNA and Cell Biology*, vol. 35, no. 5, pp. 235–240, 2016.
- [9] F. Tong, J. Wang, R. Xiao, W. Bing-Bing, Z. Chao-Chun, and W. Ding-Wen, "Application of next generation sequencing in the screening of monogenic diseases in China, 2021: a consensus among Chinese newborn screening experts [J]," *World Journal of Pediatrics*, vol. 18, pp. 1–8, 2022.
- [10] K. Choucair, B. I. Mattar, Q. Van Truong et al., "Liquid biopsy-based precision therapy in patients with advanced solid tumors: a real-world experience from a community-based oncology practice," *The Oncologist*, vol. 27, no. 3, pp. 183–190, 2022.
- [11] Z. H. Chen, L. Lin, C. F. Wu, X. Rui-Hua, and S. Ying, "Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine," *Cancer Communications*, vol. 41, no. 11, pp. 1100–1115, 2021.
- [12] M. May, "Eight ways machine learning is assisting medicine [J]," *Natura Med*, vol. 27, no. 1, pp. 2–3, 2021.
- [13] M. Wasikowski and X. Chen, "Combating the small sample class imbalance problem using feature selection [J]," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1388–1400, 2009.
- [14] W. Zhao, H. Li, Y. Hao et al., "An efficient functional magnetic resonance imaging data reduction strategy using neighborhood preserving embedding algorithm," *Human Brain Mapping*, vol. 43, no. 5, pp. 1561–1576, 2022.
- [15] Y. Y. Janjigian, M. Cecchini, K. Shitara, and P. C. Enzinger, "1416P Genomic landscape of late-stage gastric cancer [J]," *Annals of Oncology*, vol. 32, pp. S1062–S1063, 2021.
- [16] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [17] S. Ö Arık, H. Jun, and G. Diamos, "Fast spectrogram inversion using multi-head convolutional neural networks [J]," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 94–98, 2018.
- [18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, Salt Lake City, UT, USA, June 2018.
- [19] M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," in *Proceedings of the 2016 international conference on computing communication control and automation (ICCCUBEA)*, pp. 1–7, IEEE, Pune, India, August 2016.
- [20] S. Kumano and T. Akutsu, "Comparison of the representational power of random forests, binary decision diagrams, and neural networks," *Neural Computation*, vol. 34, no. 4, pp. 1019–1044, 2022.
- [21] H. S. Shon, Y. G. Yi, K. O. Kim, E.-J. Cha, and K.-A. Kim, "Classification of stomach cancer gene expression data using CNN algorithm of deep learning," *Journal of Biomedical Translational Research*, vol. 20, no. 1, pp. 15–20, 2019.
- [22] J. Liu, X. Wang, Y. Cheng, and L. Zhang, "Tumor gene expression data classification via sample expansion-based deep learning," *Oncotarget*, vol. 8, no. 65, pp. 109646–109660, 2017.
- [23] L. Rukhsar, W. H. Bangyal, M. S. Ali Khan, A. A. Ag Ibrahim, K. Nisar, and D. B. Rawat, "Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification," *Applied Sciences*, vol. 12, no. 4, pp. 1850–1867, 2022.
- [24] 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, pp. 321–357, 2002.
- [25] D. Devi, B. Biswas, and B. Purkayastha, "Redundancy-driven modified Tomek-link based undersampling: a solution to class imbalance," *Pattern Recognition Letters*, vol. 93, pp. 3–12, 2017.

Research Article

A Fast CS-Based Reconstruction Model with Total Variation Constraint for MRI Enhancement in K-Space Domain

Hongxuan Duan ¹ and Xiaochang Lv²

¹Northeast Petroleum University, Daqing 163318, China

²Northeast Petroleum University of Qinhuangdao, Qinhuangdao 066004, China

Correspondence should be addressed to Hongxuan Duan; duanhx@nepu.edu.cn

Received 27 May 2022; Revised 20 June 2022; Accepted 21 June 2022; Published 6 July 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Hongxuan Duan and Xiaochang Lv. 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 fact that Magnetic Resonance Imaging (MRI) is still a relatively slow imaging modality, its application for dynamic imaging is restricted. The total variation is introduced into the CS-based MRI reconstruction model, and three regularization conditions are adopted to ensure that a high-quality reconstructed image is produced. In this paper, a simple yet fast CS-based optimization model for noisy MRI Enhancement is proposed. The alternative direction multiplier method is chosen to optimize the model, and the k -terms power series is applied in order to derive the LogDet function into the augmented Lagrange form. Following this, an approximation of the feature vector is achieved through the iterative process. The quality of the reconstructed image was much better than that of the CS-based MRI image reconstruction algorithm, as shown by experimental results under different noise conditions. The peak signal-to-noise ratio of the reconstructed image was able to be improved anywhere from 5 to 20 percent.

1. Introduction

Since the excited hydrogen nucleus will release its energy by emitting a signal with a specific frequency, the signal can be detected and reconstructed by corresponding MRI technical so as to recover the internal image data, which makes it possible for the rapid examination and reasonable diagnosis of diseases [1]. However, MRI will invariably have issues with noise, and the reduction in the number of phase-encoded signals will result in truncation artifacts when the truncated K-space data is Fourier transformed to reconstruct the image. This will have a direct impact on how accurately the image is analyzed, as well as the results of any subsequent medical diagnosis. It is of the utmost importance to determine how to process the noisy MRI enhancement in a reasonable and accurate manner [2].

The essence of MRI enhancement is that the detailed information of the image is fully reserved, and the irrelevant information such as noise and artifacts is removed to the greatest extent. At present, most MR image enhancement methods assume that the interference signal obeys Gaussian

distribution and many effective enhancement algorithms are proposed. The enhancement algorithms for MRI images can, as of right now, be broken down into three distinct categories. The first type is known as the variational-based method, and it is distinguished by the fact that it realizes MRI enhancement by finding the numerical solution to a particular partial differential equation. Some examples of this type of method include the adaptive anisotropic diffusion enhancement method, the fractional total variation enhancement model with L1 fidelity, and the minimum unbiased risk estimation for MRI enhancement. In order to find a solution to the issue of MR image enhancement, Chen et al. utilized the total variation regularization (TV) model. The ability of total variation mode of effectively eliminating random noise while simultaneously preserving the image's edges is the primary benefit of utilizing this mode [3–5]. On the other hand, in order to use the total-variation method, MR images need to fulfill the requirements of piecewise constancy. This requirement cannot be satisfied in the actual MRI system because there is a nonuniformity between the high field system and the excitation B1 field of 3T. In

addition, the enhancement method that is based on total variation will, under the assumption of piecewise invariance, result in the emergence of an artificial effect, which will cause the reconstruction result to be artificial. This is because the assumption states that piecewise invariance exists. The name of the second approach is the transform domain method. Its most distinguishing characteristic is the ability of eliminating interference in the transform domain and then performing an inverse transform in order to obtain an image of superior quality using techniques such as the wavelet threshold transformation method, different shrinkage criteria methods of wavelet transform, and the wavelet transform approach for MRI enhancement. According to the research in [6], the MR image can be improved and denoised through the use of bilateral filtering in the wavelet domain. However, the improved result will show pseudo-Gibbs phenomenon at the point where the signal discontinues. The third one is referred to as the NLM enhancement method, and it is typically distinguished by the application of a significant quantity of spatially redundant information. Initially, the NLM algorithm was utilized primarily for the purpose of addressing Gaussian distributed noise. It is able to reserve more details and remove more noise than general methods because it has obvious advantages over those methods. In recent years, NLM and its improved algorithm have been applied to MR images and have achieved a good enhancement effect. Some examples of this include nonlocal enhancement based on the Gaussian model, maximum likelihood nonlocal estimation method, and nonlocal enhancement based on unbiased estimation [7, 8].

Although high-quality MRI images can be obtained by enhancement algorithms, most of these algorithms still analyze and process the images and do not make use of the essential features of MRI raw data. It is well known that the generation of MRI requires three main steps: (1) the protons in the imaging area generate signals (FID signal, SE signal, STE signal, etc.) through the cooperation of RF pulse and gradient magnetic field; (2) the MRT/R coil is adopted to acquire these signals and fill the acquired signals into the K-space; (3) Fourier transform is performed on the data in K-space to obtain a magnetic resonance image. Therefore, the K-space is the space for storing the original data of magnetic resonance, and the magnetic resonance image can be obtained by performing very complex data post-processing on the original data of the K-space. Therefore, some scholars began to explore the data processing in K-space so as to improve the signal quality in the imaging process [9].

In an MRI system, image signals are recorded sequentially in K-space. The scanning efficiency is the most important factor in determining how quickly images can be acquired in this system. Because of the MRI system's slow data acquisition speed, motion artifacts and noise are both relatively simple to generate. According to CS theory, if a signal is compressible or sparse in a transform domain, the high-dimensional signal can be mapped into the low-dimensional space using an observation matrix unrelated to the transform basis, and the original signal can be reconstructed with a high probability from a small number of

projections by solving the optimization problem. This is done by mapping the high-dimensional signal into low-dimensional space. The application of CS theory should make it possible to drastically cut down on the amount of data acquired, significantly cut down on the amount of time spent acquiring data, and guarantee high-quality image reconstruction. Lustig et al. applied CS theory to MRI for the first time, where the undersampled MR image reconstruction is expressed as a l_0 norm minimization problem, and solved it by greedy algorithm. The l_0 norm minimization problem is a NP hard problem. Lustig relaxes the l_0 norm problem to the l_1 norm problem and uses the conjugate gradient method to solve it. In order to improve the speed and accuracy of image reconstruction, it applies total variation to CS-based MRI model and uses two regularization conditions to ensure high-quality reconstructed image. Wang et al. introduce the dual-tree complex wavelet to obtain the global sparsity priori, combined with CS theory to reconstruct the MR image with directional structure, but the comprehensive sparse coding phase in its solution process takes a long time, resulting in the low efficiency of MRI enhancement reconstruction. With the advent of compressed sensing (CS) theory, minimizing the recording time in K-space without affecting the image quality has become the main purpose for MRI enhancement research. Due to the effective use of signal sparsity, the K-space samples required for MR image reconstruction are far less than those of conventional methods. Furthermore, it can significantly reduce the scanning time, making it a popular fast imaging method. Nowadays, compressed sensing theory has been successfully applied to MRI reconstruction [10]. In CS-based MRI enhancement algorithm, the adaptive sparse representation of MR image plays an important role in high-quality image reconstruction. The adaptive sparse representation of MR image refers to learning by using the training samples of known MR image, so as to obtain the adaptive dictionary matching with MR image. The adaptive sparse representation of MR image can obtain accurate sparse priors and capture rich structural information of the image. In recent years, MRI enhancement algorithm based on adaptive sparse representation model has become the research direction. These improved CS-based algorithms can reconstruct MR images more accurately, but these algorithms have some defects, which still need to be further improved. As a result, with the assistance of computer science theory, the purpose of this paper is to conduct an in-depth investigation of the adaptive sparse representation model in order to increase the speed of image reconstruction while simultaneously enhancing its overall quality.

In this paper, a straightforward and speedy CS-based optimization model for noisy MRI Enhancement is proposed. In this model, the total variation is incorporated into the CS-based MRI reconstruction model, and three regularization conditions are utilized to ensure that the reconstructed image is of high quality. In order to quickly solve the objective function, the alternative direction multiplier method is utilized to optimize the model. Additionally, the k -terms power series is implemented in order to derive the LogDet function into the augmented Lagrange form, and

finally an approximation of the feature vector is achieved through the iterative procedure.

2. Radial Subsampling and Its K-Space in MRI System

As we all know, accelerating the speed of imaging has always been the focus of Magnetic Resonance Imaging research, which can not only improve the efficiency of MRI system, but also weaken or eliminate motion artifacts and noise, so that MRI can be better applied to medical diagnosis [6, 11–13]. In MRI system, there are generally two methods to shorten the time of data acquisition: one is to use multichannel parallel imaging technology; the other is to use non-Cartesian sampling trajectory to fill K-space, such as radial subsampling and sparse sampling [14–17].

Radial subsampling in MRI data acquisition is to collect magnetic resonance data in a radial trajectory rather than parallel linear manner. Radial subsampling not only changes the phase coding gradient and adopts sinusoidal gradient magnetic field, but also uses two coding gradient magnetic fields, G_x and G_y , so that the direction of the total gradient vector forms an arbitrary angle θ with the x -axis. Therefore, the total gradient intensities G and θ in the two directions are

$$\begin{aligned} G &= \sqrt{G_x^2 + G_y^2}, \\ \theta &= \arctan\left(\frac{G_y}{G_x}\right). \end{aligned} \quad (1)$$

Radial subsampling should also meet Nyquist sampling theorem. The difference of sampling steps will lead to the existence of aliasing artifacts in image reconstruction. In radial subsampling, the first two parameters G and θ cannot be directly converted into spatial resolution. Since the former depends on the number of radial lines n_s , the parameters can be set freely in actual sampling. Generally, the number of radial lines is set to n_s , so it can be denoted by

$$n_s = \frac{\pi n}{2}. \quad (2)$$

Satisfy the above formula to ensure that the maximum distance between adjacent radial subsampling lines is not greater than Δk . Since traditional MRI uses Cartesian trajectory, its reconstruction method is simple, but line by line acquisition is very sensitive to motion artifacts. Non-Cartesian sampling, such as radial sampling, has obvious advantages over Cartesian sampling. There are mainly the following two aspects: firstly, each line of radial sampling data contains the same amount of low-frequency to high-frequency information, which is conducive to the subsampling reconstruction of MRI images. Secondly, the radial sampling mode determines its oversampling of K-space center data, and K-space center data determines the main information of the image. Therefore, radial sampling is not as sensitive to the motion parameter. However, radial sampling also has some disadvantages. For example, the

radial sampling trajectory is densely sampled in the middle and sparsely sampled at the edge, and then the sampling density is uneven. Therefore, the image cannot be obtained directly by Fourier transform, so the imaging process is more complex. It is necessary to use grid interpolation method or Zero_filling strategy to interpolate the data to uniform grid points and then perform Fourier transform to obtain the final image. Radial subsampling and its MRI image are shown in Figure 1. The specific operation can be seen in [18].

K-space is the space for storing the raw data of magnetic resonance, and the magnetic resonance image can be obtained by performing very complex data postprocessing, as shown in Figure 1. After the radial sampling data set is interpolated into the Cartesian coordinate system by gridding convolution, the resampled data can be directly used to obtain the final MRI reconstruction image by inverse fast Fourier transform. It can be seen that there is no one-to-one corresponding relation between the array points in the K-space and the reconstructed pixel points in image space. But the frequency encoding direction and phase encoding direction of K-space are symmetrical. Due to the spatial positioning effect of the gradient field in the phase encoding direction, the phase encoding gradient in the center of K-space is zero, and the phase encoding on both sides increases in turn, so it is also symmetrical in t . This is because in the frequency encoding direction a continuous curve that is composed of many signal subsampling points is collected, and the curve is a symmetrical curve, so it is symmetrical in the frequency encoding direction. Each point in the K-space corresponds to all of the pixels in the MR image, and the image contrast is determined by the central part of the K-space, while the spatial resolution is determined by the peripheral part of the K-space. Theoretically, one can obtain a quarter of a subsample of the K-space, and then the remaining space can be filled mathematically, which results in an MRI image. This can result in phase errors and image distortions as a result of errors and noise in the data acquisition process. The challenge that needs to be tackled is figuring out how to carry out image reconstruction and enhancement with only a limited amount of data in order to obtain MRI data of a high quality. The CS-based MRI enhancement model is improved with the help of a powerful and efficient optimization algorithm that is used in this paper.

3. K-Space Enhancement Reconstruction Algorithm for Noisy MRI

3.1. Improved CS-Based MRI Model. In an MRI system, image signals are recorded sequentially in K-space. The scanning efficiency is the most important factor in determining how quickly images can be acquired in this system [19]. Because of the MRI system's slow data acquisition speed, motion artifacts and noise are both relatively simple to generate. According to the CS theory, if the signal in a transform domain is compressible or sparse, the high-dimensional signal can be mapped into the low-dimensional space using an observation matrix that is unrelated to the transform basis, and the original signal can then be

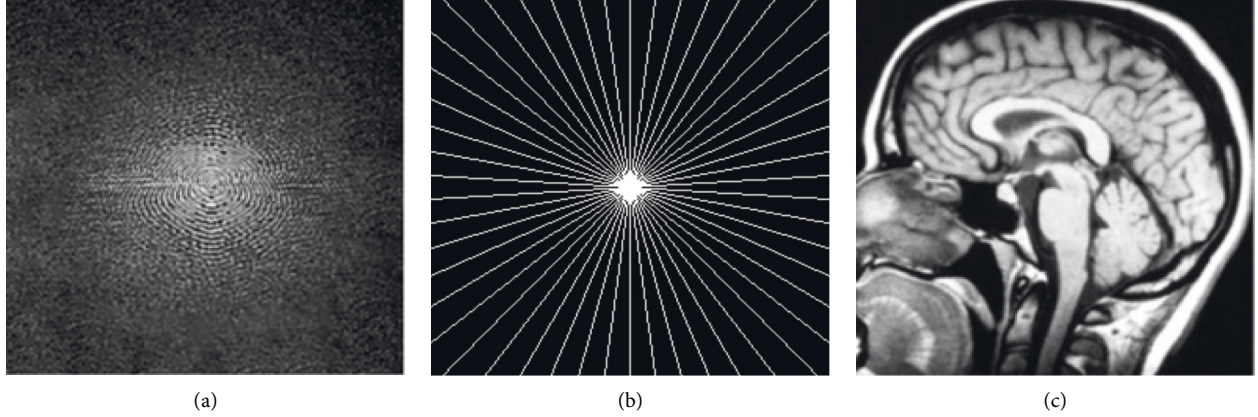


FIGURE 1: Schematic diagram of MRI reconstruction. (a) K-space (spectral) image; (b) radial subsampling; (c) MRI image.

reconstructed with a high probability from a limited number of projections by solving the optimization problem. By mapping the high-dimensional signal into the low-dimensional space, this is achieved. In the theory of CS, the image reconstruction model can be expressed as $Y = \Phi X$, where X is an image matrix; $\Phi \in R^{M \times N}$ is an observation matrix; Y is a measurement matrix. When $M < N$, this means that many solutions can be obtained for the equation $Y = \Phi X$. In order to constrain the results, it is generally necessary to add regularization terms to the equation, such as total variation, sparse, and BM3D.

If the image is sparse under the representation of a dictionary, and its model can be written as $X = DA$, where A is a sparse matrix and D is a dictionary, it can be solved by using an appropriate sparse optimization algorithm, and then the image X can be reconstructed by solving the $X = DA$. In the image inverse problem based on CS construction, selecting an appropriate dictionary plays a key role in image reconstruction.

MRI images are sparse in a specific transform domain, which meets the requirements of CS theory for signal sparsity. The convex optimization problem of l1 norm is constructed in combination with compressed sensing theory. It can be concluded that the objective function of CS-based MRI reconstruction is

$$\hat{x} = \operatorname{argmin}_x \frac{1}{2} \|Fx - y\| + \lambda L(x), \quad (3)$$

where x is the MRI image to be reconstructed, F represents the Fourier transform operator, and y is the K-space data obtained after MRI scanning; $L(x)$ is the regularization term. According to the distribution of the signals in the K-space data, the corresponding sparse transform domain, the observation matrix and the appropriate CS-based reconstruction algorithm are selected. The MRI image can be reconstructed by solving the CS-based enhancement algorithm for (3). However, to improve the reconstruction accuracy, many effective objective functions have been proposed, such as sparse and low-rank regularization constraints. The nonlocal group-sparse model proposed by Dong et al. is the most famous [2], whose global objective function is shown as follows:

$$(\hat{x}, L_i) = \operatorname{argmin}_{x, L_i} \|\Phi x - y\|_2^2 + \eta \sum_i \left(\|R_i x - L_i\|_2^2 + \lambda L(x) \right), \quad (4)$$

where $R_i x_i$ is expressed as a low-rank matrix composed of all nonlocal similar patches of the sample x_i . In order to solve the multivariable objective function, the idea of alternating solution is adopted to optimize, which fixes one variable to solve another variable and finally obtains the optimal solution. Although the objective function in (4) can achieve better performance than the traditional CS-based enhancement, it can only be effective for Gaussian distribution, and the solution process is extremely complex. TV regularization model is one of the most successful image reconstruction models, which can adapt the noisy data with different distribution for reconstruction and enhancement. In order to adapt to different noise distributions and accelerate the solution speed, the improved objective function proposed in this paper is as follows.

$$(\hat{x}, L_i) = \operatorname{argmin}_{x, L_i} \|\Phi x - y\|_2^2 + \eta \sum_i \left(\|R_i x - L_i\|_2^2 + \lambda L(x) \right) + \beta \int_{\Omega} |\Phi x - y| dx. \quad (5)$$

It can be seen that the improved model in this paper introduces the total variational regularization on the basis of (5). However, it is not easy to solve, so we relax the constraints; *i.e.*, $\int_{\Omega} |\Phi x - y| dx = \langle \Phi x - y, z \rangle$. As a consequence, we can adopt the alternative direction multiplier method for solving (6).

Firstly, split (6) and rewrite it into the augmented Lagrange form.

$$(\hat{x}, L_i) = \operatorname{argmin}_{x, L_i} \|\Phi x - y\|_2^2 + \eta \sum_i \left(\|R_i x - L_i\|_2^2 + \lambda L(x) \right) + \beta \langle \Phi x - y, z \rangle, \quad (6)$$

where z is Lagrange multiplier; γ, η, β are penalty factors. (x^k, x^k, z^k) is iterated through the alternative direction

multiplier method and get $(x^{k+1}, y^{k+1}, z^{k+1})$, whose step is shown as follows:

$$\begin{cases} (x^{k+1}, y^{k+1}) \leftarrow \operatorname{argmin}_{x \in R^n, y \in R^n} L_\gamma(x, y, z^k, \gamma), \\ z^{k+1} \leftarrow z^k - \gamma(x^{k+1} - y^{k+1}). \end{cases} \quad (7)$$

It is worth noting that (7) only regards ((6) as a general linear constrained convex programming problem without considering its separable structure, so it is necessary to solve two variables x^{k+1} and y^{k+1} at the same time. On the contrary, the alternative direction multiplier method decomposes (7) into two subproblems to solve x^{k+1} and y^{k+1} , respectively, so it has the following form:

$$\begin{cases} y^{k+1} \leftarrow \operatorname{argmin} L_\gamma(x, y^k, z^k), \\ x^{k+1} \leftarrow \operatorname{argmin} L_\gamma(x^{k+1}, y, z^k), \\ z^{k+1} \leftarrow z^k - \gamma(x^{k+1} - y^{k+1}). \end{cases} \quad (8)$$

In other words, the subproblem for solving x can be rewritten as

$$x^{k+1} = \operatorname{argmin}_{x \in R^n} \left\{ \frac{1}{2} \|\Phi x - y\|^2 + \left\| x - y^{k+1} - \frac{z^k}{\gamma} \right\|^2 \right\}. \quad (9)$$

Therefore, in order to solve the MRI enhancement problem under the background of least squares problem by using linear alternative direction multiplier method, $(\tau/2)\|x - x^k\|^2$ is introduced into the subproblem, so x -subproblem can be solved by computing

$$x^{k+1} = \operatorname{argmin}_{x \in R^n} \left\{ \frac{1}{2} \|\Phi x - y\|^2 + \left\| x - y^{k+1} - \frac{z^k}{\gamma} \right\|^2 + \frac{\tau}{2} \|x - x^k\|^2 \right\}, \quad (10)$$

where $\tau > \rho(\Phi^T \Phi)$ is positive constant. Therefore, the weighted singular value thresholding operator can be adopted to solve optimization result, as in [2].

In this paper, the total variational constraint is introduced into nonlocal group-sparse model, and the parameters of the objective function are relaxed, so that each subproblem can have analytical solutions, and the size of iterative step is greater than 1. For example, z^{k+1} has a closed-form solution; we can obtain

$$\begin{cases} y^{k+1} \leftarrow \frac{1}{\lambda^2 \Phi^T \Phi + \gamma I} (\lambda^2 \Phi^T y - z^k + \gamma x^k), \\ x^{k+1} \leftarrow P_\Omega \left\{ \frac{1}{\tau + \gamma} [\tau x^k + z^k + \gamma(\omega y^{k+1} + (1 - \omega)x^k)] - A^T(Ax^k - b) \right\}, \\ z^{k+1} = z^k - \gamma[x^{k+1} - \omega y^{k+1} - (1 - \omega)x^k], \end{cases} \quad (11)$$

where $\omega \in (0, 2)$ is the relaxation operator.

3.2. LogDet Optimization for Acceleration. To solve the improper operation of hard threshold for all singular values, the smoothing effect caused by inaccurate estimation of matrix rank, and the high time complexity of the algorithm caused by a large number of iterative optimizations, we use the LogDet optimization for acceleration to estimate the singular value, so we can directly obtain the singular value approximation based on the number of singular value iterations.

In [6], it has been proved that low-rank problem can be approximated to LogDet problem. For a general matrix L that is neither square nor positive semidefinite, it can be slightly modified as $L(X) = \logdet((XX^T)^{1/2} + \varepsilon I)$. Because of $XX^T = U\Sigma U^T$, so we can obtain $L(X) = \logdet(\Sigma^{1/2} + \varepsilon I)$, where Σ is diagonal matrix composed of singular values of XX^T . In other words, we can use the LogDet optimization to replace low-rank model, which can improve optimization speed. However, when the matrix decomposition method is used for XX^T , its computation complexity of the exact solution is $O(N^3)$, so its computational cost and additional storage requirements may limit their use in MRI real-time reconstruction.

Literature [20] proposed a uniform distributed sample selection method for the estimation of random traces in the process of Gaussian regression, whose computation complexity of the exact solution for logdet function is just a $O(N^2)$. In the decomposition of large sparse matrix, logdet(C) can realize the approximation estimation based on power series expansion. Since any matrix can be converted into a specific form $I - aD$, its eigenvalues remain less than 1 after row standardization or extraction of a sufficiently large factor, where, $a \in (0, 1)$, I and D are expressed as identity matrix and large sparse matrix, respectively. Firstly, the power series expansion method is applied to the approximation of generalized positive definite matrix, and then the power series truncation error compensation is used, and its result can be used as an effective surrogate function for accurate estimation. For a positive definite matrix $C \in R^{N \times N}$, the approximate value of the k -terms power series for logdet(C) can be denoted by

$$\logdet C \approx N \log(a) - N \varepsilon \left(\sum_{i=1}^k \frac{s^T B^i s}{i s^T s} \right), \quad (12)$$

where $a = \|C\|_\infty$, $B = I - C/a$, $s \sim N_N(0, I)$.

On basis of the infinite norm of the matrix, we can obtain

$$\begin{aligned} \log\det(C) &= \log\det(aA) = \log(a^N \det A) \\ &= N \log(a) + \log\det A, \end{aligned} \quad (13)$$

where $\alpha = \|C\|_\infty = \max_i (\sum_{j=1}^N c_{ij})$, and $A = C/a$. Therefore, $\log\det A$ can be approximately expressed as the sum of a simple polynomial

$$\log\det A \approx -N\varepsilon \left(\sum_{i=1}^k \frac{s^T B^i s}{is^T s} \right), \quad (14)$$

where ε is the mean for random number. The basic algorithm of $\log\det(C)$ is based on equations (13) and (14), where an intermediate vector ν needs to be calculated or stored; that is, $\nu^{f+1} = B\nu^f$ and $\nu^0 = s$. The above approximate estimation method only needs $O(N^2)$ operations and N vector storage. Compared with the traditional matrix decomposition, our adopted strategy can meet the needs of MRI data reconstruction in K-space.

4. Experiments

4.1. Parameter Setting. In order to verify the effectiveness of the proposed noisy MRI reconstruction algorithm in K-space, this section reports the experimental results for simulation comparison experiment. In this experiment, six MRI images are selected for testing, and the noise variance is 5, 10, 25, and 50 respectively. In the experimental parameters, the size f of image patch and the size W of search window are $\sigma^2 < 30f = 7$, $W = 21$; $\sigma^2 \geq 30$, $f = 9$, $W = 31$. Since the number of similar patches is adaptive, it overcomes the introduction of dissimilar image blocks due to the fixed number of similar patches, but too many similar blocks will increase the amount of calculation, so we set the boundary value $S_i \in [10, 60]$. The regularization parameter λ, β, η is tuned separately for different iteration times, where the initial values are set to 0.1, 0.05, and 0.01, respectively. Sensing rate in K-space is set to 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6. Noisy MRI reconstruction algorithms are programmed using MATLAB language. The experimental simulation platform is a personal computer with Intel (R) core (TM) i5 dual core CPU, 2.6 GHz frequency, 4G memory, Win10 operating system and MATLAB 2017a is selected as experimental simulation software.

4.2. Evaluation Indexes. Some common objective evaluation indexes for performance quantification have been adopted in order to perform an objective evaluation of the reconstruction performance of the improved reconstruction algorithm for noisy MRI with different noise distributions. These objective evaluation indexes include mean square error (MSE), peak signal-to-noise ratio (PSNR), and structure similarity (SSIM). Since MSE is the divergence of the mean square error between the original picture and the reconstructed image, its primary application is in the simulation experiment of a known original image.

$$MSE = \sum_{i \in I} |x(i) - \hat{x}(i)|_2^2, \quad (15)$$

where \hat{x} represents the reconstructed gray level. MSE reflects the approximation degree of the reconstructed image to the

original image. The smaller its value is, the better the noisy MRI construction effect is.

Peak signal-to-noise ratio is an image-quality evaluation index based on mean square error, which is defined as

$$PSNR = 10 \lg \left[\frac{255^2}{\sum_{i \in I} (x(i) - \hat{x}(i))^2 / M \times N} \right], \quad (16)$$

where M, N represent the size of image. It can be seen from the (15) and (16) that the evaluation of PSNR and MSE is just the opposite. The larger PSNR the better reconstruction result, and the smaller PSNR the worse reconstruction effect.

Structural similarity is to evaluate the performance of image reconstruction algorithm from the perspective of the edge structure of image, which evaluates the quality of image based on structural distortion. It is an objective evaluation method very close to human vision, which is defined as follows:

$$SSIM = \frac{4u_x u_y \sigma_{xy}}{(u_x^2 + u_y^2)(\sigma_x^2 + \sigma_y^2)}, \quad (17)$$

where u_x, u_y represent the mean value of the original image and reconstructed image, respectively; σ_x, σ_y can represent the variance of the original image and the estimated image, respectively. If the two structures are more similar, the value of SSIM is greater, and the value of SSIM is less than or equal to 1.

4.3. Reconstruction Performance for Different Noises. To verify the performance of the improved algorithm for noisy MRI reconstruction with different noise distributions, Rayleigh noise, Gaussian white noise, and random non-uniform noise are selected for reconstruction analysis, where the standard deviation σ_n of Gaussian noise is 5, 15, 25, 20, and 35, respectively. All of the comparison tests had their noisy images treated using the same settings. We also performed a comparison study of the experimental outcomes, looking at them from both a subjective and an objective perspective. During the subjective evaluation, the smooth area and the area with rich texture information are primarily selected for analysis, and the performance of the reconstruction method is evaluated based on the visual effect. During the objective evaluation, PSNR and SSIM are utilized in order to evaluate the performance of the reconstruction effect.

Uniform noise is defined as noise that obeys the same distribution at different pixel positions of the image; namely, the distribution of any position (i, j) is the same in $y = x + n$. Nonuniform noise is defined as the noise that either obeys different distributions at different pixel positions or obeys the same distribution but has different corresponding parameters. The noise $n_{ij} = n\delta_{ij}$ at any position (i, j) in the image $y = x + n$ obeys different distributions or has the same distribution but different parameters, where the value of the sampling function $n_{ij} = n\delta_{ij}$ at any position (i, j) is 1 and 0 at other positions. For the convenience of discussion, the standard deviation σ of nonuniform Gaussian

TABLE 1: Reconstruction performance for different noise distribution.

Images		Rayleigh			Gaussian			Nonuniform		
		NL-CS	GSCS	Ours	NL-CS	GSCS	Ours	NL-CS	GSCS	Ours
1	PSNR	31.51	30.23	33.71	35.54	34.29	37.43	29.73	30.11	33.12
	SSIM	0.928	0.934	0.942	0.912	0.920	0.941	0.951	0.955	0.961
2	PSNR	33.11	33.55	33.94	34.06	33.09	33.97	34.53	35.03	35.31
	SSIM	0.889	0.922	0.929	0.825	0.924	0.933	0.9822	0.9125	0.9825
3	PSNR	32.90	33.48	34.02	33.98	33.08	33.91	33.91	34.48	34.74
	SSIM	0.824	0.797	0.763	0.705	0.500	0.561	0.707	0.758	0.825
4	PSNR	30.99	32.42	32.43	32.57	31.55	32.62	33.21	33.82	33.80
	SSIM	0.784	0.815	0.833	0.825	0.817	0.833	0.841	0.8525	0.874
5	PSNR	33.16	34.42	34.97	34.97	33.03	34.93	34.21	31.51	31.85
	SSIM	0.650	0.597	0.663	0.705	0.714	0.561	0.707	0.751	0.821
6	PSNR	37.06	38.11	38.33	38.48	36.61	38.51	35.21	35.11	35.74
	SSIM	0.787	0.815	0.848	0.785	0.819	0.848	0.852	0.8458	0.852

noise follows the uniform distribution of $[1 : 100]$, while the standard deviation of uniform Gaussian noise is $\sigma = 20$. Table 1 shows the reconstruction results under different noise distributions. NL-CS and GSCS are the selected comparison algorithms, where NL-CS is a reconstruction algorithm based on nonlocal compressed sensing and GSCS is a group-sparsity based reconstruction algorithm. The best results in Table 1 have been bold. It can be seen that the reconstruction performance of the reconstruction algorithm proposed in this paper is the best under different types of noise, because the regularization constraints used in this paper are independent of the noise model, while the comparison models are optimized based on Gaussian noise model. The processing ability of NL-CS and GSCS for Gaussian noise is significantly better than that of the other two kinds of noise in Table 1.

It is evident from the results of the reconstruction shown in Figure 2 that our approach does a better job of preserving the features. When compared to NL-CS, it is only slightly poorer in the bright places, but when compared to other comparative algorithms it performs significantly better in the bright spots. The noise in the image is exceedingly chaotic as a result of the interference caused by the non-uniform Gaussian noise. The noise can be effectively separated using the model that was proposed in this study. The reconstruction performance is capable of preserving the detailed signal, particularly in the edge region, but there are also fake effects in some locations. This is especially true for the edge region. Some of the highlighted noise spots in the amplification region are smoothed down in the salt and pepper noise area; however the comparison process has a lot of interference, and the reconstruction impact is not very good. Figure 3 is the comparison of reconstruction residual for Gaussian noise with $\sigma = 20$, where the proposed model has the least texture detail and the best reconstruction performance.

4.4. Comparison of Reconstruction Performance for Different Algorithms. To verify the performance of noisy MRI enhancement reconstruction in K-space, we selected some

representative comparison algorithms, such as KLLD, SAIST, LSSC, NL-CS, and GSCS. The low-rank algorithms of KLLD reconstruction in solving RPCA model are singular value threshold operations based on hard threshold, and the low-rank matrix is obtained through repeated iterations. Our method is mainly to solve the improper operation of hard threshold for all singular values, the smoothing effect caused by inaccurate estimation of matrix rank, and the high time complexity of the algorithm caused by a large number of iterative optimizations. We use the LogDet optimization for acceleration to estimate the singular value, so we can directly obtain the singular value approximation based on the number of singular value iterations. SAIST is to solve the problem that LSSC needs to establish a very large redundant dictionary, resulting in too high spatial complexity and too high time complexity by solving it through low-rank algorithm on the basis of dictionary. Since the signal space can be divided into noise-free signal space and noise space, and the peak of the singular value difference curve is the critical point of the signal space, we used the LogDet optimization to divide the similarity matrix into low-rank part and difference part. The singular values corresponding to the low-rank part account for the vast majority of the sum of all singular values, and the proportion of noise pollution is small. Based on the prior information, we use the k -terms power series to split and rewrite the LogDet function into the augmented Lagrange form and then approximate the real vector in the iterative process. Finally, it can be seen that the improved CS-based MRI reconstruction algorithm can approximate the real noise-free data. Table 2 shows that, compared with NL-CS, KLLD, and GSCS, our algorithm can effectively reduce noise, improve denoising performance, and save texture edge and other structural information. When the noise is less than 15, our algorithm PSNR is much higher than NL-CS, KLLD, and GSCS about 0.2 dB. With the increase of noise intensity, the advantage of our algorithm weakens, but it is also equivalent to the best algorithm, which is due to the preservation of texture and the smoothing effect of smooth area. However, since our algorithm is based on LogDet optimization for

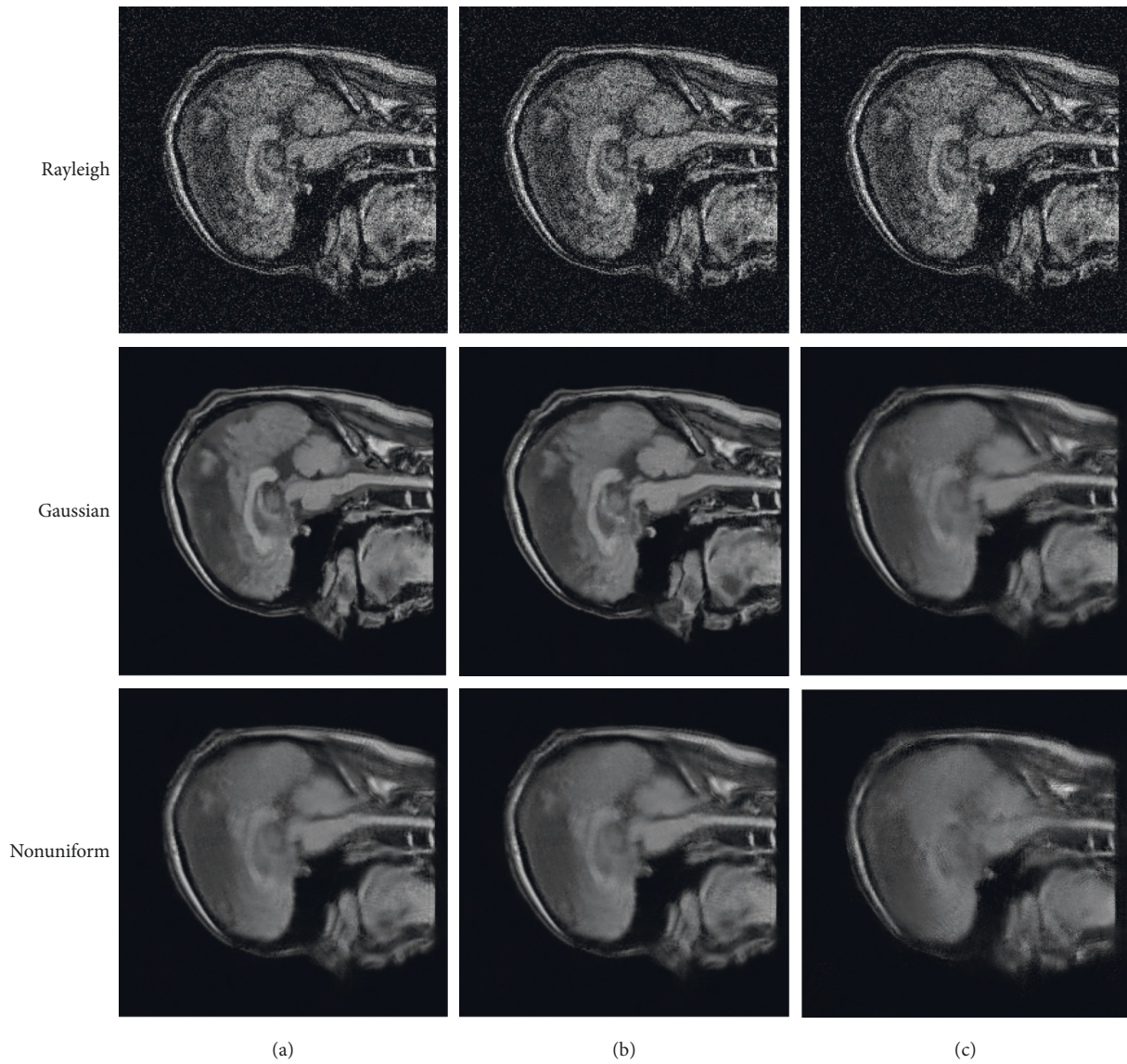


FIGURE 2: Comparison of reconstruction performance for different noise. (a) The proposed algorithm; (b) GSCS; (c) NL-CS.

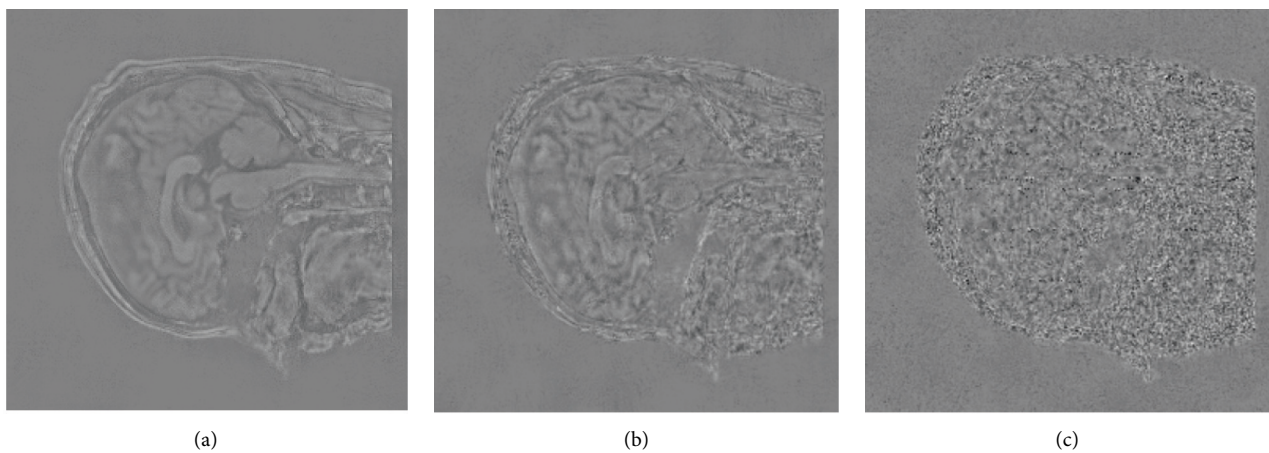


FIGURE 3: Comparison of reconstruction residual for Gaussian noise with $\sigma = 20$. (a) NL-CS, (b) GSCS; (c) the proposed algorithm.

TABLE 2: Reconstruction of average performance for different algorithm.

Average	σ	KLLD	SAIST	LSSC	NL-CS	GSCS	The proposed algorithm
SSIM	5	0.914	0.964	0.965	0.923	0.951	0.966
	10	0.908	0.934	0.941	0.890	0.942	0.941
	20	0.817	0.881	0.883	0.819	0.851	0.919
	25	0.815	0.850	0.887	0.793	0.804	0.892
	50	0.671	0.713	0.794	0.562	0.572	0.805
PSNR	5	36.75	37.30	37.49	37.45	35.89	37.39
	10	32.90	33.48	34.02	33.98	33.08	33.91
	20	29.01	30.02	30.76	30.69	29.92	30.77
	25	27.99	28.88	28.72	29.61	29.72	29.71
	50	24.58	25.29	26.46	26.30	26.20	26.37
Time/s	10	63.12	30.74	41.75	31.5	75.52	28.11

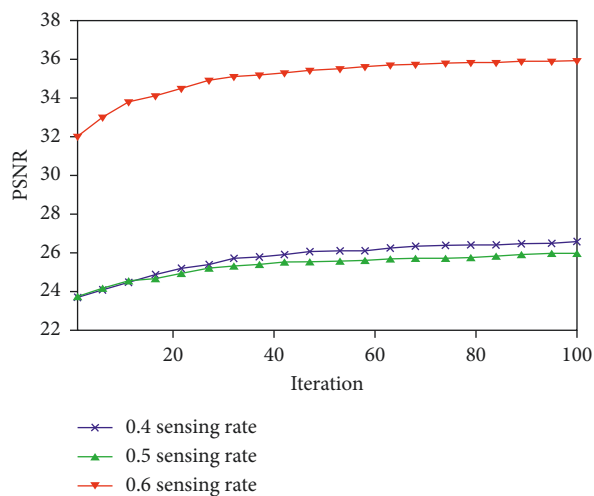


FIGURE 4: Comparison of PSNR of the reconstructed image with the proposed algorithm.

acceleration, the reconstruction time of the algorithm is the best.

For the irregular texture image, such as head MRI, our algorithm can retain the texture structure but ignore the small details. Even so, our algorithm is better than the comparison algorithm. NL-CS algorithm is not denoised enough in the texture area. Although the texture information is still saved, the reconstruction result is incomplete. Although the evaluation index of NL-CS and GSCS is greatly improved compared with KLLD, both methods are based on the sparse coefficients under the CS dictionary. Due to the pseudo-Gibbs effect, there are more scratches in the smooth area. Our method mainly deals with LogDet function, and Gibbs effect is inevitable, but we adjust the threshold according to the mode characteristics of the image patch to further improve our reconstruction quality; especially the TV model is introduced to separate the residual matrix of the similar patches. Our proposed algorithm performs much better than the comparison algorithms on all noisy images and sensing rates.

4.5. Ablation Analysis. The comparison of the PSNR of the reconstructed picture using the proposed technique for various sensing rates of an MRI image is shown in Figure 4.

The ordinate represents the number of iterations, and there are a total of one hundred iterative tests performed. Figure 5 demonstrates that the quality of the reconstructed picture for 0.4 sensing rate and 0.5 sensing rate is comparable to one another; however the quality of the reconstructed image for 0.6 sensing rate is noticeably superior to that of the other two sensing rates. A sensing rate of 0.6 is chosen for the radial subsampling lines because the goal of the ablation analysis is to lower the sensing rate in order to enhance the reconstruction quality and efficiency under the conditions without compromising the performance of the hardware.

In addition to the number of iterations, take the size parameter of overlapping patches as an example. We also analyze the results of reconstruction experiments under three different overlapping patch sizes 6×6 , 7×7 , and 8×8 . It can be seen that the PSNR of reconstructed images corresponding to the three patch sizes is not much different, but the SSIM of 8×8 is the best, and the growth range of PSNR is decreasing. Considering the calculation cost, 8×8 is selected as the size of overlapping patches. By experiment contrast, our proposed algorithm can preserve the edges and local structures better than comparison algorithms. This paper adopts LogDet optimization for acceleration. In order to analyze the reconstruction efficiency, the algorithm using

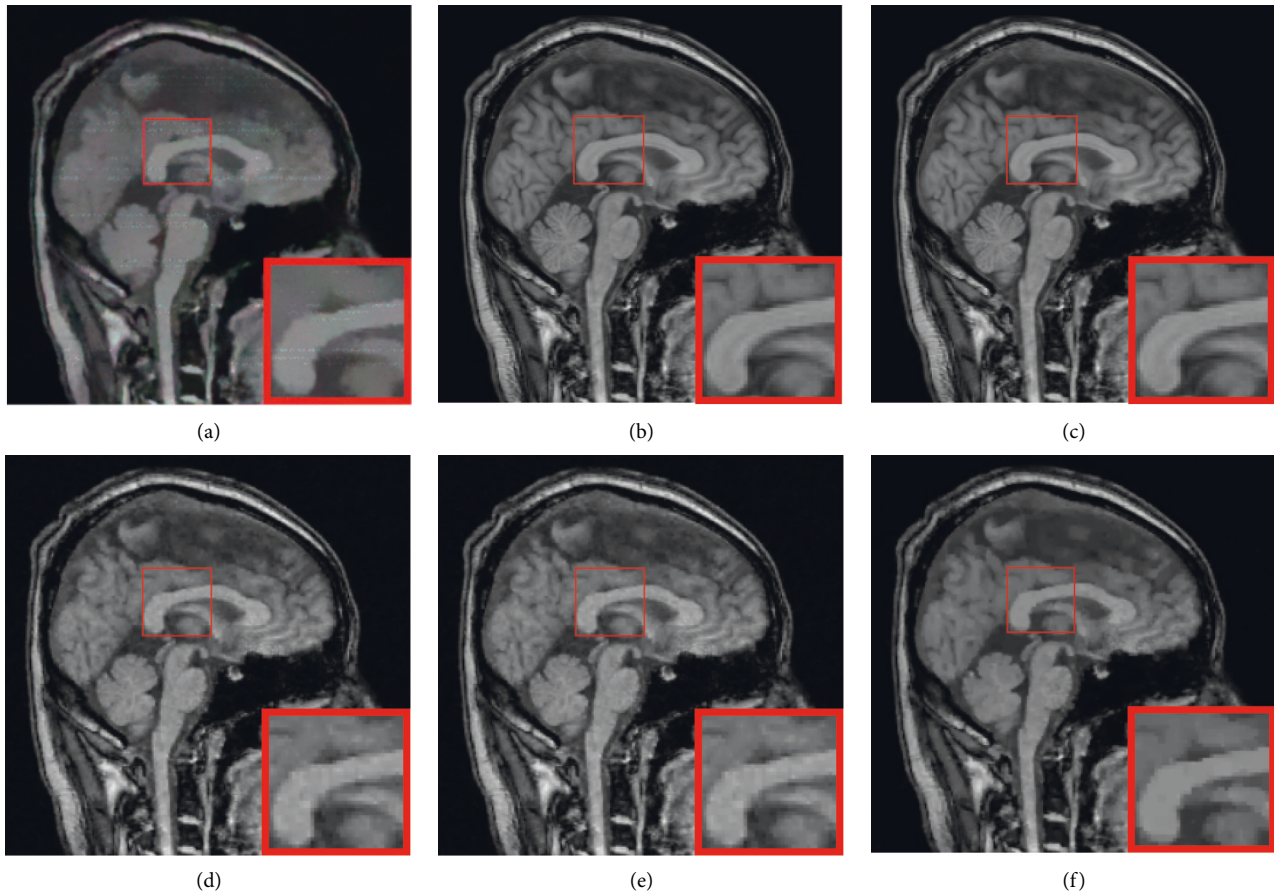


FIGURE 5: Comparison of reconstruction performance for different algorithm. (a) KLLD; (b) SAIST; (c) LSSC; (d) NL-CS; (e) GSCS; (f) the proposed algorithm.

low-rank decomposition is recorded as LR-CS, and the nonconvex LogDet surrogate of the rank is recorded as Logdet_CS. In order to facilitate ablation analysis, the algorithm model is consistent except that different methods are used to solve the low-rank module. We also analyze the reconstructed PSNR and reconstruction time under different iterations. It can be seen that the PSNR of different comparison algorithms changes little, but the time of each iteration is different. Therefore, the comparative experiment fully shows that the acceleration strategy proposed in this paper has high efficiency without loss of accuracy.

5. Conclusion

Due to the fact that Magnetic Resonance Imaging (MRI) is currently a very slow imaging technique, its application for dynamic imaging is restricted. The total variation is included into the CS-based MRI reconstruction model, and three regularization requirements are used to ensure that a high-quality reconstructed picture is produced. In this article, a simple yet efficient CS-based optimization model for noisy MRI enhancement is provided. The alternate direction multiplier approach is used to optimize the model, and the k -terms power series is applied in order to extract the LogDet function into the augmented Lagrange form. Following this,

an approximation of the feature vector is achieved through the iterative process. The quality of the rebuilt picture was substantially better than that of the CS-based MRI image reconstruction method, as shown by experimental results under varied noise settings. The peak signal-to-noise ratio of the reconstructed image was able to be enhanced anywhere from 5 to 20 percent.

Data Availability

The datasets 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 Northeast Petroleum University.

References

- [1] G. R. Matcuk, J. S. Gross, B. K. K. Fields, and S. Cen, "Compressed sensing MR imaging (CS-MRI) of the knee:

- assessment of quality, inter-reader agreement, and acquisition time,” *Magnetic Resonance in Medical Sciences*, vol. 19, no. 3, pp. 254–258, 2020.
- [2] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, “Compressive sensing via nonlocal low-rank regularization,” *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.
 - [3] Y. Chen, W. Qian, W. Liu et al., “Feasibility of single-shot compressed sensing cine imaging for analysis of left ventricular function and strain in cardiac MRI,” *Clinical Radiology*, vol. 76, no. 6, pp. 471.e1–471.e7, 2021.
 - [4] M. Sandilya, S. R. Nirmala, and N. Saikia, “Compressed sensing MRI reconstruction using generative adversarial network with rician de-noising,” *Applied Magnetic Resonance*, vol. 52, no. 11, pp. 1635–1656, 2021.
 - [5] K. Koolstra and R. Remis, “Learning a preconditioner to accelerate compressed sensing reconstructions in MRI,” *Magnetic Resonance in Medicine*, vol. 87, no. 4, pp. 2063–2073, 2022.
 - [6] B. Aska, B. Nnv, and B. Rr, “Accelerated dynamic contrast enhanced MRI based on region of interest compressed sensing,” *Magnetic Resonance Imaging*, vol. 67, pp. 18–23, 2020.
 - [7] A. Joy, M. Jacob, and J. S. Paul, “Compressed sensing MRI using an interpolation-free nonlinear diffusion model,” *Magnetic Resonance in Medicine*, vol. 85, no. 3, pp. 1681–1696, 2021.
 - [8] E. Shimron, A. G. Webb, and H. Azhari, “CORE-deblur: parallel MRI reconstruction by deblurring using compressed sensing,” *Magnetic Resonance Imaging*, vol. 72, pp. 25–33, 2020.
 - [9] C. G. Glessgen, H.-C. Breit, T. K. Block, E. M. Merkle, T. Heye, and D. T. Boll, “Respiratory anomalies associated with gadoxetate disodium and gadoterate meglumine: compressed sensing MRI revealing physiologic phenomena during the entire injection cycle,” *European Radiology*, vol. 32, no. 1, pp. 346–354, 2021.
 - [10] H. Wang, D. Liang, S. Su et al., “Improved gradient-echo 3D magnetic resonance imaging using compressed sensing and Toeplitz encoding with phase-scrambled RF excitation,” *Medical Physics*, vol. 47, no. 4, pp. 1579–1589, 2020.
 - [11] R. Liu, Y. Zhang, S. Cheng, Z. Luo, and X. Fan, “A deep framework assembling principled modules for CS-MRI: unrolling perspective, convergence behaviors, and practical modeling,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4150–4163, 2020.
 - [12] M. V. W. Zibetti, P. M. Johnson, A. Sharafi, K. Hammernik, F. Knoll, and R. R. Regatte, “Rapid mono and biexponential 3D-T1 ρ mapping of knee cartilage using variational networks,” *Scientific Reports*, vol. 10, Article ID 19144, 2020.
 - [13] Z. Zhuo, D. Yunyun, Z. Jie et al., “Accelerating brain 3D T1-weighted turbo field echo MRI using compressed sensing-sensitivity encoding (CS-sense),” *European Journal of Radiology*, vol. 131, no. 8, Article ID 109255, 2020.
 - [14] A.-I. Iuga, N. Abdullayev, K. Weiss et al., “Accelerated MRI of the knee. Quality and efficiency of compressed sensing,” *European Journal of Radiology*, vol. 132, Article ID 109273, 2020.
 - [15] C. A. Varela-Ma Baron and R. S. Menon, “Automatic determination of the regularization weighting for wavelet-based compressed sensing MRI reconstructions,” *Magnetic Resonance in Medicine*, vol. 86, no. 3, pp. 1403–1419, 2021.
 - [16] S. Yang, M. Wang, Y. Sun, F. Sun, and L. Jiao, “Compressive sampling based single-image super-resolution reconstruction by dual-sparsity and non-local similarity regularizer,” *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1049–1059, 2012.
 - [17] W. L. Li, X. Q. Yin, B. Wang, M. J. Zhang, and K. Tan, “Laser active image-denoising based on principal component analysis with local pixel grouping,” *Applied Mechanics and Materials*, vol. 571–572, pp. 753–756, 2014.
 - [18] M. G. Zeilinger, M. Wiesmüller, C. Forman et al., “3D Dixon water-fat LGE imaging with image navigator and compressed sensing in cardiac MRI,” *European Radiology*, vol. 31, no. 6, pp. 3951–3961, 2021.
 - [19] 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.
 - [20] A. Yi, L. A. Jie, B. Fm, D. Shuangli, and L. Yiguang, “High quality and fast compressed sensing MRI reconstruction via edge-enhanced dual discriminator generative adversarial network,” *Magnetic Resonance Imaging*, vol. 77, pp. 124–136, 2021.

Research Article

Exploration of Emotion Perception in Serious Interactive Digital Narrative

Jie Zhang  and Yaqian Liu

School of Design, Jiangnan University, Wuxi 214122, Jiangsu, China

Correspondence should be addressed to Jie Zhang; zhangjie@jiangnan.edu.cn

Received 22 May 2022; Revised 14 June 2022; Accepted 15 June 2022; Published 28 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Jie Zhang and Yaqian 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 procedural process of children's emotional involvement in the interactive digital narrative conforms to children's emotional attachment to the story and enhances the mediating nature of learning through forced interactivity. This study explores how compelling arcs influence learners' preference for serious story content by using a combination of natural language processing methods and statistical analysis methods. By analyzing 474 Chinese short serious stories, the emotional trajectory of each story is generated. Then, the obtained trajectories are combined into clusters of serious story emotional groupings through supervised learning. The study results found that the emotional arc in serious stories can be divided into six basic shapes, and the serious story with the highest preference is the "N"-shaped emotional arc. Emotional ups and downs characterize the emotional narrative aspect of this type of serious story as the story progresses but with an apparent emotional uptick towards the end of the story. Based on experimentally derived emotional topology and narrative generation methods, this paper proposes the design strategies for future emotional arcs to apply to serious interactive digital narratives.

1. Introduction

Storytelling is an essential way of communicating thoughts and ideas, going through stages of development from spoken language to written records. At all stages of its development, it is an important art form that inspires human emotions. This makes readers have an attachment to the world depicted in the story. In reading, readers are emotionally attached to the story that matches their current psychology [1]. Similarly, reading is accompanied by the process of continuous satisfaction of psychological needs and gradually develops a knowledgeable and confident self-identity. These psychological needs are mainly expressed as an emotional will, and interaction plays a vital role in satisfying the emotional will. Interactive digital narrative systems provide a more comprehensive storytelling process that integrates systems, users, processes, and frameworks [2]. This technology may generate a customizable narrative environment for users through prototype tales, narrative design, and narrative vectors. Interactive digital storytelling sits at the crossroads

of the humanities and computer science study domains, integrating computational storytelling and generative systems, with a great potential for varied applications and aesthetic approaches [3]. It resolves the contradiction between linear narrative and interaction in a "system-interaction-output" way, [4] producing instantiated narrative products through participatory process experiences. Serious Interactive Digital Narrative creates knowledge and wisdom in a specific context by harnessing the strengths of interactive storytelling to construct and convey serious messages. For example, "Global Conflicts: Palestine" is a serious educational game in which the users learn about the natural world by interacting as a journalist and forming a narrative report during the game. This approach provides children with new learning opportunities that are different from the classic linear narrative and is helpful for children's cognition and learning.

Some serious cultural knowledge imparted to readers should not simply be compiled into compact truths but should be conveyed in a serious narrative. The content of

serious stories is a module of social reasoning and environmental knowledge, and most of them show the behaviour of human character: aspects of integrity, honesty, and perseverance. The main reason why a serious narrative can move people more than “information” is that the story has the triggering power of emotion, and the microscopic individual narrative carries the life and cultural conditions experienced by the nation’s history, highlighting the national solid narrative tendency. Invoking the concept of emotional arcs in narratives, transforming emotional fluctuations into time-series data, is closely related to narrative structure. Whether it is an early automatic storytelling system based on emotional arcs and symbolic planning [5] or a neural story generation method that models emotional trajectories with the support of deep learning techniques [6], they both show that introducing emotion into plot construction results in diverse and exciting stories. By introducing emotion into constructing a serious interactive digital narrative system, personal stories are shaped by the trajectory of emergent emotions. Narratives are organized by emotional experiences, in which personal intentions, goals, and hopes are expressed, and the resulting personal stories contribute to the construction of personal meaning and self-identity. From the perspective of data science, this study takes the texts currently applied to serious narratives as examples to explore which narratives are more acceptable to the readers under the influence of emotional arcs. Furthermore, a future-oriented design strategy for serious interactive digital narratives is proposed based on the experimental results.

Section 2 provides a comprehensive introduction to previous related literature research, Section 3 presents the processing of datasets and sentiment arc analysis methods based on natural language processing and statistical methods, and Section 4 points out the design strategies for emotional arc-perceived serious interactive digital narratives based on experimental results, Section 5 discusses the limitations of the study, and Section 6 summarizes the conclusions and future work.

2. Related Work

2.1. Serious Interactive Digital Narrative. The narrative itself is a storytelling artifact consisting of a sequence of narrative streams of events that take place in time and space that move the story forward [7]. The interactive digital narrative is an expressive narrative form in digital media; it also is conceived of as a computer system comprising possible storylines [4]. As a systematic framework, the interactive digital narrative has three components, including the drama manager, user model, and agent model [8]. This paper only reviews the research part related to narrative generation in the serious interactive digital narrative.

Interactive digital storytelling has been around for nearly 30 years. In terms of narrative generation, early interactive digital narratives relied on intelligent planning, programming the content of interactive digital narratives as program nodes, and abstracting them into planning problems. Ensure a coherent experience by balancing a coherent story progression and user autonomy in a systematic way [9]. For

example, the children’s book series “choose your adventure” splits the experience into interaction points with decision-making. Computer intervention in the interactive narrative attempts to use the model of artificially written narrative program node with a model generator based on planning domain language [10]. For example, the Automated Story Director, [11] Player-Specific Stories via Automatically Generated Event (PaSSAGE) [12], Player-specific Automated Storytelling (PAST) [13], these models are based on a specific story (The RED), programming narrative into nodes, study and explore author intent, virtual agent and user modeling in interactive digital narratives to generate emergent narratives, and automatically generate narrative content by modeling drama management, AI experience managers. The researchers further pointer out the possibility of interactive storytelling in the fields of entertainment, training, and education [11]. However, this kind of interactive digital narrative with computer planning as the core is still subject to the strong story script written in advance by the author/designer and the narrative plan that needs to be foreseen in advance, which can only operate in the pre-defined field, and it is difficult to generate a more complex narrative in line with the human cognition.

This problem has been mitigated by the increasing sophistication of machine learning techniques, and the training of large-scale neural language models has brought more attention to computers’ understanding, telling, and creating stories. Computational narrative generates coherent story content through a story-generation approach based on common-sense reasoning; for example, the transformer-based large-scale language model GPT-2 [14] generates fluent text by training on the WebText corpus, a collection of texts scraped from the internet. The computational narrative is not only about the generation of commonsense knowledge but also about the controllability of story generation. Model Fusion improves the quality of randomly produced stories by combining a convolutional sequence to sequence model with a self-attention mechanism.

With the development of computational storytelling, the possibility of interactive digital storytelling applied to serious and non-entertainment environments has emerged. Lugmayr et al. originally defined serious storytelling as storytelling with a purpose beyond enjoyment and identified applications of serious storytelling in fields such as education [7]. The first attempt in this field was serious games, aiming to convey serious messages through constructing the digital media of games. James Lester et al. have been actively studying the construction of a catalog of narrating-centric learning environments in serious games [15]. These include the influence of story on the learning process and offering automated evaluations and individualized feedback from a data-driven perspective. Most of these studies are from a pedagogical perspective and are beyond the current focus of this paper. The SIREN project has not completed the final design and pilot study [16]. However, the study proposes using the computational emotional player model as the pillar of the generation system to ensure an efficient adaptive narrative experience for users, which can provide a reference for our study. Arash et al. focused on player modeling in

serious interactive digital narratives, explored how to use artificial intelligence technology to provide personalized experiences to consolidate the influence of serious discourse and proposed a prototype for the game design.

2.2. Emotional Arc in Storytelling. Emotion is a vital element of storytelling, a component of the cognitive portion of the tale that inspires in the listener, who sees the narrative as an emotional experience. With the increase of computer power, individuals begin to analyze this emotional experience from the standpoint of big data. Moreover, the current breakthroughs in natural language processing and computational narratology make it feasible to examine the emotion of the text. From the standpoint of computational narrative, Chen et al. have proved that emotional development is important for story understanding [17].

Kurt Vonnegut was one of the first researchers to use data science to analyze the emotional content of stories. He coined the term “emotional arc” in the narrative. He defines the emotional arc as the relationship between the horizontal storytime “start-end” and the vertical emotional “Ill Fortune-Great Fortune.” By identifying emotional arcs in a sample of 1327 stories in the Project Gutenberg fiction collection, Reagan et al. discovered a set of six fundamental emotional arcs that are also the basic building blocks of complicated emotional trajectories [18]. Del et al. derived emotional arcs for each film by analyzing 6,174 films using a combination of natural language processing and econometrics. By plugging these emotional clusters into the econometric model, it is concluded that films affected by the U-shaped emotional arc can achieve box office success regardless of genre, helping to improve the productivity of the entertainment industry [19]. On the opposite, namely, by beginning with the emotional arc and contributing in the production of fascinating stories. During the planning language coding period, Sergio Poo et al. created the interactive narrative experience manager PACE, which generated narrative nodes by rewriting the ballet story Gisele, used PACE to predict the user’s emotional response in the narrative process, and used the emotional response to shape the subsequent narrative content, so as to keep the user on the target emotional arc. In the era of computational narrative, EC-CLF, a story generation method based on the emotional arcs of the protagonist proposed by Brahman and Chaturvedi, allows users to present the story’s progress by entering the title and the emotional arc of the protagonist [20]. The ViNTER (Visual Narrative Transformer with Emotional arc Representation) model proposed by Kohei et al. can automatically generate image narratives represented by “emotional arcs” as time series. The above researches have analyzed the primary form of an emotional arc and the advantages in the creative entertainment industry from data science. Computational narrative studies the impact of emotional arcs on narrative generation from the perspective of text and image content generation. However, in these studies, the generation of narrative lines still relies on the emotional arc input by the user in advance, and there is a lack of pre-research on the emotional arc of user

preferences. The application of this method in the serious interactive digital narrative is difficult to balance the needs of narrative consistency and perceived self-agency. Users are given too much autonomy, and it is difficult to accomplish the educational purpose of narrators in a serious interactive digital narrative. At the same time, the emotional arc does not intervene and influence the narrative content in the narrative process, which is difficult to reflect the positive influence of the process exploration of interactive digital narrative on learners’ acquisition.

Through the systematic review of narrative generation in interactive digital narrative and the analysis of serious interactive digital narrative used in serious non-entertainment occasions, the paper combines the computational generation of emotional arc in data science and the application of emotion arc in narrative generation. A two-stage study of our approach to serious interactive digital tales is offered. (i) In the first stage, natural language processing and sentiment analysis are used to acquire the emotional topology of serious narratives based on emotional arcs, and the emotional arcs that learners favor based on the quantity of reading and likes are obtained. (ii) In the second step, we construct our design approach for serious interactive digital tales based on the existing narrative creation methodologies and the emotional topology established in the first stage.

3. Method

3.1. Dataset. The dataset for this project comes from a general application, and there is a separate section in the application called “Red China.” The serious narrations below the catalog are mainly dominated by typical cases, historical figures, and wonderful stories from different periods, comprehensively demonstrating the connotation of the national spirit in Chinese culture for publicity and education. The number of reads and likes for each serious narrative after reading is also available. As a learning platform, Learning Power has collected a lot of knowledge information, including documentaries, concept explanations, micro-videos, and classic works. For this project, we focused on the Serious Stories section, which contains 1,826 pieces of information, in order to filter the quality and dependability of serious tale text material and assure the link between narrative text and learner choice. Deleting the section of the narrative material that only comprises micro-videos decreases the total number of serious tales to 1,043 and then deletes serious stories without clear emotional inclinations, resulting in 563 results. Finally, the dataset is linked with the user reading information on serious tales from the site, including readings and likes. Matching and further cleaning the data resulted in a final dataset of 474 serious stories in total, as shown in Figure 1.

3.2. Methodology. We use the resulting dataset of 474 serious stories for sentiment analysis, and the specific methods used are shown in Figure 2. To analyze Chinese sentiment arcs, we used and modified the *syuzhet* R package to uncover the underlying structure of narratives through sentiment

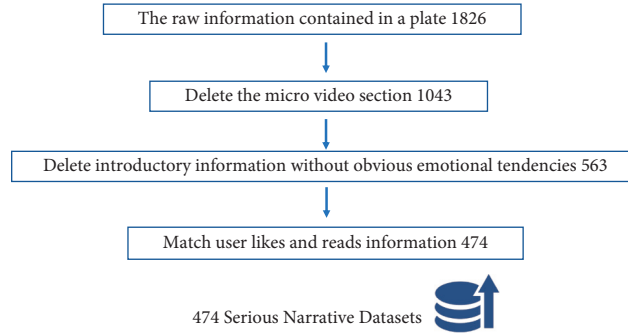


FIGURE 1: Methods of data processing and cleaning.

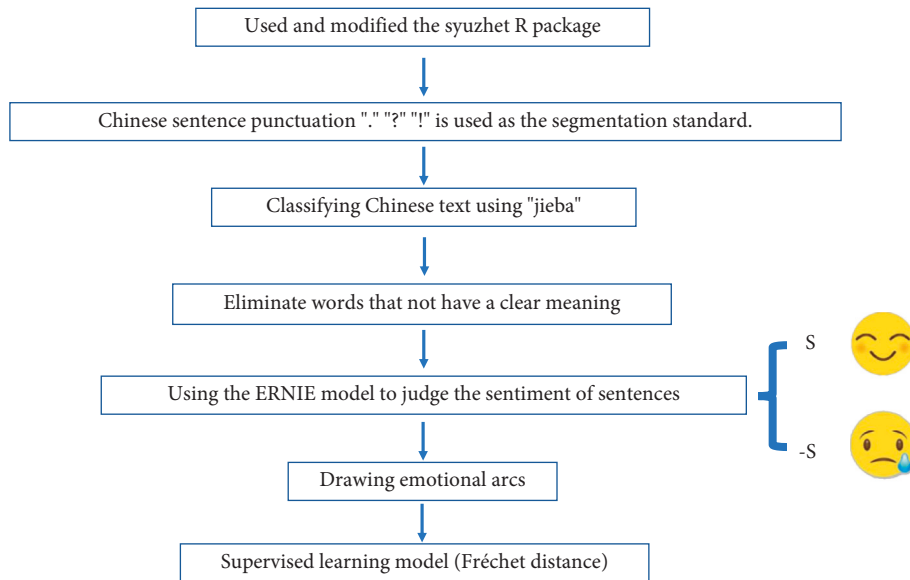


FIGURE 2: Sentiment analysis and emotional arc drawing methods.

analysis in Chinese, revealing sentimental changes in narratives that are proxies for narrative movement between conflict and conflict resolution. Unlike the sentiment analysis method based on the sentiment dictionary used by Reagan et al., this study uses the text sentiment analysis method based on deep learning. First, the collected data is preprocessed in a loop, and the Chinese sentence punctuation “.” “?” “!” is used as the segmentation standard. Based on regular expressions, fine Chinese sentences are divided into n different sentences for each short story. Then, the jieba word segmentation tool is used to divide each sentence. In word segmentation, the Chinese public stoplist eliminates words that appear frequently but have no clear meaning. The embedding vector of the spliced sentence is obtained through the ERNIE model [21]. The embedding vector is linearly transformed and classified into two categories to judge its emotional tendency. According to the content of the text, the system gives the corresponding scores of positive and negative emotions as s ; when it is determined to be a positive emotion, it outputs s . When it is determined to be a negative emotion, the output is $-s$, and the range of positive and negative emotions is scaled to $[-1, 1]$; treat each sample in the dataset as a dyad $\langle d, s \rangle$. This study uses a

window-based approach to analyze the sentiment of each serious narrative, first extracting the text covered by a separate window, and each time the loop is executed, the window slides forward until the end of the text is reached.

The generated emotional arcs are clustered according to the existing research on emotional arcs in novels and movies by Regan and Marco et al. [18, 19]. We hypothesized that the emotional arcs of serious narratives conform to the six emotion categories studied by Regan et al. The six categories are Rags to riches (rise), Riches to rags (fall), Man in a hole (fall-rise), Icarus (rise-fall), Cinderella (rise-fall-rise), and Oedipus (fall-rise-fall). The resulting 474 emotion arcs were clustered using supervised learning. Specifically, the Fréchet distance algorithm calculates the maximum difference between the two sequences and aligns each generated emotional arc with the six emotional arc templates calibrated in advance. Statistical analysis was then used to introduce the ratio of likes (a number that reflects learners’ affirmation of a serious narrative after reading) to the number of readings. The likes rate of serious narrative users under different emotional arc categories is calculated separately. In this way, the learners’ preference for serious narratives of different emotional arcs is known.

3.3. Result. Through systematic analysis using supervised learning, it is found that all 474 emotional arcs of serious narratives in the dataset can be matched with the six basic emotional arcs proposed by Regan et al. All serious narrative story texts can be divided into six main emotional trajectories. Figure 3 shows the clustering results of the experiment, where the red line represents the mean of each clustering result. Figure 4 shows a collection of six arc means. In each figure, the length of serious narration from beginning to end is displayed between the horizontal axis [0, 100]; and the change of emotion is displayed between the vertical axis [-1, 1].

Our 474 serious narrative screening datasets include: “Rags to riches” (rise) 46, “Riches to rags” (fall) 35, “Man in a hole” (fall-rise) 176, “Icarus” (rise-fall) 64, “Cinderella” (rise-fall-rise) 90, and “Oedipus” (fall-rise-fall) 63. There are more than 30 in all categories, and the rest of the summary statistics are shown in Table 1. Detailed serious narrative topics and related information have been stated in Data Availability. According to the data analysis in Table 1, among the six categories, “Man in a hole” and “Cinderella” are the two types of clusters that account for the largest proportion. The reading volume of users is relatively obvious. It is not hard to find common ground between these two emotional arcs in that no matter how the story develops, their results are positive. However, the number of likes after reading can more intuitively reflect the learners’ feedback on the story’s content after reading from a data perspective. Comparing the average like rates of the six different types of serious narratives found that the Cinderella cluster was the most popular, far exceeding serious narratives influenced by other emotional arcs. As shown in Figure 4, the main feature of this emotional arc is the ups and downs in the middle story but a clear emotional rise at the end of the story. Furthermore, the negative feeling b_2 of plot emotions in the middle of the story is lower than the negative value at the beginning of the story, and the positive value at the end of the narrative is higher than the positive value in the development process story. By analyzing these six arcs and their readers’ preferences, in general, “Man in a hole” and “Cinderella” emotional arcs, which have a good ending g after the story’s tortuous development, have higher preferences. In contrast, fewer readers prefer “Rags to Riches,” with rising emotions without twists and turns. Compared with serious narratives with rising emotions, “Riches to Rags” serious narratives with declining emotional arcs are more popular. It may be that the serious narrative of the dataset contains many stories about the protagonist going through a stubborn struggle with fate and finally dying generously. Such stories follow the “fall”-shaped emotional arc in the narrative and, at the same time, can influence readers to achieve the purpose of education. Three groups of serious narratives ending in tragedy, the “Riches to Rags” and the “Oedipus” serious narratives both start with a happy state and end with a sad emotional state, while the “Icarus” narrative start with a sad state and ends with a sad state. Comparing these three types, users have a higher preference for the “Riches to Rags” and the “Oedipus” emotional arcs, while users have a lower preference for the “Icarus.” This result suggests that when a serious narrative ends with a tragedy, readers are more

receptive to an emotional arc that begins with a positive state. It allows the reader to start the narrative with joy and a sad ending. Meanwhile, the word count difference in serious narrative content under different emotional arcs is also an important factor influencing readers’ preferences. Counting the number of words in each category of narrative content and calculating the average, the average number of words in serious narratives is between 1300 and 1550 words. The popularity of serious narratives is higher when the number is around 1300 words. With 1310 words as a boundary, the preference for more than 1310 words is generally lower than that of the serious narratives with less than 1310 words.

4. Serious Interactive Narrative Strategies for Emotional Arc Perception

Artificial intelligence and natural language processing offer many opportunities for generating intelligent feedback in interactive digital narratives, making the overall structure of the narrative personal, more relevant to the reader on an emotional level, and inducing a more profound sense of belonging to the narrative. Based on the findings above, we found that learners prefer a serious narrative emotional arc of “rise-fall-rise.” Under the influence of this emotional arc, we propose a serious interactive digital narrative system for emotional arc perception. It is expected to use artificial intelligence technology to assist the design of interactive narratives in consolidating the influence of serious discourse. The serious interactive digital narrative guided by this emotional arc is not just about the story’s content being kept in an “N” (Cinderella) arc. However, it emphasizes that the interaction becomes a decision-making process and changes the narrative flow to achieve important situational goals. It is shifting the serious narrative from an output-centric point of view to a procedural narrative point of view, focusing on how the output process is emotionally oriented. It turns from the narrative itself to the process of its construction and finally generates a serious narrative generation mechanism of “data + human + algorithm.” In the proposed serious interactive digital narrative strategy, the “N”-shaped emotional curve in the results is taken as a reference. In the proposed serious interactive digital storytelling strategy, the “N”-shaped emotional arc in the results is used as a reference. The optional emotional dimension question and answer is generated through programmatic content generation and user data; narrative content is jointly generated through emotional interaction between users and the system at nodes and emotional monitoring of narrative generation. Let users invisibly fit the “N”-shaped emotional arc in the emotional experience of the entire serious interactive narrative. We propose only a design prototype, and its realization still requires the joint efforts of the humanities, computing, and design disciplines.

4.1. Program Content Generation. In order to complete the intelligent content generation of serious interactive narratives, we need to build a large corpus of serious narratives

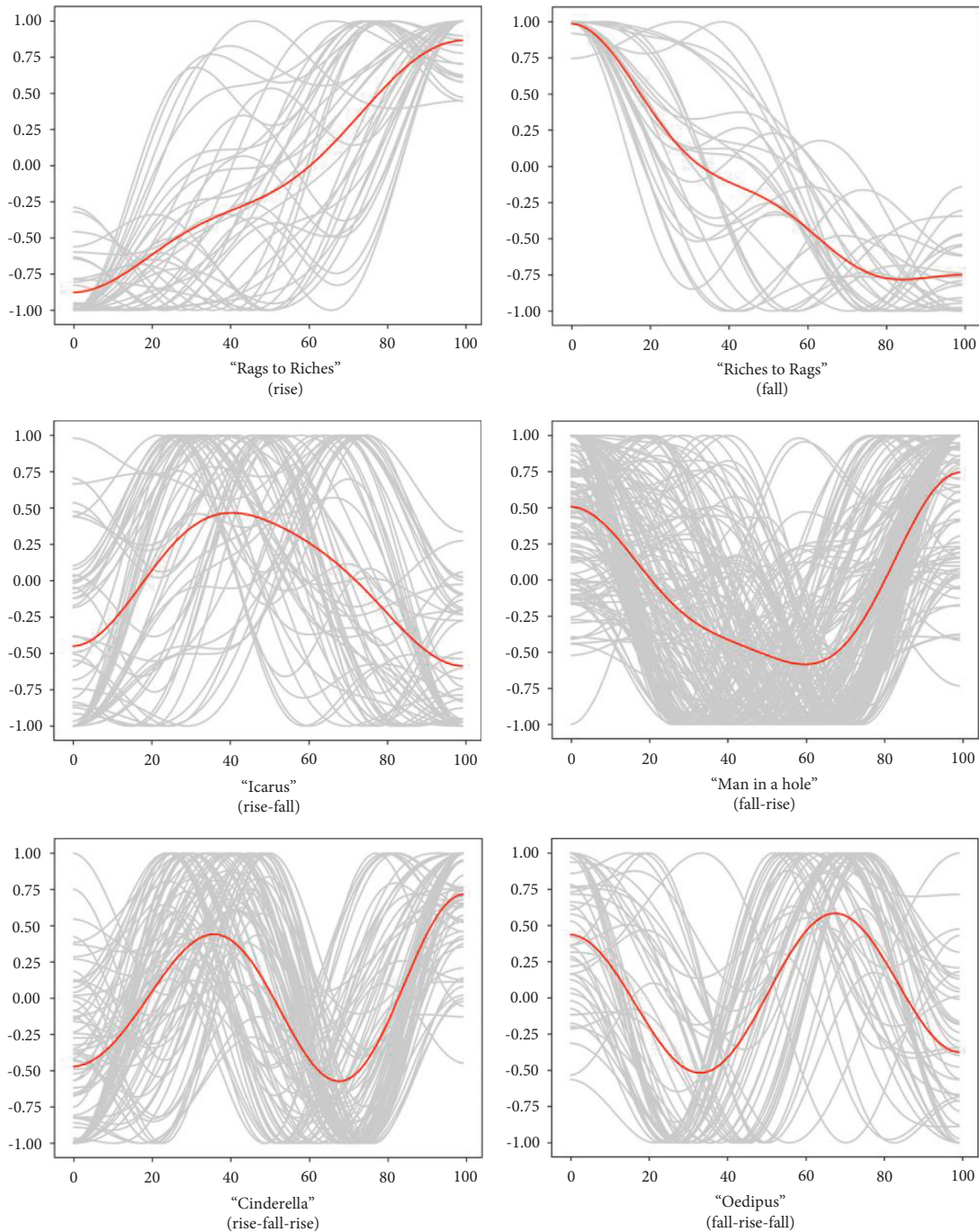


FIGURE 3: Six emotional trajectories of serious narrative.

and train the corpus. Ensuring that the generated stories are organized around themes and maintaining the consistency of entities and events in narrative generation using existing text generation methods. The quality of generated stories can be improved by combining the wander sequence to sequence model and self-attention mechanism to ensure that the generated content is the logical continuation of the narrative. Moreover, it eventually reached the state of generating continuous text content related to the input topic, just like the Reddit WritingPrompt [22] project implemented by Fan et al. as shown in Figure 5.

4.2. Emotional Dimension Question and Answer. This part of the questions is organized under specific categories of different emotional levels. The detailed part is shown in Figure 6. Several emotional dimension question templates are preset in the emotional selection pool, corresponding to the three dimensions of joy, general, and sadness, respectively. When narrative generation and interaction occur, the emotion selector modifies the question template in the emotion selection pool at the corresponding node according to the narrative content of the previous stage, the survey of the user's style in the previous stage,

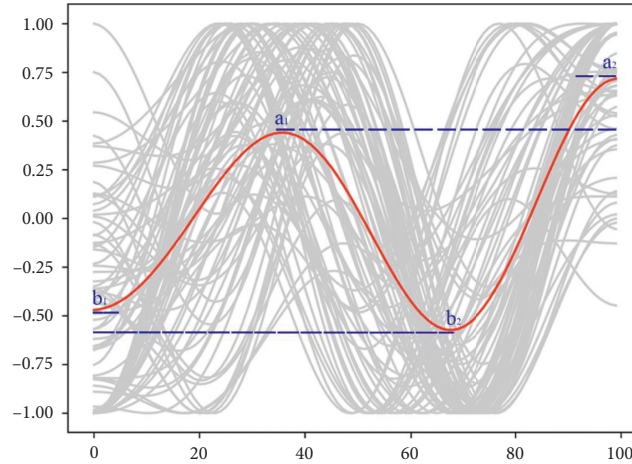


FIGURE 4: Node analysis of “cinderella” emotional arc.

TABLE 1: Summary statistics of analysis results.

	Rags to riches	Riches to rags	Man in a hole	Icarus	Cinderella	Oedipus
Number	46	35	176	64	90	63
Average reads	24401.7	22246.89	46692.81	36798.40	48236.34	41877.84
The average ratio of reads to likes	3.88%	4.28%	4.42%	4.07%	5.13%	4.13%
Average word count	1397.9	1301.8	1298.7	1547.7	1310.6	1455.3

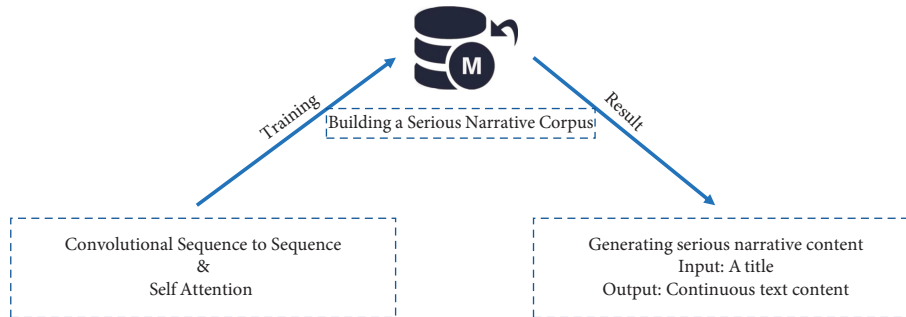


FIGURE 5: Program content generation.

and the remote sensing monitoring of the user. The remote sensing monitoring of users during the reading process mainly collects user reading data, including reading time, facial expressions [23] while reading, eye movement data, and heart rate data. The user’s style is divided into storytelling, show-off, and modesty. Then, push the question to the user in an orderly manner. In this way, effective emotional questions are selected. An appropriate emotional experience is maintained for users, keeping the serious narrative interesting for the user based on the corresponding emotional arc. The key motivation for this design decision is to focus on delivering several aspects of emotional options for the user’s reading experience. The scope and style of questions can be defined according to narrative material and the reader’s style. Provide the user with emotional choice while keeping her inside the confines of the narrative structure to produce the effect of the serious story of emotion perception.

4.3. *Emotion-Driven Program Content Generation.* In emotion-driven program content generation, we need to introduce the emotion monitoring and the emotional reward models, use emotion monitoring to supervise the content generated in the process of program generation, and use emotional dimension questions and answers to complete emotional rewards. The two work together to approximate an “N”-shaped emotional arc. The detailed system process is shown in Figure 7. After the user enters the corresponding topic, the system provides narrative content with rising emotions by default. When the content is generated to a turning point, the emotion selector executes the command to provide the user with emotion questions of three different dimensions. The different choices of users correspond to three possible narrative emotional trajectories: in the picture, *a*-continue to rise, *b*-stay level, and *c*-begin to fall. When the user’s choice is “*a*,” the narrative emotion continues to rise until the system determines that the emotional

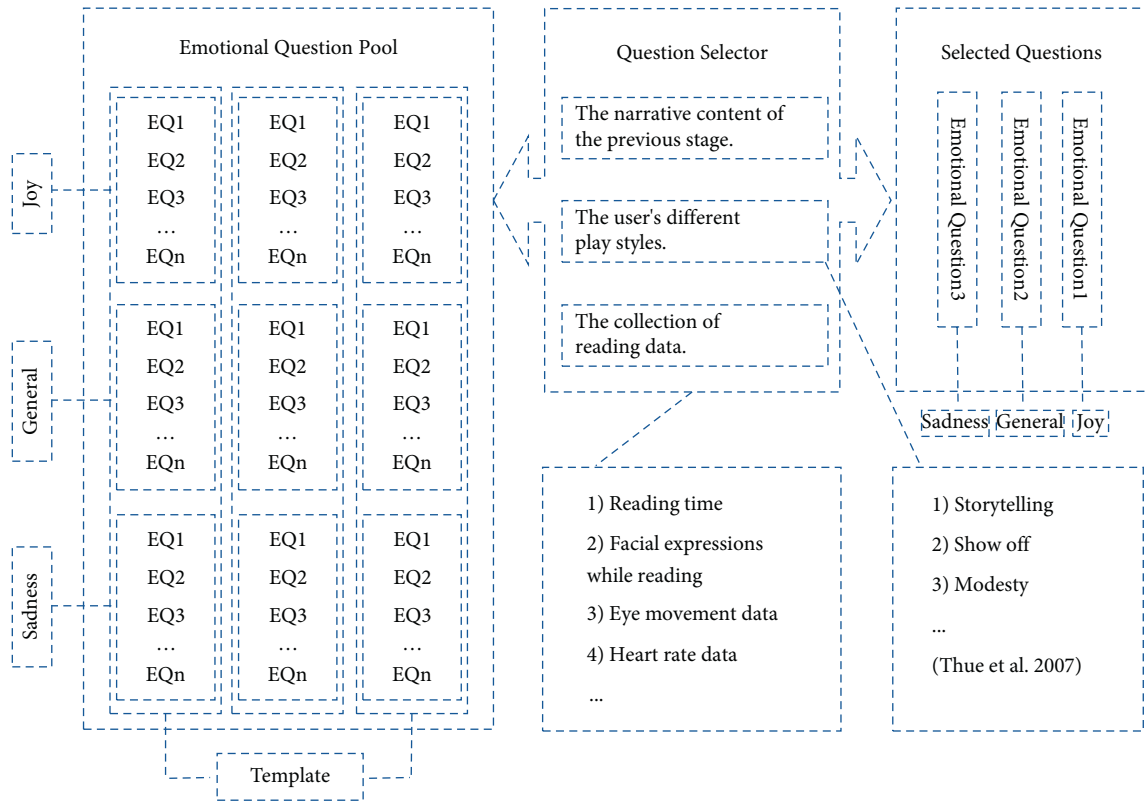


FIGURE 6: Emotional dimension question and answer.

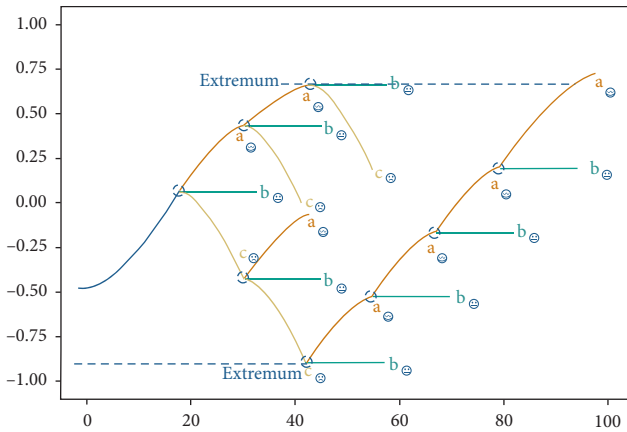


FIGURE 7: Emotion-driven program content generation.

arc reaches the extreme value, and the negative intervention of the narrative content needs to be enforced. When the user chooses “b,” it may indicate that the user’s emotional state is not high at this time, and a relatively gentle narrative needs to be generated to help the user complete the transition. When the user selects “c,” it has begun to enter the negative stage of the narrative. The emotion needs to be questioned again at the node where the narrative progresses until the emotion monitoring detects that the emotional state generated during the narrative process is lower than the negative state at the beginning. At the node of narrative generation, the generation of narrative content is affected by the way “a, b, and c” appear alternately until the narrative generation

begins to enter the final emotional rising stage. The emotional Q&A at the node in the final stage should be dominated by positive questions, constantly guiding users to reach the final positive emotional state of the serious narrative, and surpass the positive emotional peak in the narrative process. In a sense, this method achieves the purpose of personalizing the narrative with the user’s more preferred emotional arc and reflects the exploration and influence of the user’s emotional perception in the serious interactive narrative.

5. Discussion

5.1. Analysis of the Main Reasons for the Preference for “N” (Cinderella)-Shaped Emotional Arcs in Serious Interactive Digital Narratives. Through the application of natural language processing technology, it is concluded that users have the highest preference for “N”-shaped emotional arcs in serious narratives, and it is concluded that 1310 words are the best word count for serious narrative texts. Our findings are inconsistent with the novel emotional arcs of Regan et al. that “Icarus,” “Oedipus,” and “Man in a hole” are the most downloaded e-books and are more likely to lead to success. There is also a difference with the “U”-shaped emotional arc being more productive in the film industry. This phenomenon may suggest that the non-entertainment purpose of the serious narrative makes the desire for emotional expression different from the entertainment in novels and movies. Reading novels, serious narratives, and watching movies, users experience varying lengths of time. Reading a novel, a

serious narrative, and watching a film also take different lengths of time, with the average reader spending hours or even days reading a novel and watching a film. In our dataset, the average word count of serious narratives is 1300–1500 words, and readers only need 10–15 minutes to read these texts. Therefore, the serious narrative guided by “cinderella” is more popular than the movies and novels guided by the emotional arc of “cinderella” may be different from the fact that the background and the foreshadowing can be set up through large paragraphs in novels and movies. A serious narrative needs to express one’s mind directly, convey ideas, and achieve the purpose of conveying knowledge in a short period [24]. Under the influence of this reason, people are more willing to experience more emotional narratives with limited text and time. The protagonist’s bumpy experiences and ups and downs in the narrative can arouse the readers’ emotional resonance. In particular, the endings of the narrative under the emotional arc are more positive, and people are pleased to learn and appreciate the spirit and energy conveyed by the protagonist in such a short period.

5.2. The Influence of Selected Text Narrative Research on Multimedia Narrative. Due to the higher availability of the textual content of serious narratives in this study, serious narratives in micro-video, audio, and other media are not included. In the narrative environment of digital and multimedia fusion, the text is still the primary way of carrying narrative content in video, animation, games, virtual reality, and other forms of expression. Video and animation production requires scripts provided by text content, and the completion of games and virtual reality requires the world view and interaction logic of text content architecture. Text is a succession of several phases in the development of narrative material. The research on the emotional arc of textual story will serve as the basis for multimedia research, and utilize text to lead the production of future multimedia tales. The serious story is always about the transfer of information and wisdom, and the study of the emotional arc of serious text narrative as an example will become the emotional support of serious narrative, enabling it to enrich the media world with a completely different narrative approach [7].

5.3. Is it Valuable and Meaningful for the Emotional Arc to Intervene in Serious Interactive Narrative. The fundamental contribution of this research is to expand on earlier work in a way that enhances emotional perception, presenting a mechanism that allows individual users to develop growing serious stories through more favored emotional exchanges. This tailored method to story development offers them with an adaptable series of emotional experiences and learning events [25], allowing learners to become knowledge creators in the learning process rather than passive recipients of the information. Such a method can help groups with higher requirements for interactivity in the narrative and are more eager to acquire specific knowledge [26]. For example, it can help children

complete the learning of knowledge and the construction of self-cognition in the serious narrative of emotional perception.

6. Conclusion

Humans are part of the system in a serious interactive digital storytelling system, and knowledge creation and meaningful interaction are the main goals of human involvement in this system. This research focuses on the emotional arc in the narrative, a fundamental aspect of the story that evokes the audience’s cognition. Firstly, the preference degree of learners’ emotional arcs in a serious narrative is discussed. In general, learners prefer serious narratives with cheerful endings with twists and turns, and they have the highest preference for “N”-shaped emotional arcs. Nevertheless, when a serious narrative ends tragically, learners prefer a positive beginning regardless of the twists and turns. Our goal is to intervene in serious narratives with readers’ preferred emotional arcs and achieve serious contexts. Based on this research combining data science and humanistic perspectives, we propose a serious interactive digital narrative design strategy for emotion perception assisted by the artificial intelligence technology. The elements and methods of emotional arc intervention in a serious interactive narrative are discussed. It is expected to have particular guiding significance for creating a serious interactive digital narrative in the future. The main innovation of this research is to obtain readers’ preference for narrative emotional arcs in serious narratives through experiments and then use the experimental conclusions to guide the generation method of serious interactive narratives. It applies emotional arc in a narrative “forward and rear.” This paper proposes a feasible future way of serious interactive narrative driven by emotional arcs that combine existing technology and previous research in the strategy part. For future work, it is necessary to design a serious interactive digital narrative system through practice, and conduct more in-depth research on the effectiveness of serious interactive digital narrative with emotional perception by combining the pilot research methods of qualitative and quantitative measurement and more experimental evaluation.

Data Availability

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number (s) can be found below: <https://github.com/undo123/Serious-Narrative-Dataset>.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the Science and Technology Demonstration Project of Social Development of Jiangsu Province under Grant BE2019631.

References

- [1] K. J. Alexander, P. J. Miller, and J. A. Hengst, "Young children's emotional attachments to stories," *Social Development*, vol. 10, no. 3, pp. 374–398, 2001.
- [2] M. Gil and C. Sylla, "A close look into the storytelling process: the procedural nature of interactive digital narratives as learning opportunity," *Entertainment Computing*, vol. 41, Article ID 100466, 2022.
- [3] C. Roth and H. Koenitz, "Evaluating the user experience of interactive digital narrative," in *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, pp. 31–36, NewYork, NY, USA, October 2016.
- [4] H. Koenitz, "Towards a Specific Theory of Interactive Digital Narrative," *Interactive digital narrative*, Routledge, Milton Park, Abingdon-on-Thames, Oxfordshire, England, USA, pp. 91–105, 2015.
- [5] M. Theune, S. Rensen, R. op den Akker, D. Heylen, and A. Nijholt, "Emotional Characters for Automatic Plot creation," *International Conference on Technologies for Interactive Digital Storytelling and Entertainment*, vol. 3105, pp. 95–100, 2004.
- [6] L. Yao, N. Peng, R. Weischedel, K. Knight, D. Zhao, and R. Yan, "Plan-and-Write: towards better automatic storytelling," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7378–7385, 2019.
- [7] A. Lugmayr, E. Sutinen, J. Suhonen, C. I. Sedano, H. Hlavacs, and C. S. Montero, "Serious storytelling - a first definition and review," *Multimedia Tools and Applications*, vol. 76, no. 14, pp. 15707–15733, 2017.
- [8] M. O. Riedl and V. Bulitko, "Interactive narrative: an intelligent systems approach," *AI Magazine*, vol. 34, no. 1, 67 pages, 2013.
- [9] M. Riedl, C. J. Saretto, and R. M. Young, "Managing Interaction between Users and Agents in a Multi-Agent Storytelling environment," in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 741–748, NewYork, NY, USA, July 2003.
- [10] M. O. Riedl and R. M. Young, "Narrative planning: balancing plot and character," *Journal of Artificial Intelligence Research*, vol. 39, pp. 217–268, 2010.
- [11] M. O. Riedl, A. Stern, D. Dini, and J. M. Alderman, "Dynamic experience management in virtual worlds for entertainment, education, and training," *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning*, vol. 4, no. 2, pp. 23–42, 2008.
- [12] D. Thue, V. Bulitko, M. Spetch, and E. Wasylshen, "Interactive storytelling: a player modelling approach[C]//Proceedings of the aaai conference on artificial intelligence and interactive," *Digital Entertainment*, vol. 3, no. 1, pp. 43–48, 2007.
- [13] A. J. Ramirez and V. Bulitko, "Telling Interactive Player-specific Stories and Planning for it: ASD+ PaSSAGE= PAST," in *Proceedings of the Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*, Stanford, California, USA, October 2012.
- [14] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.
- [15] J. Sabourin, "Affective support in narrative-centered learning environments," *Affective Computing and Intelligent Interaction*, vol. 6975, pp. 280–288, 2011.
- [16] C. Grappiolo, Y.-G. Cheong, J. Togelius, R. Khaled, and G. N. Yannakakis, "Towards Player Adaptivity in a Serious Game for Conflict resolution," in *Proceedings of the 2011 Third International Conference on Games and Virtual Worlds for Serious Applications*, pp. 192–198, Athens, Greece, May 2011.
- [17] J. Chen, J. Chen, and Z. Yu, "Incorporating structured commonsense knowledge in story completion," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 6244–6251, 2019.
- [18] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds, "The emotional arcs of stories are dominated by six basic shapes," *EPJ Data Science*, vol. 5, no. 1, pp. 1–12, 2016.
- [19] M. Del Vecchio, A. Kharlamov, G. Parry, and G. Pogrebna, "Improving productivity in Hollywood with data science: using emotional arcs of movies to drive product and service innovation in entertainment industries," *Journal of the Operational Research Society*, vol. 72, no. 5, pp. 1110–1137, 2021.
- [20] F. Brahman and S. Chaturvedi, "Modeling Protagonist Emotions for Emotion-Aware storytelling," 2020, <https://arxiv.org/abs/2010.06822>.
- [21] Y. Sun, S. Wang, Y. Li et al., "Ernie 2.0: a continual pre-training framework for language understanding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 8968–8975, 2020.
- [22] A. Fan, M. Lewis, and Y. Dauphin, "Hierarchical Neural story generation," 2018, <https://arxiv.org/abs/1805.04833>.
- [23] N. Rao, S. L. Chu, R. W. Faris, and D. Ospina, "The Effects of Interactive Emotional Priming on Storytelling: An Exploratory study," *International Conference on Interactive Digital Storytelling*, vol. 11869, pp. 395–404, 2019.
- [24] E. M. Raybourn, "A new paradigm for serious games: transmedia learning for more effective training and education," *Journal of computational science*, vol. 5, no. 3, pp. 471–481, 2014.
- [25] A. Moradi-Karkaj, "Serious interactive digital narrative: explorations in personalization and player experience enrichment," in *Proceedings of the 2021 International Serious Games Symposium (ISGS)*, pp. 35–42, Tehran, Iran, November 2021.
- [26] M. Fan, J. Fan, S. Jin, A. N. Antle, and P. Pasquier, "EmoStory: A Game-Based System Supporting Children's Emotional Development," in *Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, NewYork, NY, USA, April 2018.

Research Article

An Intelligent System for Detecting Abnormal Behavior in Students Based on the Human Skeleton and Deep Learning

Yourong Ding , **Ke Bao** , and **Jianzhong Zhang**

Wuxi Institute of Technology, Wuxi, Jiangsu 214121, China

Correspondence should be addressed to Yourong Ding; dingyr@wxit.edu.cn

Received 25 May 2022; Revised 7 June 2022; Accepted 14 June 2022; Published 27 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yourong Ding 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 use of an intelligent video system, this research provides a method for detecting abnormal behavior based on the human skeleton and deep learning. To begin with, the spatiotemporal features of human bones are extracted through iterative training using the OpenPose deep learning network and the redundant information of human bone facial features is reduced in the feature extraction process, effectively reducing the time it takes to identify and analyze abnormal behavior. The collected human skeleton features are then classified using a graph convolution neural network to reduce the computational complexity of the behavior identification algorithm, and the sliding window voting method is used to further improve the accuracy of the behavior classification in practical application, resulting in the diagnosis and classification of abnormal behavior of students under video surveillance. Finally, using the self-built student trajectory data set and the INRIA data set, simulation analysis is performed, and the practicality and superiority of the proposed method for abnormal behavior detection is confirmed by comparing it to the existing abnormal behavior recognition methods. The proposed method for detecting anomalous behavior in a self-built database and INRIA data set has a high accuracy of more than 99.50 percent and a high processing efficiency rate.

1. Introduction

The society is developing rapidly, and the population is large and increasingly dense. Traffic accidents, fights, and other socially unstable incidents occur from time to time, and even terrorist attacks occur. Social security needs to be greatly enhanced. The number and coverage of surveillance cameras in transportation systems and public places are also increasing year by year, and cameras are basically found in every aspect of people's daily lives. In order to ensure people's safety, the camera can simulate human eyes so that they have the ability to "see." Computers simulate human brains with decision-making abilities. The computer obtains the video data through the surveillance camera for calculation and analysis, so as to understand the picture content in the surveillance scene, so as to realize the detection, identification, early warning, and alarm of abnormal behavior. As a kind of transportation equipment, escalators are widely used in shopping malls, office buildings, schools, and other public places to facilitate the people's travel. Especially

for students, as a specific group of the society, because of their own sense of autonomy and physical condition is still immature, a variety of hand lift safety accidents are prone to occur. At the same time of enhancing students' safety awareness, it is more necessary to monitor the escalator to stop the occurrence of safety accidents in time; in addition, by monitoring whether there are students on the escalator, it can also avoid no idling of the escalator, so as to save energy and prolong the life of the escalator, and realize fine management by counting the passenger flow of the escalator [1].

Intelligent video monitoring system (IVMS) has the characteristics of low cost, accuracy, and stability, and has been paid more and more attention in the field of public security [2, 3]. Passenger abnormal behavior recognition is an important application in IVMSs, which can detect and track moving targets through video sequences to analyze the target behavior [4–6], detect abnormal behavior fragments, and then identify abnormal behavior categories. When students take the escalator, abnormal behaviors such as

falling, climbing the handrail, probe, and hand probing can cause serious safety accidents. Therefore, it is of great significance to apply IVMs to accurately and stably identify various abnormal behaviors [7, 8].

Traditional abnormal behavior recognition methods such as hidden Markov models can only recognize specific actions in a single, simple environment, and are easily affected by environmental interference in a complex environment, which will reduce the recognition rate. Based on image acquisition technology and artificial intelligence technology, the collected images are input into a multi-layer network model composed of convolution for training, feature extraction of image signal data, and continuous learning based on their own network to improve student shape recognition performance and effectively guarantee the safety of students [9, 10].

2. Related Works

Due to a lack of safety awareness, students as a distinct group of society cause escalator accidents. Many researchers have conducted study on deviant behavior analysis when riding escalators as the precision of intelligent monitoring systems has improved and the maturity of image analysis algorithms has grown.

Traditional aberrant behavior recognition is constrained by ambient elements like light and shadow, and has issues like imprecise recognition and low processing efficiency [11]. The direction change of human ellipse fitting can be used to identify abnormal behavior in literature [12], but it can only be used in a simple environment; in literature [13], a Gaussian mixture model and filtering method are used to detect moving targets, extract fusion features, analyze target posture, and accurately recognize indoor falls and paralysis behaviors in real time. However, precisely modeling the background in complicated scenarios is difficult, lowering the identification rate. The use of recursive filtering to get target characteristics in literature [14] can solve the problem of difficult modeling of complex backgrounds, but the amount of computation is considerable and cannot match real-time requirements. Aberrant behaviors can be detected in real-time using the Hidden Markov model; however, the sorts of abnormal behaviors cannot be identified [15]. The human body is identified by filtering channel features and features are retrieved by Hough direction calculator to identify a variety of abnormal behaviors, according to the literature [16], but the abnormal behavior sequence must be segmented in preparation.

To recognize abnormal behavior, spatial and temporal features that can represent human motion can be retrieved from the original image data using deep learning theory [17–19]. The spatiotemporal point of interest feature [20], silhouette feature [21], optical flow feature [22], depth feature [23], and human two-dimensional skeleton feature are some of the most widely used features. Kinect [24], a prominent abnormal behavior analysis technology at the moment, can easily extract the two-dimensional skeleton of the human body. Through iterative training and learning, a

deep convolution neural network can efficiently extract feature information from the processing data set [25, 26] and realize behavior recognition and analysis. A dual residual convolutional network-based fall recognition algorithm was proposed in the literature [27]. The shallow and deep visual characteristics are fully integrated by nesting the residual network in the residual network, which reduces the impact of gradient disappearance during model training and improves the model's performance. Literature [28] calculates the optical flow field of sparse feature points using the Lucas–Kanade method, performs temporal and spatial filtering on the optical flow field, and detects anomalous behavior for the moving population using the graph convolutional neural network mode. A deep learning-based technique has been proposed in the literature [29]. The feasibility test was conducted using the VGG-16 model, which was trained on the open benchmark population data set. Through a cascaded network topology, literature [30] converts pretrained supervised FCN to unsupervised FCN based on convolution neural networks, which decreases the computational cost and enhances the real-time and accuracy of aberrant behavior detection.

The implementation of an intelligent video detection system and the use of intelligent approaches are critical for detecting inappropriate behavior in pupils when using escalators. However, contextual circumstances limit classic abnormal behavior, which has issues with identification accuracy and processing speed. This paper presents a method for detecting anomalous behavior in students based on the human skeleton and deep learning, based on previous anomaly detection research. The following are the major contributions:

- (1) The spatiotemporal properties of the human skeleton are retrieved using an OpenPose deep learning network to improve the accuracy and real-time of behavior recognition in escalator operation. The redundant information of face characteristics is eliminated during feature extraction, and the input original image is processed through the network to achieve end-to-end skeleton extraction results and effectively shorten the identification time.
- (2) This paper proposes a method based on graph convolution neural network to classify the collected human skeleton features, and uses sliding window voting method to further improve the classification accuracy in actual application, and finally realizes the video sequence diaphragm.

The remainder of this article is structured in the following manner. The second section introduces the abnormal behavior detection method's network model; the third section introduces the specific theoretical content and method implementation of the abnormal behavior detection method based on human skeleton and deep learning; the fourth section introduces the feasibility and optimality experimental simulation analysis of the proposed method using self-built and INRIA data sets; and the fifth section is the paper's conclusion.

3. The Proposed Model

Convolutional neural networks have difficulty extracting video features from huge numbers of frames and long-time sequences, but long-term and short-term memory networks have difficulty processing time sequence data in parallel and are slower. As a result, this article provides a skeletal action recognition model based on spatiotemporal relationship in order to better handle long-time video and meet real-time performance requirements, along with the characteristics of the two networks. The network may be used to recognize skeleton actions in long-term video and to recognize multi-person scenario behavior. The proposed method's flowchart is shown in Figure 1.

The abnormal behavior detection algorithm's ultimate purpose is to binary classify video sequences. First, the OpenPose pose estimation algorithm [31] extracts the 2D skeleton coordinates of the human body, and then the depth information of joint points is obtained using a monocular camera-based depth estimation method; then, the depth information and two-dimensional skeleton coordinates are combined to form three-dimensional skeleton data, and behavior recognition is performed using the skeleton data; finally, the skeleton recognition model is proposed based on the spatiotemporal relationship method.

4. Method and Implementation

4.1. Video Image Capture. The installation position of the escalator surveillance camera in a school is shown in Figure 2. Use a 3.6 mm focal length camera to shoot from diagonally above the escalator to ensure a clearer image.

The real video data are all scenes of passengers taking the escalator normally. There are many videos, and some video frames are intercepted for transfer learning of pedestrian detection models. The abnormal behavior was simulated by student volunteers on escalators in different scenes (air-floor, semi-outdoor). Affected by the camera's shooting angle and viewing angle, the maximum number of passengers in the escalator monitoring image is 5. The algorithm in this paper is not applicable to extreme situations with severe occlusion. For example, in the case of too many people, the passengers behind are blocked by a large area. And in a two-person scene, the person in front is relatively large, completely obscuring the person behind, etc. These situations will lead to missed detection of blocked passengers or most of the key point extraction results are missing (people who are not blocked in front have little influence). Therefore, in the volunteer simulation, this article only considers sparse scenes and crowded scenes where the occlusion is not serious. The videos simulated by the experimental volunteers in this article include 7 types of behaviors in different environments: standing normally, falling forward, falling backward, climbing the handrail, extending the hand to the escalator, and leaning against the handrail. Environmental variables are light intensity and passenger density.

The movie is initially divided into 1613 segments for the anomalous behavior data set. Each segment lasts 20–30 seconds and covers the entire process of passenger behavior

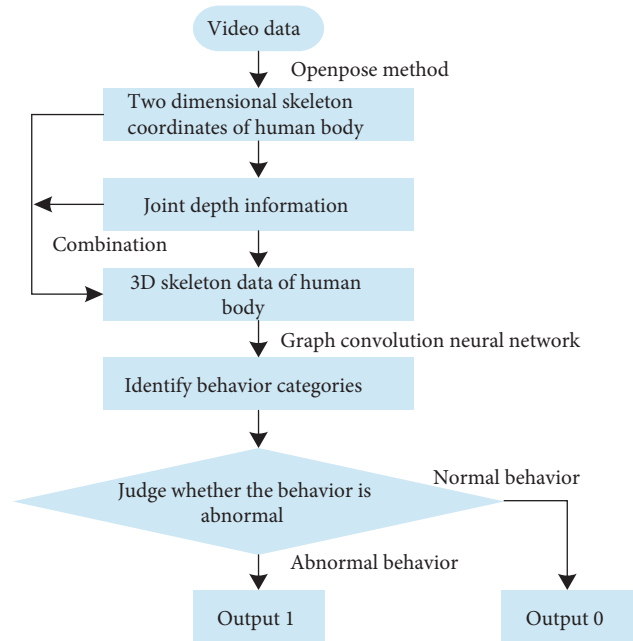


FIGURE 1: Flowchart of the proposed method.

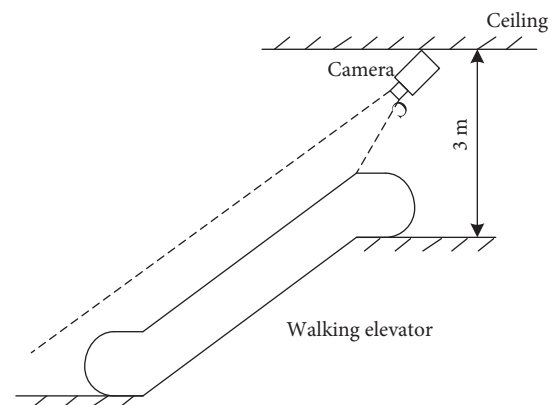


FIGURE 2: Installation diagram of the camera.

in various contexts. The short video is then separated into a training set and a validation set based on behavior and environmental characteristics in a 3 : 1 ratio. The training set and verification set of the graph convolutional neural network are then recovered from the key behavior frames. In this approach, the operation of dividing the video data set first and then capturing the picture is compared to capturing the image first and then dividing the image data set, which can prevent using the same short video for both the training and verification sets. It also guarantees that the training model does not overfit the validation data.

4.2. OpenPose Deep Learning Network. According to the method of skeleton extraction, the skeleton extraction network can be divided into top-down and bottom-up extraction. A human body detector must be used to determine the position of the human body in order to extract the

skeleton from top to bottom. The skeleton is then extracted by detecting key points of the human body in each human body area. This method relies on the human body detector's performance, and the speed of skeleton extraction slows dramatically as the number of people in the image grows. The bottom-up skeleton extraction does not require the detection of the human body, instead detecting all of the key points in the image directly. Then, using the same person's key points, create a human body skeleton. This method's skeleton extraction speed is unaffected by the number of people present, and the skeleton can be extracted quickly even when there are many. However, determining the relationship between the key points and the human body to which they belong is difficult. OpenPose presents Part Affinity Fields (PAFs) to communicate the relevant information between key points of the human body and the human body to which it belongs as a solution to this challenge. Each pixel corresponds to a two-dimensional vector in the PAFs, which are the same size as the original image. By connecting two adjacent key points in a straight line, you can encode the position and direction of the torso. The likelihood that the two key points can be joined to generate a human body torso is then calculated by adding the inner product of the PAFs vector and the connecting vector of all pixels on the segment connected by any two key points. The foundation for subsequent abnormal behavior detection and recognition is accurate, real-time, and stable skeleton extraction. Deep learning methods extract human skeletons more accurately and consistently than traditional image processing or machine learning methods. It creates a skeleton extraction network by iterative training, processes the input original images through the network, and outputs the skeleton extraction results from start to finish. Skeleton extraction networks have been regularly enhanced and put forward one after another as deep learning technologies have progressed. The OpenPose deep learning network used in this paper is one of them. It is a deep learning network that considers real-time performance and can accurately and consistently extract the human skeleton. It is the standard skeleton extraction network at this time, and it is widely used in the engineering area.

The network structure of OpenPose is shown in Figure 3. First, the first 10 layers of the VGG network are used as a pretrained convolutional neural network to generate a feature map set F . Then, input it into two branch networks, each branch network contains T stages. Each stage t of the first branch outputs a set of key point confidence maps S^t . Each key point confidence map is a heat map corresponding to the key points of the human body, which is the same size as the original image. Each pixel value represents the confidence that the point belongs to the corresponding key point. Each stage t of the second branch outputs a set of PAFs map L^t , corresponding to each segment of the human torso connected by key points. The input of the first stage is F and the output is S^1 and L^1 . Starting from the second stage, the input of each stage t is the fusion feature map of F and the previous stage S^{t-1} and L^{t-1} , and the output of S^t and L^t .

At each stage, calculate the L_2 norm of S^t , L^t and S^* , L^* as the loss function. Here, S^* and L^* are the real key point

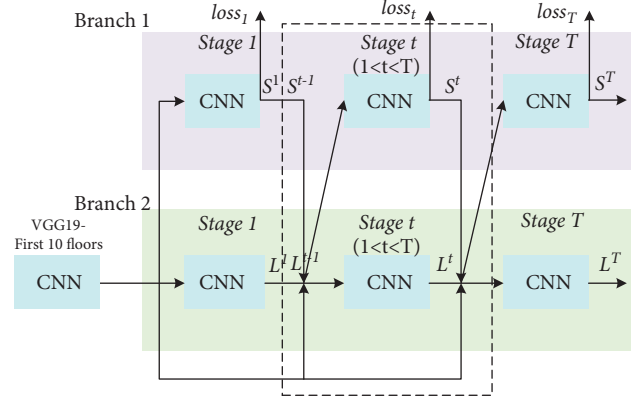


FIGURE 3: OpenPose network structure.

confidence map and real PAFs. Using the real label data, calculate according to

$$S_{j,k}^*(p) = \exp\left(\frac{-|p - x_{j,k}|_2^2}{\sigma}\right), \quad (1)$$

$$S_j^*(p) = \max_k S_{j,k}^*(p), \quad (2)$$

$$v = \frac{(x_{j_2,k} - x_{j_1,k})}{|x_{j_2,k} - x_{j_1,k}|_2}, \quad (3)$$

$$L_{c,k}^*(p) = \begin{cases} v, 0 \leq v \cdot (p - x_{j_1,k}), \\ \leq l_{c,k} \cup |v_{\perp} \cdot (p - x_{j_1,k})| \leq \sigma_{c,k}, \\ 0, \text{ otherwise,} \end{cases} \quad (4)$$

$$L_c^*(p) = \sum_k L_{c,k}^*((p)/n_c(p)), \quad (5)$$

where $x_{j,k}$ and $S_{j,k}^*(p)$ are the real position of the j key point of the k th person and the real confidence of the pixel point p , respectively. σ controls the smoothness of the distribution. $L_{c,k}^*(p)$, $l_{c,k}$, and $\sigma_{c,k}$ are the PAFs vector, torso length, and width of the torso of the k th person's section c , respectively. v and v_{\perp} are the torso unit vector and the vertical unit vector, respectively. $n_c(p)$ is the number of people with non-zero $L_{c,k}^*(p)$. Accumulate all stages to obtain the total loss function. Continuously optimize the total loss function through iterative training until the model converges to obtain the final network model. The network output is the J key point confidence level and the C segment trunk PAFs graph. The key points can be used as nodes in the bipartite graph, and the possibility E of connecting the two key points d_{j_1} and d_{j_2} into the trunk can be calculated according to

$$E = \int_{u=0}^{u=1} L_c((1-u)d_{j_1} + ud_{j_2}) \cdot \frac{d_{j_2} - d_{j_1}}{\|d_{j_2} - d_{j_1}\|_2} du. \quad (6)$$

Then, using E as the corresponding edge weight, the problem of optimal connection of key points is transformed into the problem of optimal bipartite graph matching.

4.3. Human Body Two-Dimensional Skeleton. The COCO training data set is used to train the OpenPose deep learning network in this article. The output skeleton extraction result is the two-dimensional coordinate locations (x, y) and confidence c of the 18 human body key points that make up the skeleton. The value of c ranges from 0 to 1. The 18 key points are the nose, neck, right shoulder, right elbow, right wrist, left shoulder, left elbow, left wrist, right marrow, right knee, right ankle, left marrow, left knee, left ankle, right eye, left eye, right ear, left ear, right eye, left eye, right ear, left ear, right eye, left eye, and right ear. Head movements can be represented by the key points of the left and right eyes, left and right ears, and nose. On the basis of the above skeleton extraction results, this study discards the key points of the left and right eyes and left and right ears in order to remove redundant information, leaving only the key points of the nose. The nose, neck, right shoulder, right elbow, right wrist, left shoulder, left elbow, left wrist, right marrow, right knee, right ankle, left marrow, left knee, and left foot are all included in the skeleton extracted in this paper. They also connect the 13 torso parts.

4.4. Optimize Joint Depth Information. The abnormal behavior skeleton sequence is the identification object for students' aberrant behavior recognition. To produce the abnormal behavior skeleton sequence, it is necessary to detect the abnormal behavior skeleton from the passenger human skeleton sequence and merge them in chronological order. When riding an escalator, travelers normally stand on the escalator with their hands on their sides and their heads up to look forward. It has distinct traits as compared to abnormal conduct. As a result, different passengers with varied distances are picked in several operating phases of diverse escalator environments to create 20 normal behavior templates based on the features of normal behavior. The skeletons in the passenger human skeleton sequence are template-matched, and anomalous behavior skeletons in the skeleton sequence are discovered.

In order to adapt to the size changes caused by the distance of the human body and the differences of individual body types, when performing template matching, the human body posture feature vectors of the passenger skeleton and the template skeleton are extracted, respectively. Then, the matching similarity between the two is calculated based on the Euclidean distance of the vector [32]. If the matching similarity between the passenger skeleton and all template skeletons is greater than the normal threshold, it is judged as a normal behavior skeleton. Otherwise, it is judged as an abnormal behavior skeleton. When calculating the human body pose feature vector of the skeleton, the 13 bones of the human skeleton are regarded as a sequence $\{J^1, J^2, \dots, J^{13}\}$ containing 13 two-dimensional vector elements. Where J^i is the i -th bone formed by connecting the starting joint point B^i and the ending joint point E^i . The starting point of the bone vector is (B_x^i, B_y^i) and the confidence is C_B^i . The end point coordinates are (E_x^i, E_y^i) and the confidence level is C_E^i . The horizontal direction angle is α^i , and the vertical direction angle is β^i . Figure 4 is a schematic diagram of the human

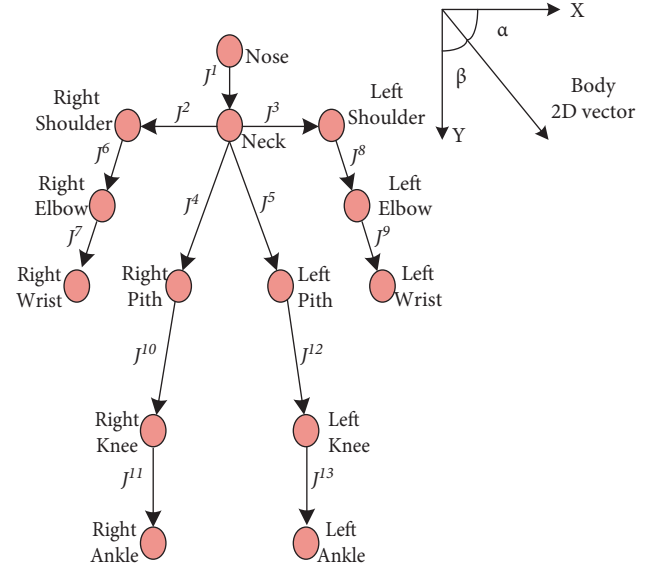


FIGURE 4: Vector diagram of human skeleton.

skeleton bone vector. The bone vector is denoted as $(E_x^i - B_x^i, E_y^i - B_y^i)$. The horizontal cosine value and the vertical cosine value are, respectively,

$$\begin{cases} \cos \alpha^i = \frac{(E_x^i - B_x^i)}{\sqrt{(E_x^i - B_x^i)^2 + (E_y^i - B_y^i)^2}}, \\ \cos \beta^i = \frac{(E_y^i - B_y^i)}{\sqrt{(E_x^i - B_x^i)^2 + (E_y^i - B_y^i)^2}}. \end{cases} \quad (7)$$

Calculate the horizontal and vertical cosine values of 13 bone vectors in sequence, and arrange to obtain a 26-dimensional feature vector $(\cos \alpha^1, \cos \beta^1, \dots, \cos \alpha^{13}, \cos \beta^{13})$. And use it as the human body posture feature, and then calculate the matching similarity $O(S_D, S_T)$ between the skeleton S_D to be matched and the template skeleton S_T as

$$O(S_D, S_T) = \exp \left(\sqrt{\sum_{i=1}^{13} \zeta \left((\cos \alpha_D^i - \cos \alpha_T^i)^2 + (\cos \beta_D^i - \cos \beta_T^i)^2 \right)} \right). \quad (8)$$

Here, $\zeta_i = C_{B,D}^i + C_{E,D}^i + C_{B,T}^i + C_{E,T}^i$ is the confidence coefficient of the i segment bone; $\cos \alpha_D^i, \cos \beta_D^i$, and $C_{B,D}^i, C_{E,D}^i$ is the direction cosine value and the end point confidence of the i -th segment of the skeleton to be matched; $\cos \alpha_T^i, \cos \beta_T^i$, and $C_{B,T}^i, C_{E,T}^i$ is the direction cosine value and the end point confidence of the i segment bone of the template skeleton.

4.5. Abnormal Behavior Recognition. Based on the above graph convolution operation, a graph convolution neural network for passenger behavior recognition can be constructed, and its structure is shown in Figure 5. Here, cn ($n \in \mathbb{Z}$) means that the number of channels is n . First, the

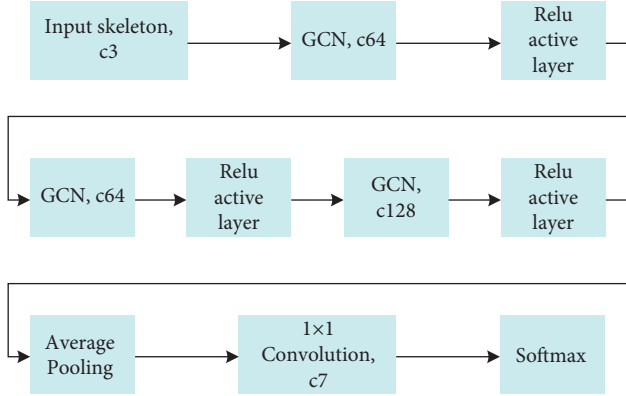


FIGURE 5: Structure of behavior recognition neural network.

coordinates and confidence of 14 key points are connected into a 3-channel graph through human bones as the input skeleton. After the input skeleton undergoes 3 times of graph convolution and ReLU activation function, the depth map features of 128 channels are extracted. Then, perform global average pooling on each channel, and then reduce the number of channels to 7 through 1×1 convolution. Finally, the probability of the occurrence of seven passenger behaviors is returned through the Softmax layer.

The behaviors of students when riding the escalator are divided into 7 types of behaviors: normal standing, falling forward, falling backward, climbing the handrail, reaching out the escalator, reaching out the escalator, and leaning against the handrail. Other behaviors can be classified into the above 7 categories. At time t , the detected human skeleton coordinates and confidence are used as the input skeleton diagram in Figure 6. After passing through the network in Figure 6, the behavior with the highest probability is selected as the output. Suppose that the skeleton of the k th person at time t is determined to be the behavior $B_t(k)$ after the behavior recognition neural network. In practical applications, due to interference factors such as illumination and occlusion, there will be noise in the extraction of individual frame skeletons, leading to incorrect behavior classification. Therefore, if $B_t(k)$ is output as the final decision-making behavior, the recognition rate will be greatly reduced. Because the behavior of passengers on the escalator often lasts for a period of time (ranging from more than ten frames to more than a hundred frames, most of the behavior decision result $B(k)$ of the k th passenger during this period is the same behavior, but there is noise). Therefore, this paper uses the sliding window voting method to count the multi-frame behavior classification result $B(k)$ of each passenger to obtain the final behavior decision result of the passenger. This can effectively reduce the classification errors caused by skeleton noise.

The length of the sliding window is preset to T . For all passengers k of sequence length $|B(k)| \geq T$, their behavior decisions are as follows: Take the behavior of the most recent T times (i.e., $(t - T, t]$ interval) for voting analysis. Suppose the number of votes for 7 behaviors is $d_1 - d_7$, $d_1 + d_2 + \dots + d_7 = T$. If the maximum number of votes is greater than the set threshold T_{th} ($T_{th} < T$), it can be

determined that the behavior has occurred. The statistical formula for sliding window voting is as follows:

$$\text{action}_t(k) = \begin{cases} \arg \max(d_1, d_2, \dots, d_7), \\ \max(d_1, d_2, \dots, d_7) > T_{th}, \\ \text{action}_{t-1}(k), \text{ others.} \end{cases} \quad (9)$$

The sliding window voting method greatly improves the classification accuracy of behavior in practical applications by slightly sacrificing the detection time [33], which has the effect of a low-pass filter. High-frequency noise caused by behavior recognition errors in individual frames can be filtered out. When $T = 10$, $T_{th} = 5$ achieve the best results.

In actual application scenarios, there may be serious occlusion due to crowding. At this time, when the algorithm in this paper uses GCN for forward inference, it needs to filter out some severely occluded skeletons. Only the skeletons whose key point confidence sum $\sum_{k=1}^{14} P_c^k$ exceeds the threshold P_c^T are used for behavior prediction. The skeleton with a confidence lower than P_c^T has low reliability due to occlusion, so its behavior recognition is not performed. Good results are achieved when $P_c^T = 5$ is in the text. The passengers behind were severely obscured, and even only one head was exposed. If this kind of uncertain noise is input into GCN, random behavior recognition results will be obtained. Because during training, such noise samples are not trained, and this noise cannot be labeled as a certain type of behavior. Therefore, it can only be eliminated in training and actual application scenarios at the same time, and behavior recognition is not performed on it.

5. Experiments

Experiments on the Windows 10 platform using MATLAB are carried out to validate the feasibility and effectiveness of the suggested strategy (R2016a). The video files were shot with a Canon HF R806 megapixel digital camera that has a resolution of 350 320 pixels and a frame rate of 32 frames per second. The footage is then fed into the regular CAMS algorithm and the new tracking system to see how well it detects and recognizes the objects.

The self-built data set and INRIA pedestrian data set described in Section 3 of this work are the simulated data sets.

Dalal et al. compiled people's images from photographs and videos into the INRIA data collection, which is currently the most extensively utilized. In the INRIA data set, the majority of pedestrians are standing. The most notable feature of the INRIA pedestrian data set is its complex background, which poses a significant challenge to researchers studying pedestrian detection. Because the image in the INRIA pedestrian data set is so similar to a genuine situation, it is frequently used to train real-world detection models.

The INRIA data collection divides the training and verification sets by providing the original image and the relevant annotation information. 614 pedestrian photos (a total of 2416 pedestrians) and 1218 backdrop images make up the training set. 288 pedestrian photos (1126 total

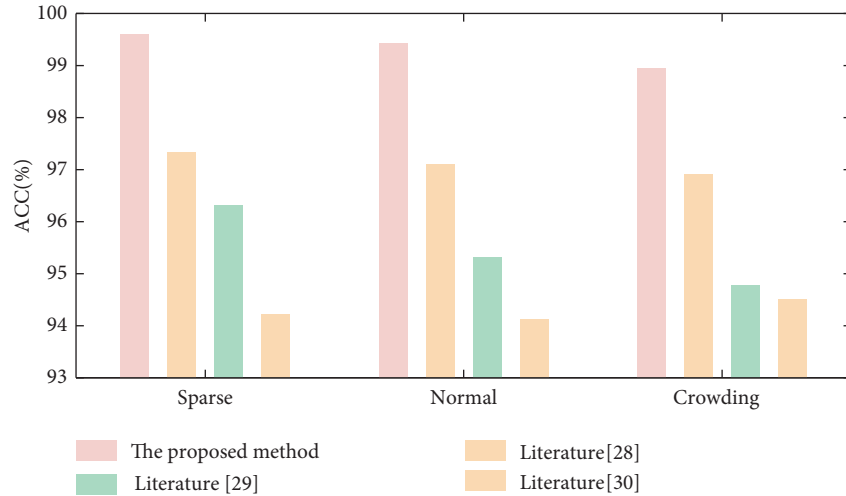


FIGURE 6: The accuracy of different methods in different scenarios.

pedestrians) and 458 backdrop images make up the verification set. The majority of the labeled pedestrian detection frames have a height of over 100 pixels, with a width-to-height ratio of 0.25–0.5.

5.1. Simulation Analysis of Self-Built Data Set

5.1.1. Experimental Results of Student Trajectory Construction. Starting from the students entering the escalator monitoring area, the construction of students' track is stopped after they leave the escalator or have abnormal behavior. The number of frames of students in this period is N_p . The accuracy rate PR , recall rate RE , and harmonic mean $F1$ ($F1$ score) of reference [33, 34] were used to analyze the effect of trajectory construction, in which $PR = (TP / (TP + FP))$, $RE = (TP / (TP + FN))$ and $F1 = (2TP / (2TP + FP + FN))$. If the IOU (intersection and union ratio) of the target tracking frame and the real marker frame is greater than 0.6, the target is considered to be successfully tracked in that frame. If the number of consecutive tracking frames of the student exceeds $0.95 N_p$, the student track is successfully constructed and the number of successful frames is recorded as TP . Otherwise, the target is missed and the number of missed frames is recorded as FN . At the same time, the number of wrong tracking frames is recorded as FP and the tracking speed is recorded as TI . Because the escalator is located outdoors, the light intensity will cause insufficient illumination. If the light is not uniform, there will be shadows and crowding when students overlap in the image. The above factors will affect the construction of students' trajectory. Table 1 shows the performance index of student trajectory construction in self-built data set.

The performance index of student trajectory construction shows that the algorithm can continuously track the students who appear in the escalator monitoring area under different lighting conditions, different crowding degree, and with/without shadow. The harmonic mean value of this algorithm is more than 92%, and the average harmonic mean

value of the aforementioned cases is 95.96%. This algorithm has the best performance in the environment of sufficient illumination, sparse students and no shadow, with a success rate of 99.50%.

The results show that the success rate is reduced by 1.19% and the tracking speed is reduced by 0.21 frames per second when other conditions are the same, which shows that the algorithm can effectively resist the global environmental disturbance caused by the change of light intensity. The success rate caused by shadow is reduced by 2.18%, and the tracking speed is reduced by 0.39 frames per second, which shows that the algorithm is more sensitive to shadows that cause local environment changes than light intensity. The success rate is reduced by 3.71%, and the tracking speed is reduced by 1.21 frames per second, which shows that the congestion caused by local face occlusion is the biggest reason for the performance degradation of the algorithm.

It is worth noting that, even in the case of insufficient light, crowded students and shadows, the success rate of student trajectory construction is maintained at 92.42%, which indicates that the algorithm can construct student trajectory robustly and stably in different environments. This lays a good foundation for the detection of students' abnormal behavior.

5.1.2. Experimental Results and Analysis of Students' Abnormal Behavior Recognition. According to the above five kinds of abnormal behaviors, the number of students' abnormal behaviors in the experimental video is counted as TG , and the number of successful detection through the abnormal behavior detection is recorded as TP . The recall rate $RE1 = (TP / TG)$ of literature [34] was used as the performance index to analyze the detection effect of each abnormal behavior. It represents the proportion of abnormal behavior skeleton sequence detected by abnormal behavior detection in the total abnormal behavior. The confusion matrix of five kinds of abnormal behavior skeleton sequences is used to analyze the classification effect of the successfully detected abnormal behavior skeleton sequences.

TABLE 1: Performance index of student trajectory construction.

Influence factor	TP	FN	FP	F1 (%)	TI/(frame × second ⁻¹)
Sufficient light, crowding, no shadow	165	7	10	95.79	26.71
Sufficient light, crowding, shadow	140	9	15	93.61	26.32
Sufficient light, sparse, no shadow	172	1	3	99.5	27.92
Sufficient light, sparse, shadow	148	5	7	97.32	27.53
Insufficient light, crowding, no shadow	136	7	14	94.6	26.5
Insufficient light, crowding, shadow	140	9	15	92.42	26.11
Insufficient light, sparse, no shadow	162	3	8	98.31	27.71
Insufficient light, sparse, shadow	169	9	11	96.13	27.32
Average				95.96	27.015

TABLE 2: Confusion matrix of the classification results of the 5 kinds of abnormal behavior skeleton sequences.

The real situation	Forecast results				
	Fall forward	Fall back	Climbing	Probe	Explore the hand
Fall forward	191	10	1	5	2
Fall back	4	138	0	2	1
Climbing	6	1	131	1	1
Probe	1	0	0	231	0
Explore the hand	1	0	0	1	199

In the confusion matrix, the number of abnormal behavior prediction results consistent with the real situation is recorded as TR, and the recall rate is recorded as $RE2 = (TR/TP)$. It represents the proportion of the abnormal behavior skeleton sequence which is successfully identified by this method in the abnormal behavior skeleton sequence. Finally, the recognition accuracy is defined as the performance index to analyze the recognition effect of the algorithm. $ACC = RE1 \times RE2 = (TR/TG)$ indicates the possibility of correctly identifying the type of abnormal behavior from the skeleton sequence of total abnormal behavior. Table 2 is the confusion matrix of the classification results of the 5 kinds of abnormal behavior skeleton sequences, and Table 3 is the performance of abnormal behavior recognition.

The algorithm can accurately recognize a range of abnormal behaviors in the process of students taking the escalator, according to the results of students' abnormal behavior recognition and performance indicators, with a total recognition accuracy of 93.2 percent. The recognition accuracy of hand probing, probe, ascending, back falling, and front falling is strong, and the recall rate of five kinds of abnormal behavior is about 96 percent, according to the analysis of performance indicators of students' abnormal behavior recognition.

The deep learning algorithm has difficulty detecting anomalous behavior in an escalator scene from beginning to end. Three existing end-to-end abnormal behavior recognition methods are utilized to examine the abnormal behavior in this study [28, 29], and [30], and the recognition results are compared with those of the algorithm in this work. The comparison results of various approaches are shown in Figure 7.

The testing results demonstrate that the abnormal behavior recognition algorithm based on human skeleton sequence has faster operation time and greater recognition

TABLE 3: Performance of abnormal behavior recognition.

Abnormal behavior	TG	TP	TR	RE1 (%)	RE2 (%)	ACC (%)
Fall forward	201	192	171	97.23	95.21	90.4
Fall back	154	132	131	97.57	94.67	93.3
Climbing	123	131	121	97.12	96.72	94.6
Probe	241	211	195	97.51	99.12	96.2
Explore the hand	198	192	216	97.69	99.68	96.8
Total	917	858	834	97.42	97.08	94.3

accuracy than the abnormal behavior recognition algorithm based on single frame image. The sliding window voting approach considerably enhances the classification accuracy of behavior in practical applications by somewhat sacrificing the detection time, which has the effect of a low-pass filter. High-frequency noise caused by behavior recognition failures in individual frames can be filtered out. This approach does not need to develop a classifier or sophisticated model, so the running time is faster; at the same time, compared with the single frame behavior, the behavior sequence can better explain the aberrant behavior of students, so the identification rate of abnormal behavior is greater.

5.2. Simulation Analysis of INRIA Data set. The simulation experiment first divides the used INRIA data set into 3 different crowded scenes, and detects abnormal behaviors in 3 different scenes, respectively. The proposed algorithm is compared with literature [28], literature [29], and literature [30] in three scenarios, respectively. The experimental results of different methods in each crowded scene are shown in Figure 6.

The suggested method has a greater effect of identifying abnormal behavior in diverse settings, as shown in Figure 6. The accuracy rate of various congestion scenarios is above

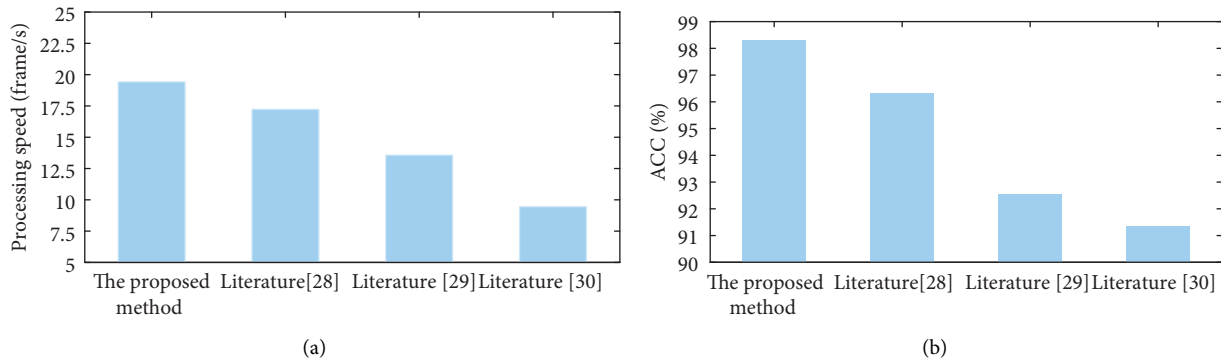


FIGURE 7: Comparison results of various methods. (a) Processing speed of each method. (b) The accuracy of each method.

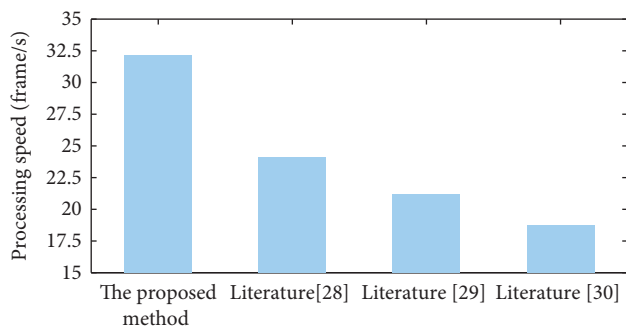


FIGURE 8: Processing speed of each method for INRIA data set.

99.5 percent, which is clearly superior to other methods. Because the suggested method aims to reduce the impact of background features on positioning accuracy, this is the case. The method described in this work describes the feature difference between the target object and the background, and it increases the target object's tracking accuracy. GCN, on the other hand, can better characterize the passenger's behaviors and provide a higher behavior recognition rate because it uses the important points of the human body and their relationships as the graph's input. Furthermore, the sliding window voting statistics method has an effect on the recognition accuracy's further increase.

Literature [28], literature [29], literature [30], and the method of this paper are also employed to classify passenger behavior on the same short video data set, with the results displayed in Figure 8. As shown in the image, the behavior recognition algorithm based on video surveillance suggested in this research has a processing speed of 32.2 frames per second, which is faster than the other methods. Because this paper compares the literature [29] VGG-16 and the literature [30] FCN method, the GCN algorithm has the advantage of fewer network layers. The GPU is also employed for graph convolution forward inference, which ensures that the anomalous activity is detected quickly. Demonstrate that its graph convolutional neural network application is feasible. During the feature extraction procedure, the features of the human skeleton model are integrated and simplified at the same time. As a result, the comparative literature [28] has an advantage in terms of speed. In conclusion, the

abnormal behavior identification approach in this study, which is based on human bones and a graph convolutional neural network, can increase the efficiency of the detection process behavior while maintaining the accuracy.

6. Conclusion

The traditional, limited environmental variables of abnormal behavior have the drawbacks of low recognition accuracy and processing speed. As a solution to this issue, the author of this research suggests a method of student anomalous behavior detection that is based on deep learning and the human skeleton. Iterative training is used in this technique, which is based on the OpenPose deep learning network [35]. The goal of the technique is to extract the spatiotemporal properties of human bones. This will improve the effectiveness of the identification and analysis of abnormal behaviors. In addition, on the basis of the graph convolutional neural network, the features of the acquired human skeleton are categorized properly, and this helps to reduce the amount of calculation that is required by the behavior recognition algorithm. Continue to increase the categorization accuracy of actions in practical applications, and strive to achieve efficient recognition of anomalous behaviors exhibited by pupils while they are being filmed. Based on the analysis of the results of the experiments, it has been determined that the suggested technique is capable of maintaining an accuracy of aberrant behavior identification of self-built databases and INRIA data sets that is greater than 99.50 percent and possesses outstanding processing efficiency.

However, the scene of the self-built data set of this system is relatively single, and the sample size is small, so we need to continue to expand the data set, collect training samples from different environmental conditions, and for the task of passenger abnormal behavior recognition, we need to collect more abnormal behaviors to increase the diversity of samples. With the continuous advancement of national modernization and intelligence, the escalator intelligent monitoring video system with many advantages will play an increasingly important role in the field of public security. Future research will focus on the platform of the proposed method, and strive to achieve the

commercialization of the proposed method. The focus of future research will be to explore the platformization and to realize the commercialization of the proposed method.

Data Availability

The data sets 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 in part by the Jiangsu Education Science “13th Five Year Plan” Project (Research on the Construction Path of Smart Campus under the Background of Double Universities-Research on Micro Service Architecture Based on the Theory of Middle Platform) under grant B-B/2020/03/29.



References

- [1] Y. Qi, P. Lou, J. Yan, and J. Hu, “Surveillance of abnormal behavior in elevators based on edge computing,” in *Proceedings of the The Second International Conference on Image, Video Processing and Artificial Intelligence*, Bellingham, WA, U.S.A, November 2019.
- [2] A. Ben Mabrouk and E. Zagrouba, “Abnormal behavior recognition for intelligent video surveillance systems: a review,” *Expert Systems with Applications*, vol. 91, no. 1, pp. 480–491, 2018.
- [3] Z. Jin, W. Cheng, W. Yiming, and P. Wang, “Detection of abnormal behavior in narrow scene with perspective distortion,” *Machine Vision and Applications*, vol. 30, no. 5, pp. 987–998, 2018.
- [4] R. A. Shatalin, V. R. Fidelman, and P. E. Ovchinnikov, “Abnormal behaviour detection method for video surveillance applications,” *Computer Optics*, vol. 41, no. 1, pp. 37–45, 2017.
- [5] R. A. Shatalin, V. R. Fidelman, and P. E. Ovchinnikov, “Abnormal behavior detection based on dense trajectories,” *Computer Optics*, vol. 42, no. 3, pp. 476–482, 2018.
- [6] M. George, B. R. Jose, and J. Mathew, “Abnormal activity detection using shear transformed spatio-temporal regions at the surveillance network edge,” *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 37–38, 2020.
- [7] J. Zhang, C. Wu, and Y. Wang, “Human fall detection based on body posture spatio-temporal evolution,” *Sensors*, vol. 20, no. 3, pp. 1–21, 2020.
- [8] Z. Huang, Q. Niu, and S. Xiao, “Human behavior recognition based on motion data analysis,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 9, pp. 1–13, 2019.
- [9] S. Xu, E. S. L. Ho, N. Aslam, and H. P. H. Shum, “Unsupervised abnormal behaviour detection with overhead crowd video,” in *Proceedings of the 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, Malabe, Sri Lanka, December 2017.
- [10] N. Marir, H. Wang, G. Feng, B. Li, and M. Jia, “Distributed abnormal behavior detection approach based on deep belief network and ensemble SVM using spark,” *IEEE Access*, vol. 6, no. 1, pp. 59657–59671, 2018.
- [11] O. P. Popoola and K. Kejun Wang, “Video-based abnormal human behavior recognition-A review,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 865–878, 2012.
- [12] A. K. S. Kushwaha, S. Srivastava, and R. Srivastava, “Multi-view human activity recognition based on silhouette and uniform rotation invariant local binary patterns,” *Multimedia Systems*, vol. 23, no. 4, pp. 451–467, 2017.
- [13] H. Rajabi and M. Nahvi, “An intelligent video surveillance system for fall and anesthesia detection for elderly and patients,” *IEEE*, in *Proceedings of the International Conference on Pattern Recognition & Image Analysis*, pp. 1–6, Rasht, Iran, March 2015.
- [14] D. Nehab and A. Maximo, “Parallel recursive filtering of infinite input extensions,” *ACM Transactions on Graphics*, vol. 35, no. 6, pp. 1–13, 2016.
- [15] T. Fuse and K. Kamiya, “Statistical anomaly detection in human dynamics monitoring using a hierarchical dirichlet process hidden Markov model,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3083–3092, 2017.
- [16] T. Wang, Q. Li, Y. Liu, and Y. Zhou, “Abnormal human body behavior recognition using pose estimation,” *Chinese Journal of Scientific Instrument*, vol. 37, no. 10, pp. 2366–2372, 2016.
- [17] C. Yuan, X. Li, W. Hu, H. Ling, and S. Maybank, “3D R transform on spatio-temporal interest points for action recognition,” in *Proceedings of the 2013 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 724–730, IEEE, Piscataway, NJ, U.S.A, January 2013.
- [18] M. Ahmad and S. W. Lee, “Variable silhouette energy image representations for recognizing human actions,” *Image and Vision Computing*, vol. 28, no. 5, pp. 814–824, 2010.
- [19] C. D. Geddes, P. Douglas, C. P. Moore, T. J. Wear, and P. L. Egerton, “A compact optical flow cell for use in aqueous halide determination,” *Measurement Science and Technology*, vol. 10, no. 4, pp. N34–N37, 1999.
- [20] H. Rabiee, H. Mousavi, M. Nabi, and M. Ravanbakhsh, “Detection and localization of crowd behavior using a novel tracklet-based model,” *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 12, pp. 1999–2010, 2018.
- [21] Y. Iwashita, S. Takaki, K. Morooka, T. Tsuji, and R. Kurazume, “Abnormal behavior detection using privacy protected videos,” in *Proceedings of the Fourth International Conference on Emerging Security Technologies*, pp. 55–57, IEEE, Cambridge, U.K, September 2013.
- [22] J. Liu, H. Tao, L. Luo, L. Zhao, and C. Zou, “Video image abnormal behavior detection algorithm based on gradient histogram and optical flow feature fusion,” *Signal Processing*, vol. 32, no. 1, pp. 1–7, 2016.
- [23] F. Hui, N. Peng, S. Jing, Q. Zhou, and S. Jia, “Clustering and anomaly detection method of driving behavior based on aggregation hierarchy,” *Computer Engineering*, vol. 44, no. 12, pp. 196–201, 2018.
- [24] A. Franco, A. Magnani, and D. Maio, “A multimodal approach for human activity recognition based on skeleton and RGB data,” *Pattern Recognition Letters*, vol. 131, no. 1, pp. 293–299, 2020.
- [25] K. Pawar and V. Attar, “Deep learning approaches for video-based anomalous activity detection,” *World Wide Web*, vol. 22, no. 2, pp. 571–601, 2019.
- [26] E. K. Kwang and K. B. Sim, “Deep convolutional framework for abnormal behavior detection in a smart surveillance

- system,” *Engineering Applications of Artificial Intelligence*, vol. 67, no. 1, pp. 226–234, 2018.
- [27] X. Wang, L. Xie, and L. Peng, “Double residual network recognition method for abnormal fall behavior,” *Computer science and exploration*, vol. 14, no. 09, pp. 1580–1589, 2020.
- [28] X. Hu, C. Yi, Q. Chen, X. Chen, and L. Chen, “Abnormal behavior detection based on motion saliency map,” *Computer applications*, vol. 38, no. 04, pp. 1164–1169, 2018.
- [29] W. Ullah, F. U. M. Ullah, and S. W. Baik, “Crowd behavior detection using convolutional neural network,” *The Journal of Korean Institute of Next Generation Computing*, vol. 15, no. 6, pp. 7–14, 2019.
- [30] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, “Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes,” *Computer Vision and Image Understanding*, vol. 172, no. 1, pp. 88–97, 2018.
- [31] Z. Bin, X. Ying, L. Guohu, and L. Chen, “An abnormal behavior detection method using optical flow model and OpenPose,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 28–34, 2020.
- [32] L. Hong, D. Shuo, and L. Jian, “Traffic video significance foreground target extraction in complex scenes,” *Journal of Image and Graphics*, vol. 24, no. 1, pp. 50–63, 2019.
- [33] Z. Ouyang, J. Niu, and M. Guizani, “Improved vehicle steering pattern recognition by using selected sensor data,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 6, pp. 1383–1396, 2018.
- [34] D. M. W. Powers, “Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [35] B. Zhu, Y. Xie, G. Luo, and C. Lei, “An abnormal behavior detection method using optical flow model and OpenPose,” *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 11, no. 5, pp. 1–7, 2020.

Research Article

Adaptive Weighted Strategy Based Integrated Surrogate Models for Multiobjective Evolutionary Algorithm

Ke Bao ^{1,2}, Wei Fang,² and Yourong Ding ¹

¹Wuxi Institute of Technology, Wuxi, Jiangsu 214121, China

²Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Ke Bao; baoke@wxit.edu.cn

Received 8 May 2022; Revised 7 June 2022; Accepted 14 June 2022; Published 25 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Ke Bao 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.

Although the integrated model has good convergence ability, it is difficult to solve the multimodal problem and noisy problem due to the lack of uncertainty evaluation. Radial basis function model performs best for different degrees of nonlinear problems with small-scale and noisy training datasets but is insensitive to the increase of decision-space dimension, while Gaussian process regression model can provide prediction fitness and uncertainty evaluation. Therefore, an adaptive weighted strategy based integrated surrogate models is proposed to solve noisy multiobjective evolutionary problems. Based on the indicator-based multiobjective evolutionary framework, our proposed algorithm introduces the weighted combination of radial basis function and Gaussian process regression, and U-learning sampling scheme is adopted to improve the performance of population in convergence and diversity and judge the improvement of convergence and diversity. Finally, the effectiveness of the proposed algorithm is verified by 12 benchmark test problems, which are applied to the hybrid optimization problem on the construction of samples and the determination of parameters. The experimental results show that our proposed method is feasible and effective.

1. Introduction

With the continuous upgrading of application requirements, many problems in our real-world have been abstracted into high-dimensional multiobjective optimization problems [1]. For example, regarding the Internet of Vehicles, problems such as intrusion detection faced by the in-vehicle control network, software product selection in software engineering, selection and distribution of relief supplies, and green production optimization in coal industry can be modeled as high-dimensional objective optimization problems [2]. These practical problems may involve one or more conflicting optimization objectives, which are frequently constrained by harsh constraints. The selection of optimization schemes and the performance of optimization algorithms have become constant challenges as the complexity of the problem has increased [3].

Therefore, in recent years, how to use multiobjective evolutionary algorithm to solve high-dimensional objective optimization problems has attracted more and more

attention. The existing multiobjective evolutionary algorithms are generally divided into the following three categories: Pareto domination-based [4,5], indicator-based [6], and decomposition-based [7] multiobjective evolutionary algorithm. The diversity maintenance mechanism of most Pareto domination-based multiobjective evolutionary algorithms is only applicable to Pareto-front optimization problems with regular distribution [8]. When the distribution of Pareto-front is irregular or complex, those algorithms are often difficult to obtain solutions that consider proximity, distribution, and malleability. For index-based multiobjective evolutionary algorithms, they usually choose the optimal solution that makes more contributions to the index, which makes the dominated solution more likely to be favored than the nondominated solution; decomposition-based multiobjective evolutionary algorithms often have difficulty determining the orientation vectors that are matching with the shape of irregular Pareto-front, which leads to hardly solve the complex Pareto-front multiobjective optimization problems. Thus, to solve the

previously mentioned problems, a variety of effective high-dimensional multiobjective evolutionary algorithms (MaOEAs) have emerged [9]. Some improved MaOEAs choose more loose Pareto dominance criteria or simultaneously adopt Pareto dominance criteria and convergence criteria to alleviate the selection pressure [10]. Although this kind of algorithms can better ensure the convergence of the obtained Pareto optimal solution set, it is likely to ingrain a Pareto-front with poor distribution. On the other hand, mops problems are often mixed with noise in the process of fitness evaluation, which may mislead the search direction and reduce the optimization efficiency [11]. From this point of view, noise interferes with the steps of multiobjective optimization, such as determining nondominated individuals, diversity preservation, and elitism, so that the optimal solution of multiobjective optimization problems cannot be obtained, and the optimization problems cannot be well solved [11]. Therefore, how to deal with the noise in multiobjective optimization problem becomes very critical. However, noisy multiobjective optimization problems are general in practical engineering fields.

It can be seen from the previously mentioned analysis that most of the existing classical multiobjective evolutionary algorithms show bound limitations in dealing with high-dimensional objective optimization problems [12]. Therefore, many scholars combine multiobjective optimization with denoising model to solve the noise in the multiobjective optimization problems, resulting in noise multiobjective optimization algorithm. Because the existence of noise affects the inaccuracy of individual selection, adjusting the selection process is a fateful method to deal with the noise multiobjective optimization problem [13]. General selection methods include nondominated sorting selection, probabilistic sorting selection, threshold selection, and so on. Optimization problems in real-world are often affected by noise. And with different distributions, constraints, and central trends, noise may appear in random model parameters, objective functions, and decision variables. Many literatures have recently adopted evolutionary computation method to solve the optimization problems in practical applications from different perspectives [14]. However, for the noise optimization problems existing in engineering problems, even if the variables are set to constant, different target values will be obtained from multiple target evaluation processes due to the noise doping in the evaluation process of fitness function. In this case, the individuals entering the next generation are often not high-quality individuals [15]. To solve these problems, many researchers have proposed a variety of noise optimization algorithms.

Intuitively, a common method to reduce noise interference is to average multiple object values obtained from multiple function evaluation processes and then use the final average value to approximate the real object value. Eskandari and Geiger [16] used a fixed number of samples and averaged the target values of them and then provided a new sufficient condition for the convergence of the evolutionary algorithm that is with fixed number of samples. If there are many samples, there will be a colossal computational cost in

function evaluation. Basseur and Zitzler [17] proposed a strategy to reduce the evaluation times for everyone and not the average for all individuals; they averaged the time of some of the best individuals. So, a better solution can be found by averaging a small number of function evaluations. This simple method can significantly reduce the number of function evaluations. However, self-adaptive sampling of data can reduce the influence of noise, but sometimes it may have a great computational cost.

Therefore, some scholars have adopted one threshold method in the deterministic selection of evolutionary strategies. If and only if the fitness value of a descendant is better than that of its parent, the descendant will be accepted [18]. This kind of method calculates the probability that an individual dominates all individuals, the probability that an individual is dominated by all individuals, and the probability that it has no dominance relationship with all individuals, to calculate the rank value of this individual. This mechanism greatly improves the correctness of individual ranking and proves that the hierarchical ranking scheme used by NSGA in the case of no-noise may have defects in dealing with noisy problems [19]. Shim et al. proposed a regularity model-based noise multiobjective optimization algorithm according to regularity model-based multiobjective estimation of distribution algorithm [20], where the nondominated solution is used to establish the Pareto-solution model and samples the sample solution from it. With the continuous optimization of the evolution process, the established model continues to approach the real Pareto-solution, and the quality of the sample solution on the model is gradually improved, which proves that the model has a certain ability to resist noise. Hong et al. proposed the strategy that dividing the search space into several non-overlapping hyper-spheres and moving individual solutions in each sphere, which improved the average performance of the spheres [21]. This local model can filter noise and increase the robustness of the algorithm. By combining cooperative evolutionary frame and differential evolutionary algorithm, namely, cooperative differential coevolution algorithm, this model can effectively solve large-scale multiobjective optimization problems [22]. Multilevel cooperative coevolution algorithm adopts a technology called random grouping to group interactive variables into a subcomponent. In addition, another technique called self-adaptive weighting is used to adaptively adjust the subcomponents. In the cooperative coevolution with variable interaction learning [23], a new coevolution framework is proposed. Initially, all variables are regarded as independent variables, and each variable is placed in a separate group. And in the iterative process, the relationship between variables is gradually found and merged accordingly. DG2 is an improved differential grouping algorithm, which has better efficiency and grouping accuracy [24]. In the algorithm of using variable analysis method to reduce the dimension of search space, multiobjective evolutionary algorithm based on decision variable analysis controls variable analysis to identify the conflict between objective functions [25]. The algorithm carries out which variables affect the diversity of the generated solutions and which variables play an

important role in the overall convergence. Through the analysis of interdependent variables, the original multi-objective optimization problems are transformed into a series of sublevel multiobjective optimization problems, which can improve the convergence of most difficult multiobjective optimization problems. In order to reduce the dimension of search space through problem transformation, the weighted optimization framework is one of the most representative algorithms, where decision variables are divided into multiple groups, and each group is assigned a weight vector [21–24]. Thus, the optimization of original decision variables can be transformed into the optimization of weight vector. After the problem is transformed, the decision variable space of new problems will be greatly reduced. Large-scale multiobjective optimization framework also adopts the problem transforming strategy, which firstly decompose Pareto-solution into two directions associated with the weight vector, and then the weight vector is taken as input to construct a series of subproblems and is designed to track the corresponding points on the Pareto-solution so as to reduce the dimension of decision variable space [18,21,22,24].

Although the integrated model has good convergence ability, it is difficult to solve the multimodal problem and noisy problem due to the lack of uncertainty evaluation. Therefore, an adaptive weighted strategy based integrated surrogate model is proposed to solve noisy multiobjective evolutionary problems in this paper. Based on the indicator-based multiobjective evolutionary framework, our proposed algorithm introduces the weighted combination of radial basis function and Gaussian process regression, and U-learning sampling scheme is adopted to improve the performance of population in convergence and diversity and judge the improvement of convergence and diversity. Finally, the effectiveness of the proposed algorithm is verified by 12 benchmark test problems, which are applied to the hybrid optimization problem on the construction of samples and the determination of parameters. The experimental results show that our proposed method is feasible and effective.

2. Multiobjective Optimization and Its Noisy Problem

For a n multidimensional decision variable, the mathematical model of multiobjective optimization problem with m dimension can be defined as follows:

$$\begin{aligned} \min F(x) &= (f_1(x), f_2(x), \dots, f_m(x))^T \text{ s.t. } G_i(x) \geq 0, \\ i &\in 1, 2, \dots, p, H_j(x) = 0, j \in 1, 2, \dots, q, \end{aligned} \quad (1)$$

where $x = (x_1, x_2, \dots, x_n) \in \Omega$; x is the decision variable and Ω is the decision-space $\Omega = \prod_{i=1}^n [L_i, U_i]$; L_i and U_i are the upper and lower boundaries of the x_i ; $F(x)$ is the object vector, representing the mapping relationship of $\Omega \rightarrow R^m$; $G(x)$ and $H(x)$ are the constraints of the problem. Given $x^1 = (x_1^1, x_2^1, \dots, x_n^1)$ and $x^2 = (x_1^2, x_2^2, \dots, x_n^2)$, they are two decision vectors satisfying constraints in the object space, and x^1 Pareto dominates x^2 , denoted as $x^1 \succ x^2$, which satisfied $(\forall i) f_i(x^1) < f_i(x^2), i \in \{1, \dots, m\}$.

With the increasing demand of engineering problems, the results of solving multiobjective optimization problems are not only satisfied with a Pareto solution set or a Pareto optimal solution based on the preference of decision-makers. Sometimes, due to the influence of surrounding environmental factors or noise interference, people prefer to get a more robust Pareto optimal solution set or Pareto optimal solution. The so-called robustness is simply the sensitivity of the objective function to the small disturbance in the decision parameters [25]. If a global optimal solution is very sensitive to the variable disturbance, the final optimal solution obtained in reality may correspond to a different optimal value from the theoretical solution, which means a Pareto solution set with poor robustness is generated. To facilitate analysis and description, we will introduce the noisy model for the multiobjective optimization problem and proposed a simple and novel integrated surrogate-assisted model.

Since the evolutionary algorithm is less affected by noise in the early stage, Gaussian regression value can be used as the denoising object value. In practical engineering applications, the observed response usually contains noise, namely, $y(x) = f(x) + n$. Generally, the noise is assumed to be zero mean Gaussian distribution $\varepsilon \sim N(0, \Sigma_n)$, where the common form of Σ_n is $\Sigma_n = \sigma_n \mathbf{I}$, and \mathbf{I} is the identity matrix. For the responses of all observed samples, the variance of their noise is the same. That means the noise is an independent and identically distributed Gaussian distribution.

To reduce the influence of noise, a noise estimation and denoising method is proposed. The joint Gaussian distribution of the prediction $y(x)$ and noisy response Γ at x is written as follows:

$$\begin{Bmatrix} \Gamma \\ y(x) \end{Bmatrix} \sim N \left(\begin{Bmatrix} u \\ u(x) \end{Bmatrix}, \begin{Bmatrix} \sigma^2 R + \Sigma_n \sigma^2 r(x) \\ \sigma^2 r^T(x) + \mathbf{I} \end{Bmatrix} \right), \quad (2)$$

where the variance matrix is $C = \sigma^2 R + \Sigma_n$ and the covariance vector is $c = \sigma^2 r^T(x)$.

3. Surrogate-Assisted Evolutionary Algorithm

Surrogate-assisted model is to construct a relatively simple function from a complex function by collecting feature points and optimize the new function to obtain the optimal solution of complex function, which is shown in Figure 1. Since the surrogate-assisted model can only represent the real model to a certain extent, so its optimal solution can not directly represent the optimal solution of the original objective function. It is necessary to update the surrogate-assisted model according to a certain criterion. The framework of offspring generation is shown in Figure 2.

The search strategy of surrogate-assisted evolutionary algorithm (SAEAs) in the optimization process largely depends on a surrogate model [21,26]. The reason is that the surrogate model assumes that it can provide sufficiently accurate function estimation. Since the surrogate model cannot provide the same properties as the original objective function, it may even produce the optimal value that does not actually exist. Therefore, how to select an appropriate

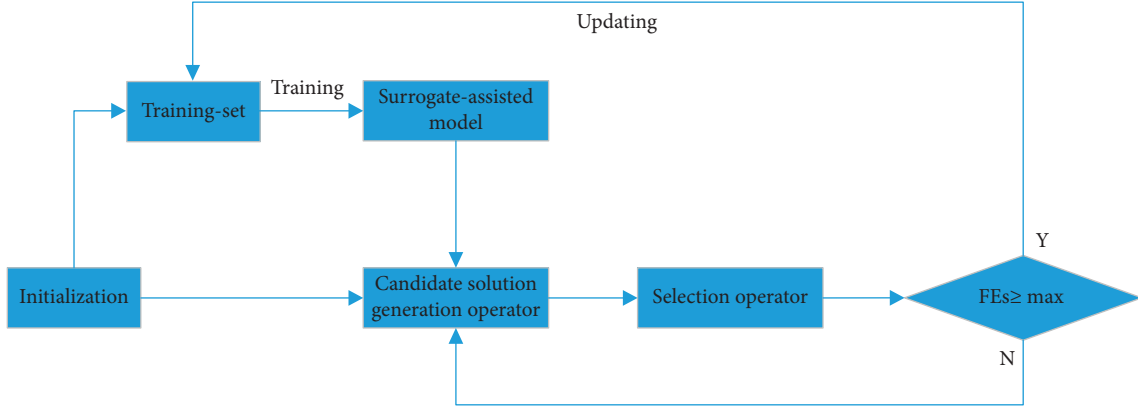


FIGURE 1: The framework of surrogate-assisted evolutionary algorithm.

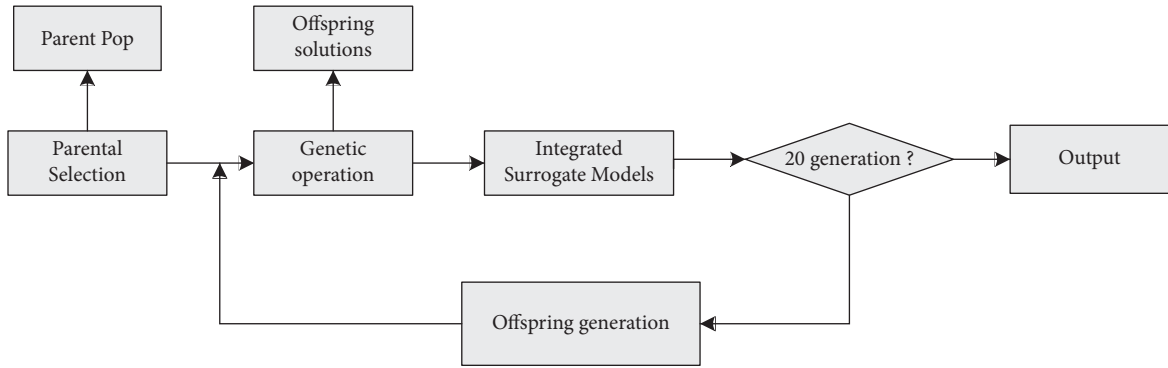


FIGURE 2: The framework of offspring generation.

surrogate model is a very important step in surrogate-assisted evolutionary algorithm. In practical application, the conventional operation is to use the surrogate model together with object-function evaluation. This section will briefly analyze two main surrogate-assisted models: Gaussian process regression(GPR) [27–30] and radial basis function (RBF) [25, 31].

3.1. Gaussian Process Regression. Gaussian process regression is a Bayesian statistical method for modeling functions, which is very suitable for objective function evaluation. Gaussian process does not need predefined data-structure and can approximate the nonlinear, discontinuous, and multimodal functions, which provides an uncertainty measurement in the form of standard deviation for the prediction function [32]. Gaussian process is the extension of multivariate Gaussian distribution to infinite dimensional random process, where any combination of finite dimensions is a Gaussian distribution. Since Gaussian distribution is a distribution of random variables, Gaussian process can be completely determined by its mean function and covariance function. Therefore, Gaussian process has also been widely used in surrogate-assisted evolutionary algorithms.

Given a dataset containing n samples $(x_i, y_i), i \in 1, 2, \dots, n$, the fitness y for any candidate solution x is regarded as $\mu + \varepsilon(x)$ in the GP model, where $\varepsilon(x)$ obeys the distribution $N(0, \sigma^2)$.

$$\begin{aligned} \mu &= k(x)K^{-1}y, \\ \sigma^2 &= \kappa(x) - k(x)^T K^{-1}k(x), \end{aligned} \quad (3)$$

K is the matrix, whose element is $K_{ij} = C(x_i, x_j)$. $k(x) = [C(x, x_1), C(x, x_2), \dots, C(x, x_n)]^T$. $X = (x_1, x_2, \dots, x_n)$ is the input value of the sample point, and $\kappa(x)$ is denoted as covariance, where $C(\cdot)$ is the covariance function. The covariance function can be written as follows:

$$C(x, x_i) = \sigma_f^2 \left(1 + \frac{\sqrt{3}r}{\sigma_l} \right) \exp\left(-\frac{\sqrt{3}r}{\sigma_l} \right), \quad (4)$$

where $r = \sqrt{(x - x_i)(x - x_i)^T}$. σ_f and σ_l are denoted as hyperparameter. Therefore, the predicted values and variances of candidate solutions x can be derived as follows:

$$\hat{y}(x) = u(x) + k(x)K^{-1}(y - Iu), \quad (5)$$

$$s^2(x) = \sigma^2 \left(1 + k^T(x)Kk(x) + \frac{(1 - I^T Kk(x))^2}{I^T K^{-1}I} \right), \quad (6)$$

where I is the identity column vector.

In (5) and (6), $\hat{y}(x)$ and $s^2(x)$ are the mean prediction function and variance prediction function of Gaussian process regression, respectively. It can be seen that the prediction value of the variance prediction function at the

point x is zero. The greater the distance from the existing sample point, the greater the prediction variance. Therefore, Gaussian process regression is an interpolation model about sample points. In addition, the distribution at prediction point x should meet the following requirements: $y(x) \sim N(u(x), s^2(x))$, so we can have

$$P(y(x) \leq t) = \Phi\left(\frac{t - u(x)}{s(x)}\right), \quad (7)$$

where $\Phi(x)$ is the Gaussian cumulative density function. $P(y(x) \leq t)$ indicates the probability that $y(x)$ is less than or equal to t at given x . Therefore, it can be given the confidence interval of $y(x)$ with probability $1 - \alpha$ at the point x .

$$y(x) \in \left[u(x) - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)s(x), u(x) + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)s(x) \right]. \quad (8)$$

Although Gaussian process has been widely used in solving expensive multiobjective optimization problems, it has encountered some bottlenecks in solving high-dimensional expensive problems. In addition, the existing infilling criterion [33] cannot be well applied to high-dimensional problems, and the time of constructing Gaussian process model will also increase relatively with the increase of training samples. Compared with low-dimensional problems, constructing Gaussian process model on high-dimensional problems requires more training samples. Therefore, the construction of GP model on high-dimensional problems becomes more time-consuming.

3.2. Radial Basis Function. Radial basis function (RBF) [25,31] is a scalar function whose value only depends on the distance from the origin. It is generally defined as a monotonic function about the radial distance between the sample and the data center. RBF kernel is one of the commonly used kernel functions. In essence, the radial basis function model is formed by linear superposition. Its formation process is denoted as follows: first, input a sample set, then select the corresponding radial basis function, and finally calculate the weighted sum of the radial basis function values between unknown points and all sample points, to obtain the predicted response value of the test point.

Since the independent variable of radial basis function is the Euclidean distance between the testing set and the sample set, this can well transform the multidimensional problem into a one-dimensional problem with only the independent variable of Euclidean distance. Given n different samples $x_1, x_2, \dots, x_n \in R^D$ and their corresponding function value $f(x_1), f(x_2), \dots, f(x_n)$ where n and D are arbitrary integers. Thus, we can have the following:

$$f(x) = \sum_{i=1}^n \lambda_i \varphi(|x - x_i|) + p(x) \quad x \in R^D, \quad (9)$$

where the coefficient $\lambda_i, i = 1, 2, \dots, n$ denotes the weight of the first i -th basis function; $|\cdot|$ is the Euclidean norm in R^D ; the degree of p is not greater than m from the polynomial space. It can be expressed as the linearity of the

function $x_1^{k_1}, x_1^{k_2}, \dots, x_1^{k_D}, x \in R^D$, where there are many choices of kernel function $\varphi(x)$ in radial basis function, such as linear, cubic, thin plate spline, and Gaussian. In our model, we adopted the Gaussian as kernel function, which can be written as follows:

$$\varphi(x) = e^{-\delta x^2}, \quad (10)$$

where $\delta > 0$ and is a constant.

The unknowns $(\lambda_1, \lambda_2, \dots, \lambda_n) \in R^D$ in the radial basis function are obtained by solving the following:

$$\begin{pmatrix} \varphi & p \\ p^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ C \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}. \quad (11)$$

In a system of linear equations, φ represents a matrix of size $n \times n$, where $\varphi_{ij} = \varphi(|x_i - x_j|)$. In (11), $\mathbf{p} = (x_1^T I, x_2^T I, \dots, x_n^T I)^T$, $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)^T$, $C = (b_1, b_2, \dots, b_n)^T$, and $F = (f(x_1), f(x_2), \dots, f(x_n))^T$.

It can be seen that the RBF model established through the previously mentioned analysis can replace the expensive function to predict the individual fitness value in the process of algorithm optimization and performs best for different degrees of nonlinear problems with small-scale and noisy training datasets. Therefore, radial basis function is used to approximate the original problem in this paper and assist evolutionary algorithm optimization, to reduce the consumption of computing resources.

4. Integrated Surrogate Models for Noisy MOPs

As we all know, radial basis function model performs best for different degrees of nonlinear problems on small-scale and noisy training datasets but is insensitive to the increase of decision-space dimension, while Gaussian process regression model can provide prediction fitness and uncertainty evaluation. Therefore, an adaptive weighted strategy based integrated surrogate models is proposed to solve noisy multiobjective evolutionary problems in this paper.

4.1. Proposed Adaptive Weighted Strategy. The choice of surrogate-assisted model is very important to the good performance of surrogate-assisted evolutionary algorithm. RBF model and GPR model are the two most popular surrogate-assisted models. RBF model performs best for different degrees of nonlinear problems on small-scale training datasets and is insensitive to the increase of decision-space dimension [34]. GPR model can provide prediction fitness and uncertainty evaluation. Therefore, the combination of these two kinds of information can prevent the search from falling into local optimization. These characteristics are very attractive and promising for solving high-cost or expensive optimization problems. Although the integrated model has good convergence ability, it is difficult to solve the multimodal problem and noisy problem due to the lack of uncertainty evaluation [35].

Integrated surrogate models (ISM) are a kind of ensemble of surrogate models (EM). It is an integrated model composed of a series of surrogate-assisted models by

weighting and combining. It can make use of the advantages of single surrogate-assisted model to effectively improve the robustness of prediction. The mathematical expression of surrogate models is described as follows:

$$\hat{y} = \sum_{i=1}^N w_i(x) y_i(x), \quad (12)$$

where $w_i(x)$ is denoted as weight and obeys $\sum_{i=1}^N w_i = 1$. \hat{y} and $y_i(x)$ represent the predicted values of EM and the integrated surrogate-assisted model, respectively.

The key step of constructing EM is to calculate the weight of each surrogate-assisted model. The prediction accuracy is positively proportional to the weight coefficient. According to the method of calculating weight, the existing weight strategies can be divided into two categories, including weighted average and point-by-point weight. The main principle of calculating the weight coefficient in the weighted average is to assign a weight to each subsurrogate-assisted model according to its global performance. Compared with the global error measurement, the weight of a single surrogate-assisted model for the point-by-point weight is calculated using the local error measurement, which means that the weight of each subsurrogate-assisted model will change in the whole sample space. Compared with the average weight method, the point-by-point weight method allows flexible adjustment of local weight coefficients in the sample space, which can better capture the local characteristics of the objective function, but it will increase the running time and is affected by noise.

Although the integrated model can obtain more accurate prediction values than a single model, and the effectiveness of the integrated model has been proved by theory, it is very difficult to fully meet the theoretical conditions of integrated surrogate model in practical applications, so it is difficult to ensure better generalization ability than a single model [36]. In our paper, an adaptive integration surrogate-assisted model is designed, which is shown in Table 1. First, calculate the number of samples in Step 1; calculate the Euclidean distance between the new sample point and all samples in the sample-set in Step 2; then, complete the division of training samples and test samples in Step 3, where 2/3 samples closest to the Euclidean distance of the new sample are selected as the training set to train each surrogate-assisted model. The remaining 1/3 samples are used as the test sample set to test the performance of the model. In Step 4, the training set divided in Step 3 is used to train RBF model and Gaussian process model, respectively. In Step 5, the performance of each surrogate-assisted model is tested on the test set, and the error of each surrogate-assisted model is calculated, respectively. In Step 6, the weight of each surrogate-assisted mode is updated. The weight of each surrogate-assisted model is equal to the ratio of the test error of the surrogate-assisted model in the test set to the sum of the test errors of all surrogate-assisted models. Update the corresponding weight according to the latest test error, where the sum of the weights of all surrogate-assisted models is 1. The method of updating the weight of the surrogate-assisted model by testing the error can adjust the weight at any time according

to the performance of the surrogate-assisted model, so that the weight of the surrogate-assisted model with smaller error is greater, so as to enhance the prediction accuracy and prediction stability of the integrated surrogate-assisted model. Finally, each surrogate-assisted model is used to predict and output the new predicted samples in Step 7, and then, the weighted sum of all the predicted output values is used as the output value of the final integrated model. It is worth noting that the training time of the integrated surrogate-assisted model can be reduced because not all samples are used in the training of the integrated model.

The next parent population need to be selected by some convergence selection strategy in the offspring generation, where we use crossover and mutation to produce the offspring solution. Table 2 shows the main steps of offspring selection.

4.2. Indicator-Based Multiobjective Evolutionary Algorithm.

The indicator-based multiobjective evolutionary algorithm adopts a fitness assignment scheme to rank the population members according to their usefulness regarding the optimization goal. The scheme is unique and simple and does not use the traditional diversity protection strategy, which makes the evolutionary algorithm have a good convergence and suitable for solving problems with a high dimension [36]. However, the algorithm performs poorly for some problems in maintaining the diversity preservation mechanism.

Performance indicator is a function that can use some preference information to assign a real number to any approximate solution set in many multiobjective evolutionary problems, so that the relative advantages and disadvantages of any two approximate solution can be judged according to the real number corresponding to each approximate solution set.

Since the binary performance index can assign a real function $I(x_i, x_j)$ to any pair of approximate solution sets (x_i, x_j) in the object space, it can be directly used for fitness calculation, but there is a prerequisite that the used index must obey Pareto rule. Fitness allocation is to grade individuals in the population according to their utilization value in the process of seeking optimization objectives [37]. Therefore, the formula for calculating individual fitness using performance indicators in this paper is as follows:

$$F(x_i) = \sum_{x_j \in S/\{x_i\}} -e^{-I(\{x_i\}, \{x_j\})/c \cdot k}, \quad (13)$$

where k is a scaling factor greater than 0.5. The experimental results show that the algorithm can achieve better results when $k = 0.05$; c is the maximum of the absolute values of all indicators, namely, $c = \max_{x_i \in S} |I(x_i, x_j)|$.

The main steps of the indicator-based multiobjective evolutionary algorithm are expressed as follows.

- (1) First, initialize the population Q , take it as the initial population, and set an empty population P ; set a variable g that holds the evolutionary generation.

TABLE 1: Main steps of integrated surrogate-assisted model.

Input: sample-set D , the predicted sample x
Output: the model output value y of the sample x to be predicted;
Step 1: calculate the number of samples in D , which is recorded as S_D ;
Step 2: for $i = 1$ to S_D do
 Calculate the Euclidean distance d_i between the sample x and the x_i ;
 End for
 Select the nearest $2/3$ samples closest to the Euclidean distance of x as the training set D_{train} and all the remaining samples as the test set D_{test} ;
Step 3: Gaussian process model M_1 and RBF neural network model M_2 are trained on the training set D_{train} , respectively
Step 4: for $i = 1$ to $2 S_D$ do
 Calculate the test error E_i of the model M_i on the test set D_{test} ;
 End for
Step 5: for $i = 1$ to 2 do
 Calculate the weight of surrogate-assisted model M_i , and $w_i = E_i / \sum_k E_k$;
 End for
Step 6: outputs the final predicted value $y = \sum_k w_k y_k$, where y_k is the predicted value of the model M_k at x ;

TABLE 2: Main steps of offspring selection.

Input: sample-set D , the predicted sample x
Output: offspring population p_0 ;
Step 1: all individuals in sample-set D are evaluated by weight model;
Step 2: $N/2$: $N/2$ individuals are selected from P_p and P_o ;
Step 3: $N/2$ individuals are selected from all the remaining individuals in P_p and P_o by the MSE of all m surrogate for uncertainty selection;
Step 4: combine two groups of individuals as new P_p for updating.

- (2) The individuals in P and Q are merged into R , and the individuals in R are processed by the indicator-based fitness assignment scheme.
- (3) Execute the environment selection operation, and continuously repeat the following two steps: (a) select the individual with the smallest fitness value from R and delete it; (b) update the fitness of the remaining individuals; repeat the previously mentioned two steps until the number of remaining individuals in R is equal to the size of P , and then put the remaining individuals in R into the P .
- (4) Judge whether the variable g is greater than the maximum evolutionary generation or meets other termination conditions. If yes, stop evolution and output the noninferior solution in P , otherwise continue to execute.
- (5) Use tournament selection to select individuals from P and copy them into Q .
- (6) Perform cross-mutation operation on the individual in Q to generate offspring individuals and replace the parent individual in Q with a new generation of individuals; add 1 to the evolutionary generation ($g = g + 1$) and turn to step (2).

4.3. U-Learning Sampling Approach. The larger the sample space, the higher the accuracy of the surrogate-assisted model. However, the construction of the surrogate-assisted model based on the larger sample space takes longer, so many scholars have successively studied the sampling size of

the initial sample space of the surrogate-assisted model [37]. It is believed that, in general, the sampling size of the initial sample space should be 10 times the dimension of the variable of the multiobjective optimization problem.

For simulation-based optimization, we usually update the surrogate-assisted model step by step by iteratively selecting promising sampling points until the stop condition is met. A good sampling strategy should consider both global and local search. In recent years, more and more methods with novel ideas have been proposed [25,35]. In addition, some researchers focus on selecting multiple sampling points in each iteration. These new acquisition points can be simulated in parallel if parallel computing resources are available, and the number of iterations will be greatly reduced.

To select the training sample points of the model efficiently, it is usually necessary to adopt an adaptive method to select the appropriate learning function and convergence conditions. In this paper, U-learning (Table 3) function $U(x) = |u(x)/\sigma(x)|$ is adopted as a sample point tool, which represents the probability that the positive and negative states of the sample output response $y = f(x)$ are misclassified. The smaller the value of U-learning function, the greater the probability of sample points being misclassified. Therefore, the sample points with smaller value are selected as new training points $x_{\text{new}} = \underset{x \in S}{\operatorname{argmin}} U(x)$, where S is the candidate sample pooling.

Literature [37] proves that the probability of sample points being correctly classified is 97.7% if $U(x)$ is equal to 2, so it can be considered that the constructed Integrated surrogate-assisted model has more than 97.7% probability of

TABLE 3: U-learning sampling strategy.

Input: sample-set S
Output: x_{new}
Step 1: set $S = p_c$
Step 2: while $ S > \eta$ do
Calculate $U(x)$ and find the smallest sample individual S from the sample set;
Delete S from the set;
Calculate the fitness of the remaining individuals.
End while

correct prediction of sample points when $U(x) > 2$. Therefore, it is determined that the convergence condition of sample point is $\min_{x \in S} U(x) \geq 2$.

5. Experiment and Simulation Analysis

5.1. Comparison Models. To verify the effectiveness of our proposed algorithm, we limit the number of real evaluations of the test problem to simulate the scenario of solving expensive MOPs. In addition, the existing popular surrogate-assisted evolutionary algorithms are also selected to evaluate its effectiveness. Next, we will briefly describe these selected comparison algorithms.

parEGO [38]: it is the first time to use EGO to solve multiobjective optimization problems parEGO. It divides the MOP problem into several subproblems evenly. In other words, the MOP problem is transformed into a single-objective optimization problem, randomly selects one subproblem at a time, and takes the expected improvement index (EI) as the sample selection strategy to select the solution for real evaluation. In addition, parEGO also limits the size of the training set to reduce modeling time.

SMS-EGO [39]: it uses the lower confidence bound to delimit the dominant relationship of the solution, and divide the obtained solution into non- ε dominant solution, ε dominant solution, and dominant solution. For non- ε dominant solution, SMS-EGO calculates its *S-metric* as the fitness. For ε dominant solution and dominant solution, SMS-EGO assigns a penalty value as fitness. The farther away from the nondominant solution, the greater the penalty. Based on this selection strategy, the solution for real evaluation is selected. Although SMS-EGO performs well, its running time is particularly long because it needs to calculate a lot of *S-metric*.

MOEA/D-EGO [40]: it clusters the solutions of the evolutionary algorithm and then uses EI as the selection strategy to select a solution in each cluster for real evaluation. In addition, it also uses the fuzzy clustering method to establish the surrogate-assisted model, which reduces the modeling time while using all the real evaluated solutions. Compared with parEGO, MOEA/D-EGO runs relatively faster because it selects solutions in batches each time, rather than just selecting a single solution.

K-RVEA [41]: the more contribution of this algorithm is in the model management strategy. It assigns the candidate solution obtained this time and the candidate solution obtained last time to a group of reference vectors, respectively. If the change in the number of inactive reference vectors is less than a certain threshold, uncertainty strategy is used as the basis for solution selection. Otherwise, the angle penalty distance is used as the basis for solution selection. In addition, when updating the model, K-RVEA will filter the solution to limit the size of the training set, to reduce the modeling time.

CSEA [42]: it uses the artificial neural network to predict the dominance relationship between candidate solutions and reference solutions. The uncertainty information in prediction is considered together with the dominance relationship to select promising solutions using the real objective functions.

5.2. Measurement Indicators. When the Pareto optimal solution set is obtained by the optimization algorithm, the advantages and disadvantages need to be analyzed by comparing the performance indexes. The performance index mainly evaluates the proximity between the nondominated solution set and the real optimal Pareto-front and the distribution and diversity of the solution set [43–45]. Therefore, some measurement indicators are proposed to evaluate the quality of the solutions obtained during optimization, to evaluate the quality of the algorithm. The performance evaluation indicators of the solution set of multiobjective optimization algorithm is mainly divided into convergence, uniformity, and spread. *Convergence* [46] reflects the difference between the solution set obtained by the optimization algorithm and the real Pareto-front. It is generally hoped that the solution set obtained is as close to the real Pareto front as possible. *Uniformity* reflects the degree of uniformity of the distribution of individuals in the solution set. Generally, it is hoped that the solution set obtained will be distributed as evenly as possible on Pareto front. *Spread* reflects the wide distribution degree of the whole solution set in the object space. Generally, it is hoped that the obtained solution set will be distributed on Pareto front as widely and completely as possible. Therefore, the performance evaluation indicators selected in this paper are shown as follows.

Generational distance (GD) is the average minimum distance from each point in the solution set P to the real

solution set P^* . The smaller the generational distance value, the better the convergence.

$$GD(P, P) = \frac{\sqrt{\sum_{y \in P} \min_{x \in P^*} \text{dis}(x, y)^2}}{|P|}, \quad (14)$$

where P is the solution set obtained by the evolutionary optimization algorithm and P^* is a set of uniformly distributed reference points sampled in the real Pareto front; $\text{dis}(x, y)$ represents the Euclidean distance between the points y in the solution set P and the point x in the sample reference set P^* .

Inverted generational distance (IGD) [47] represents the mean value of the nearest individual from the reference point. The smaller the inverted generational distance value, the better the convergence performance.

$$IGD(P, P^*) = \frac{\sum_{x \in P^*} \min_{y \in P} \text{dis}(x, y)}{|P^*|}. \quad (15)$$

Spacing is to measure the minimum distance standard deviation from each solution to other solutions. The smaller the spacing value, the more uniform the solution set.

$$\text{Spacing}(P) = \sqrt{\frac{1}{|P| - 1} \sum_{i=1}^{|P|} (\bar{d} - d_i)^2}, \quad (16)$$

where \bar{d} is denoted as the mean of all d_i .

Diversity metric (DM) [1,3,37] is designed to measure the spread of the obtained solution set.

$$\Delta = \frac{d_f + d_l + \sum_{i=1}^{N-1} |d_i - \bar{d}|}{d_f + d_l + (N-1)\bar{d}}, \quad (17)$$

where d_f and d_l represent the Euclidean distance between the extreme solution and the boundary solution of the obtained nondominated solution-set.

Hyper volume (HV) [1,3] is a performance metric for indicating the quality of a nondominated approximation set, where the super volume of the area formed by the nondominated front obtained after the optimization and the previously reference points. The larger the HV value, the better the comprehensive performance of the evolutionary optimization algorithm.

5.3. Parameter Settings. For the sake of fairness, all comparison algorithms use the original default parameters. Since the comparison algorithms selected in this paper are based on surrogate-assisted evolutionary framework [12]. Therefore, the common parameter setting is consistent, for example, the number of initial population data is set to 100. In ZDT, the number of decision variables is set to 12, while that of DTLZ is set to 10. The number of object variables of ZDT and DTLZ are set to 2 and 3, respectively. The maximum real evaluation times of ZDT and DTLZ are set to 200 and 300,

respectively. The setting probability of simulated binary crossover is 1.0 and its distribution index is 20. Polynomial mutation is selected as mutation operator, its probability is $1/d$, and the distribution index is 20. The number of reference vectors is 300 for two targets and 595 for three targets. Each algorithm runs 30 times independently for each test problem.

The key parameters of our proposed adaptive integrated surrogate model are set as follows. Population size is 100, and its fitness scaling factor is set to 0.05. In this section, the number of individuals for sampling η is set to 5. The maximum number of generations is set to 30. It is worth noting that the improved weighting strategy in this paper can adjust the weight to realize the prediction based on noise data, where the variance of noise is set to 0.2.

In this paper, all experiments on the test function were run independently for 10 times, and the average value and standard deviation (STD) [31] of the results were collected and compared. All algorithms are implemented on MATLAB 2019a and run on Intel (R) core (TM) i7-3770 CPU @ 3.40 GHz and 8 GB of RAM on personal computers.

5.4. Qualitative and Quantitative Analysis. In the comparative experiment, the selected test function is DTLZ and ZDT [12,18,22]. They are designed for multiobjective optimization problems. One of its most important features is the extensible adaptability dimension, such as variable dimension and objective function dimension. All the problems in this set of test problems are continuous n -dimensional multiobjective optimization problems with box constraints, which are scalable in the fitness. In addition, the Pareto front of each problem is different.

5.4.1. Comparison of Nondominated Solution for Different Algorithms. ZDT1 and ZDT2 are relatively simple test problems. When the number of decisions is small ($n=10$), they can achieve good convergence results. However, the performance of MOEA/D-EGO and K-RVEA decreased sharply with the increase of the number of variables. When $n=20$, parEGO performs slightly better than MOEA/D-EGO and K-RVEA, as shown in Figure 3, but IGD exceeding 10 means that its convergence is still very poor. Figure 4 shows the comparison of nondominated solution sets of different algorithms to obtain the optimal IGD value on ZDT1 ($n=50$). When $n=50$, parEGO, MOEA/D-EGO, and K-RVEA failed to find any solution on Pareto front. On the contrary, the performance of our proposed algorithm has been very stable from $n=10$ to $n=50$.

5.4.2. Comparison of Measurement Indicators for Different Algorithms. Tables 4 and 5 respectively, show the inverted generational distance (IGD) and hyper volume (HV) of different comparison algorithms, where the number in brackets represents the standard deviation of the index, and bold indicates that the value is the best on this test problem. In addition, we use the average results of 30 independent runs for performance analysis. The symbols “+,” “-,” and “≈”

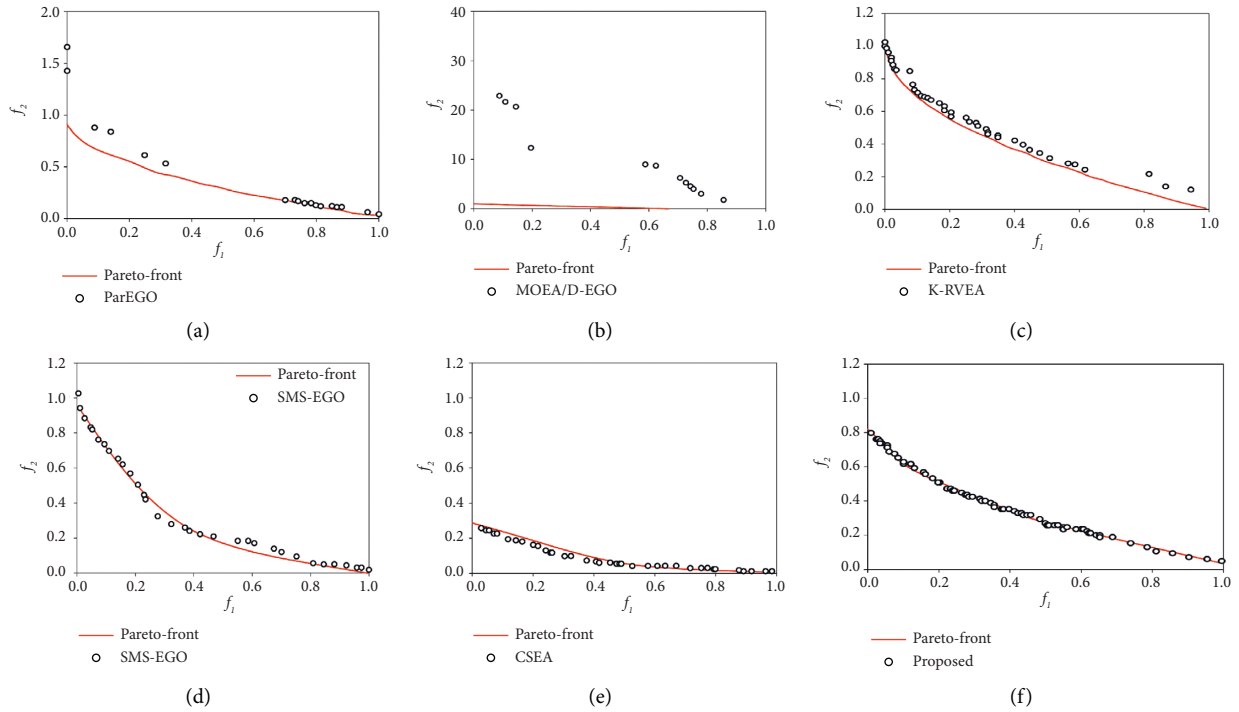


FIGURE 3: Comparison of nondominated solution for different algorithms on ZDT1. (a) ParEGO; (b) MOEA/D-EGO; (c) K-RVEA; (d) SMS-EGO; (e) CSEA; (f) our proposed.

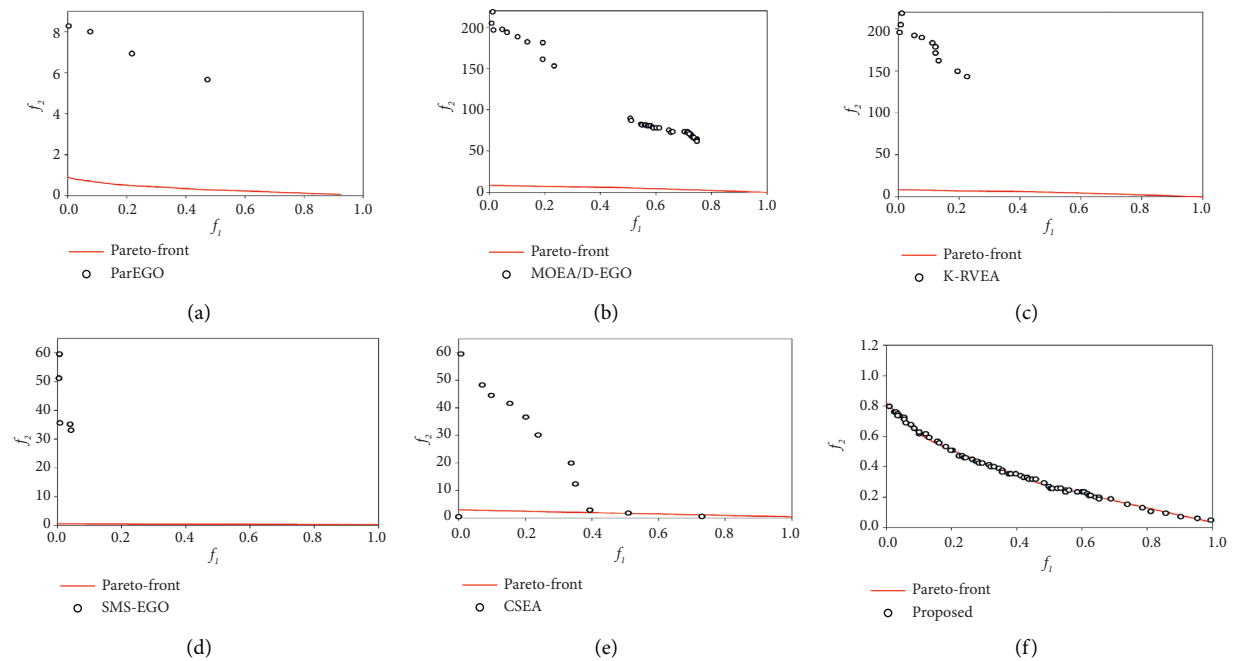


FIGURE 4: Comparison of nondominated solution for different algorithms on ZDT2 ($n = 50$). (a) ParEGO; (b) MOEA/D-EGO; (c) K-RVEA; (d) SMS-EGO; (e) CSEA; (f) our proposed.

indicate that our proposed weighted integrated surrogate model is statistically significantly superior to, inferior to, and almost equivalent to comparison models, respectively. The significance level of rank sum test is 0.05.

As shown in Table 4, only our proposed algorithm reached the near convergence state on DTLZ 1, while parEGO and MOEA/D-EGO performed poorly. On DTLZ2, K-RVEA performs best, and all algorithms can converge. On

TABLE 4: Comparison of inverted generational distance (IGD) for different algorithms.

Problems	ParEGO	MOEA/D-EGO	K-RVEA	SMS-EGO	CSEA	Proposed
ZDT1	4.7243e-1 (2.08e-1)-	1.7097e+0 (3.74e+0)-	8.2035e-1 (1.15e-1)-	3.7101e+0 (3.14e+0)-	8.2035e-1 (2.14e-1)-	6.2887e-2 (2.73e-2)
ZDT2	5.9465e-1 (1.43e-1)-	1.3118e+0 (1.55e+0)-	7.4677e-1 (1.37e-1)-	2.3611e+0 (1.54e+0)-	7.41257e-1 (1.20e-1)-	1.1052e-1 (1.07e-1)
ZDT3	536781e-1 (1.61e-1)-	1.4879e+0 (1.73e+0)-	8.1794e-1 (1.59e-1)-	3.4209e+0 (1.63e+0)-	8.17078e-1 (1.60e-1)-	1.9822e-1 (1.88e-1)
ZDT4	8.5326e+1 (1.37e+1)-	1.0688e+2 (1.28e+1)-	7.3940e+1 (2.27e+1)-	1.0711e+2 (1.30e+1)-	7.7943e+1 (1.25e+1)-	6.3493e+1 (1.78e+1)
ZDT5	6.4443e+0 (8.96e-1)-	6.2323e+0 (1.99e+0)-	5.4347e+0 (2.27e+0)-	5.2085e+0 (1.87e+0)-	5.8307e+0 (2.17e+0)-	3.2883e+0 (8.31e-1)
ZDT6	8.9775e+1 (2.24e+1)-	8.7589e+1 (1.64e+1)-	7.6495e+1 (1.87e+1)-	6.7327e+1 (1.65e+1)-	7.6085e+1 (1.87e+1)-	5.4726e+1 (1.71e+1)
DTLZ1	2.6802e-1 (2.14e-2)-	3.3203e-1 (2.55e-2)-	1.8420e-1 (1.88e-2)-	2.3251e-1 (2.57e-2)-	1.0320e-1 (1.18e-2)-	1.4977e-1 (3.47e-2)
DTLZ2	2.6622e+2 (6.05e+1)-	2.1185e+2 (4.05e+1)-	2.1205e+2 (7.18e+1)-	4.0287e+2 (4.11e+1)-	2.2505e+2 (6.18e+1)-	1.3896e+2 (2.36e+1)
DTLZ3	4.1291e-1 (1.12e-1)-	6.4259e-1 (7.17e-2)-	3.622e-1 (9.93e-2)≈	4.022e-1 (7.07e-2)-	3.6171e-1 (9.93e-2)≈	3.3670e-1 (1.01e-1)
DTLZ4	1.7326e-1 (3.06e-2)-	2.5429e-1 (3.03e-2)-	7.5002e-2 (1.28e-2)-	1.5019e-1 (2.83e-2)-	7.5662e-2 (1.28e-2)-	2.5984e-2 (6.28e-3)
DTLZ5	4.1258e+0 (4.68e-1)-	1.8576e+0 (5.81e-1)+	3.8069e+0 (4.72e-1)-	5.5814e+0 (4.99e-1)+	2.8071e+0 (4.72e-1)≈	2.7593e+0 (4.34e-1)
DTLZ6	3.3278e-1 (6.28e-2)-	2.3411e-1 (9.23e-2)-	1.0011e+0 (1.10e-1)-	3.0452e-1 (8.93e-2)-	1.8091e+0 (1.10e-2)≈	1.7815e-1 (3.03e-2)
+/-/≈	0/12/0	1/11/0	0/11/1	0/11/1	0/10/2	—

TABLE 5: Comparison of hyper-volume (HV) for different algorithms.

Problems	ParEGO	MOEA/D-EGO	K-RVEA	SMS-EGO	CSEA	Proposed
ZDT1	2.0311e-1 (1.29e-1)-	3.4343e-1 (2.33e-1)-	8.9466e-2 (7.60e-2)-	1.0688e-2 (1.28e+1)-	6.3940e-1 (2.27e+1)-	6.4567e-1 (1.48e-2)
ZDT2	4.0968e-2 (3.81e-2)-	9.4252e-2 (1.02e-1)-	4.9670e-3 (1.05e-2)-	3.2323e+0 (1.99e+0)≈	5.4347e+0 (2.27e+0)-	3.2907e-1 (8.19e-2)
ZDT3	1.9277e-1 (1.16e-1)-	2.4279e-1 (2.35e-1)-	1.0257e-1 (1.19e-1)-	8.7589e-1 (1.64e+1)-	5.1495e+1 (1.87e+1)-	5.2323e-1 (1.39e-1)
DTLZ2	1.8637e-1 (3.65e-2)-	1.3774e-1 (4.64e-2)-	3.5085e-1 (2.65e-2)-	4.3203e-1 (2.55e-2)≈	4.8420e-1 (1.88e-2)≈	4.3094e-1 (4.84e-2)
DTLZ4	2.1018e-1 (6.82e-2)≈	8.5741e-3 (1.49e-2)-	1.2883e-1 (1.17e-1)-	7.1185e+2 (4.05e+1)-	2.1205e-1 (1.18e-1)≈	2.0978e-1 (1.07e-1)
DTLZ5	5.9209e-2 (2.35e-2)-	2.5012e-2 (2.17e-2)-	1.4107e-1 (1.17e-2)-	6.4259e-1 (7.17e-2)-	1.622e-1 (6.93e-3)≈	1.7894e-1 (6.78e-3)
DTLZ7	1.5591e-1 (2.40e-2)-	2.1222e-1 (1.71e-2)-	1.5166e-1 (1.28e-2)-	3.5429e-1 (3.03e-2)-	3.5002e-1 (1.28e+2)-	2.2066e-1 (1.15e-2)
+/-/≈	0/6/1	0/7/0	0/7/0	0/5/2	0/4/3	—

DTLZ3, all algorithms fail to converge, and our proposed algorithm is close to convergence. The performance of DTLZ4 is similar to that of DTLZ2. Our proposed algorithm is close to the convergence state, while other comparison algorithms have converged; on DTLZ5, all comparison algorithms can converge, where MOEA/D-EGO performed better. In addition, all comparison algorithms failed to converge on DTLZ6.

In Table 5, on the DTLZ1 problem, only our proposed algorithm performs best, but it fails to converge and it is relatively close. On DTLZ2, parEGO performs best and completes convergence, and our proposed algorithm is still close to convergence. DTLZ3 fails to reach the convergence state. The performance of DTLZ4 is similar to that of

DTLZ2. On the DTLZ5 problem, the performance of all algorithms is similar, and only parEGO is slightly worse. All algorithms on DTLZ6 fail to converge.

Overall, our proposed algorithm in this paper performs better on the five-dimensional problem, especially on the more difficult DTLZ1 and DTLZ3 problems. Compared with the other two classical algorithms, our proposed algorithm still has an order of magnitude advantage in IGD performance. Our proposed algorithm is not satisfactory on the relatively simple DTLZ2 and DTLZ4 problems. In most cases, it is inferior to K-RVEA and CSEA. It is believed that our proposed algorithm has high convergence rate in population evolution, and has high IGD in most cases, which is consistent with the iterative curve.

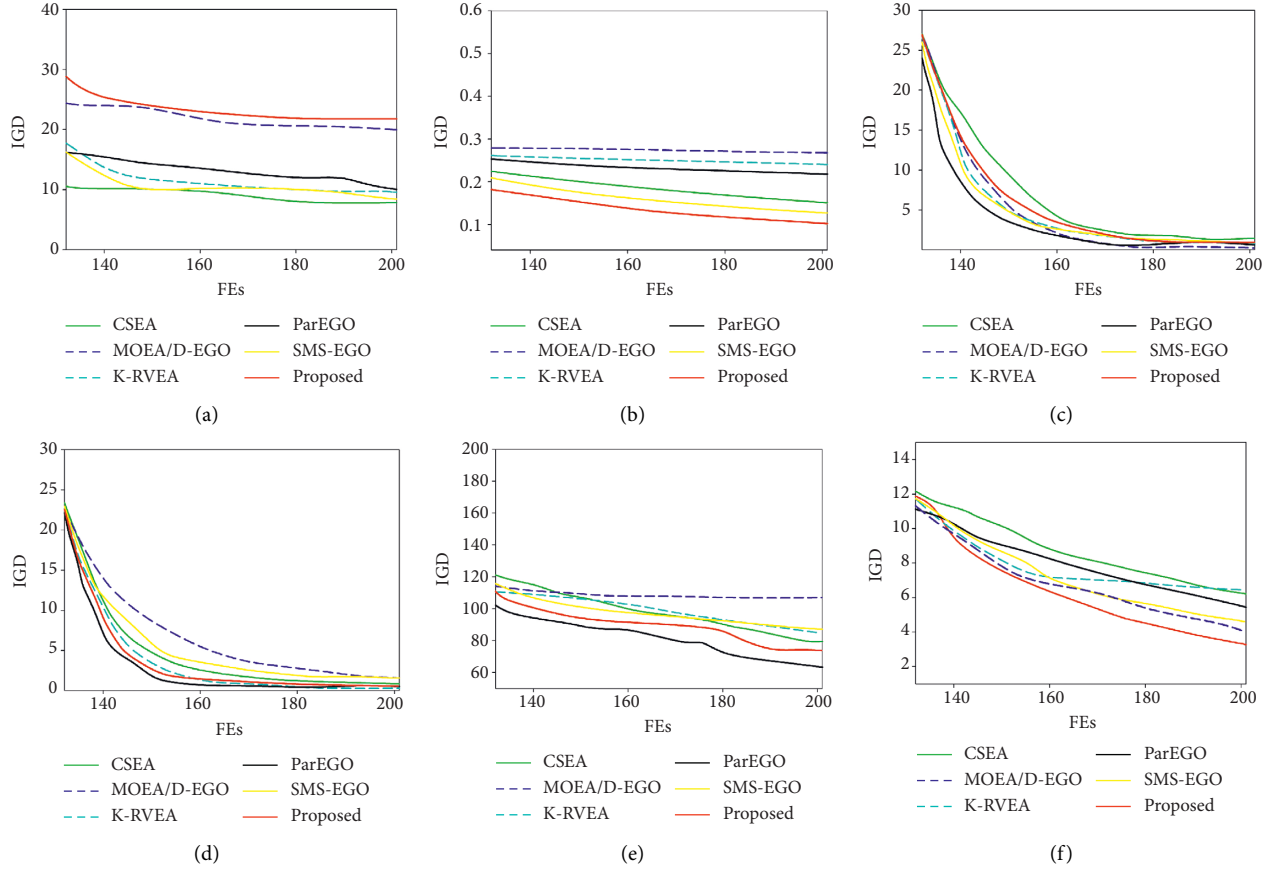


FIGURE 5: Comparison of convergence curve for different algorithms on ZDT problems. (a) ZDT1; (b) ZDT2; (c) ZDT3; (d) ZDT4; (e) ZDT5; (f) ZDT6.

Figures 5 and 6 show the convergence curves of our proposed algorithm and comparison algorithms on ZDT and DTLZ problems, where the abscissa is the real function evaluation times (Fes) and the ordinate is the IGD index. It is worth noting that, due to space constraints, we have only selected some problems for analysis. For ZDT test problem, the results of IGD and HV show that our proposed algorithm performs better when the experimental settings are consistent. The convergence curve shows that the convergence effect of our proposed algorithm is better than the comparison algorithm in most cases. Only on ZDT 1–3 problems, the convergence effect of our proposed algorithm is as small as parEGO before 170 real evaluations. Since the ZDT1–3 problems are relatively simple, our proposed algorithm is easy to find a better solution than the existing population in the initial stage. If the comparison algorithms select the solution according to the convergence, the convergence speed will be very fast. However, it is relatively easy to find a better solution than the existing population in the region close to the real PF, so the model needs to be able to better simulate the region near the real PF. Different from parEGO, our proposed algorithm selects solutions based on diversity, which is more inclined to increase the diversity of solution set, so that the surrogate-assisted model can better simulate the region near the optimal solution of the current population. Therefore, the convergence effect of the first

30th times of the convergence curve of our proposed algorithm is as small as that of parEGO, MOEA/D-EGO, and CSEA, and the later convergence effect is better than it.

For the DTLZ test problem, IGD, HV results and convergence curves show that our proposed algorithm performs better on most test problems when the experimental settings are consistent. For the DTLZ4 test problem, our proposed algorithm has the same effect as K-RVEA, but our proposed algorithm is not as good as MOEA/D-EGO for the DTLZ6 test problem.

5.4.3. Ablation Analysis for Noisy Treatment. Based on the indicator-based multiobjective evolutionary framework, our proposed algorithm introduces the weighted combination of radial basis function and Gaussian process regression. In other words, two different surrogate-assisted models are linearly combined to improve the optimization performance. To analyze the performance of this optimization strategy, we used ablation analysis to explain this difference. The comparative experimental curve is shown in Figure 7, where RBF-EA, GPR-EA, and Both-EA denote a radial basis function, Gaussian process regression, and the weighted combination, respectively.

Since DTLZ belong to the test problems that are sensitive to noises, we only choose DTLZ to analyze the performance

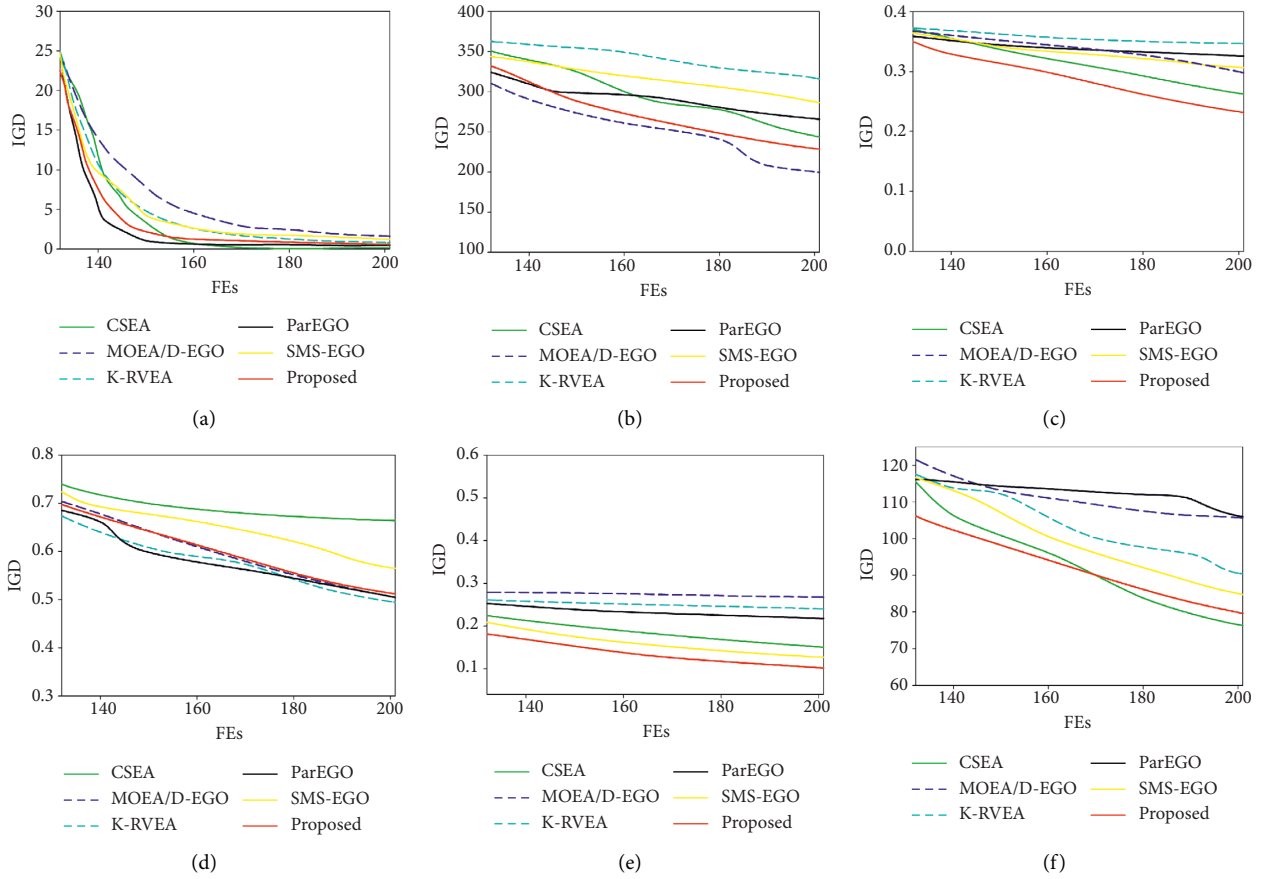


FIGURE 6: Comparison of convergence curve for different algorithms on DTLZ problems. (a) DTLZ 1; (b) DTLZ 2; (c) ZDT 3; (d) DTLZ 4; (e) DTLZ 6; (f) DTLZ 6.

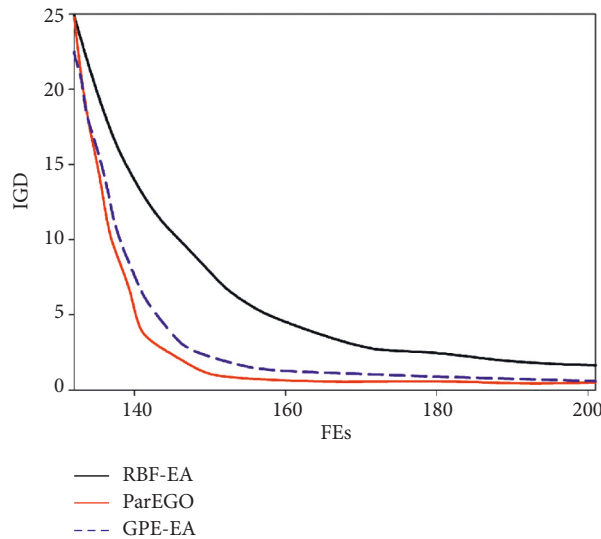


FIGURE 7: Comparative experimental curve for ablation analysis.

of our algorithm. The results are shown in Table 6. N-Both-EA is denoted as the version without noise treatment. It can be seen that the processing results of noisy data and clean data are quite different; that is to say, noise has a great impact on the performance of the evolutionary algorithm. However,

our proposed algorithm has higher precision than other comparison algorithms, which is attributed to our proposed algorithm with the ability of noise removal. Therefore, it is very necessary to deal with the noises for multiobjective evolutionary algorithm.

TABLE 6: Results of noisy treatment on the DTLZ problems.

Problems	RBF-EA	GPR-EA	N-Both-EA	Both-EA
DTLZ-2	$1.373e-1 (2.55e-2)\approx$	$4.8420e-1 (1.88e-2)\approx$	$4.3094e-1 (4.84e-2)-$	$1.3094e-1 (1.84e-2)$
DTLZ-3	$4.1185e+2 (4.05e-2)-$	$2.1274e-1 (1.18e-1)\approx$	$2.0914e-1 (1.07e-2)-$	$1.5978e-1 (9.07e-1)$
DTLZ-4	$6.7259e-1 (7.17e-2)-$	$1.672e-1 (6.93e-3)\approx$	$1.8694e-1 (6.17e-3)\approx$	$1.7490e-1 (5.78e-3)$
DTLZ-5	$3.5729e-1 (3.03e-2)-$	$3.7002e-1 (1.28e+2)-$	$2.2066e-1 (1.15e-2)\approx$	$2.210e-1 (1.75e-2)$
DTLZ-6	$3.5429e-1 (3.03e-2)-$	$3.5002e-1 (1.28e+2)-$	$2.2066e-1 (1.15e-2)-$	$1.2021e-1 (1.15e-2)$

The previously mentioned analysis shows that the designed model in this paper has higher performance than that using only a single strategy, which is helpful to improve the performance. In addition, through the comparison between our proposed algorithm and the other comparison algorithms, it is found that our proposed algorithm achieves better performance than the compared algorithms on noisy DTLZ and ZDT problems.

6. Conclusion

Radial basis function model performs best for different degrees of nonlinear problems on small-scale and noisy training datasets but is insensitive to the increase of decision-space dimension, while Gaussian process regression model can provide prediction fitness and uncertainty evaluation. Therefore, an adaptive weighted strategy based integrated surrogate models is proposed to solve noisy multiobjective evolutionary problems in this paper. Based on the indicator-based multi-objective evolutionary framework [7], our proposed algorithm introduces the weighted combination of radial basis function and Gaussian process regression, and U-learning sampling scheme is adopted to improve the performance of population in convergence and diversity and judge the improvement of convergence and diversity. Finally, the effectiveness of the proposed algorithm is verified by 12 benchmark test problems, which are applied to the hybrid optimization problem on the construction of samples and the determination of parameters. The experimental results show that our proposed method is feasible and effective [48–51].

Data Availability

The labeled datasets 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 by 2021 Jiangsu Higher Vocational College Teacher Professional Leader High-End Training Project (project no. 2021GRFX072).

References

- [1] Y. Jin, "Surrogate-assisted evolutionary computation: recent advances and future challenges," *Swarm and Evolutionary Computation*, vol. 6, no. 3, pp. 189–197, 2011.
- [2] D. Lim, Y. Jin, Y. S. Ong, and B. Sendhoff, "Generalizing surrogate-assisted evolutionary computation," *IEEE Transactions on Evolutionary Computation*, vol. 14, no. 3, pp. 329–355, 2021.
- [3] Y. S. Ong, P. B. Nair, and Y. L. Kai, "Max-min surrogate-assisted evolutionary algorithm for robust design," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 392–404, 2015.
- [4] Y. S. Ong, P. B. Nair, A. J. Keane, and K. W. Wong, "Surrogate-assisted evolutionary optimization frameworks for high-fidelity engineering design problems," *Springer Berlin Heidelberg*, vol. 23, no. 1, pp. 74–88, 2015.
- [5] Y. S. Ong, K. Y. Lum, P. B. Nair, D. M. Shi, and Z. K. Zhang, "Global convergence of unconstrained and bound constrained surrogate-assisted evolutionary search in aerodynamic shape design," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 400–404, Canberra, ACT, Australia, December 2003.
- [6] T. Chugh, K. Sindhya, K. Miettinen, J. Hakanen, and Y. Jin, "On constraint handling in surrogate-assisted evolutionary many-objective optimization," in *Proceedings of the Parallel Problem Solving from Nature (PPSN) 2016*, pp. 1255–1263, Edinburgh Scotland, September 2016.
- [7] N. Zheng, H. Wang, and B. Yuan, "An adaptive model switch-based surrogate-assisted evolutionary algorithm for noisy expensive multi-objective optimization," *Complex & Intelligent Systems*, 2022.
- [8] Z. Liu, H. Wang, and Y. Jin, "Performance indicator-based adaptive model selection for offline data-driven multi-objective evolutionary optimization," *IEEE Transactions on Cybernetics*, vol. 25, pp. 659–664+12, 2022.
- [9] A. Massaro and E. Benini, "A surrogate-assisted evolutionary algorithm based on the genetic diversity objective," *Applied Soft Computing*, vol. 36, pp. 87–100, 2015.
- [10] A. Xw, A. Yj, B. Ss, B. Mo, and C. Ra, "Transfer learning based surrogate assisted evolutionary bi-objective optimization for objectives with different evaluation times," *Knowledge-Based Systems*, vol. 23, no. 1, pp. 74–88, 2021.
- [11] Q. Liu, X. Wu, Q. Lin, J. Ji, and K. C. Wong, "A novel surrogate-assisted evolutionary algorithm with an uncertainty grouping based infill criterion," *Swarm and Evolutionary Computation*, vol. 60, no. 2, pp. 1067–1087, 2021.
- [12] M. Zhang, H. Li, S. Pan, J. Lyu, and S. Su, "Convolutional neural networks based lung nodule classification: a surrogate-assisted evolutionary algorithm for hyperparameter optimization," *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 99, p. 1, 2021.
- [13] X. Wang, Y. Jin, S. Schmitt, and M. Olhofer, "Transfer learning based co-surrogate assisted evolutionary bi-objective optimization for objectives with non-uniform evaluation times," *IEEE Trans Syst Man Cybern Part C (Applications and Reviews)*, vol. 32, no. 4, pp. 460–473, 2021.
- [14] E. J. Hughes, "Evolutionary multi-objective ranking with uncertainty and noise," in *Proceedings of the International*

- conference on evolution-ary multi-criterion optimization, pp. 329–343, Springer, Zurich, Switzerland, March 2001.
- [15] P. Boonma and J. Suzuki, “A confidence-based dominance operator in evolutionary algorithms for noisy multi-objective optimization problems,” in *Proceedings of the 2009 21st IEEE international conference on tools with artificial intelligence*, pp. 387–394, IEEE, Newark, NJ, USA, November 2009.
- [16] H. Eskandari and C. D. Geiger, “Evolutionary multiobjective optimization in noisy problem environments,” *Journal of Heuristics*, vol. 15, no. 6, pp. 559–595, 2009.
- [17] M. Basseur and E. Zitzler, “A preliminary study on handling uncertainty in indicator-based multiobjective optimization,” in *Proceedings of the Workshops on applications of evolutionary computation*, pp. 727–739, Springer, Budapest, Hungary, April 2006.
- [18] M. Babbar, A. Lakshmikantha, and D. E. Goldberg, “A modified NSGA-II to solve noisy multi-objective problems,” in *Proceedings of the 2003 genetic and evolutionary computation conference. Late-Breaking Papers*, pp. 21–27, Chicago, IL, USA, July 2003.
- [19] H. Tang, S. V. Ann, T. K. Chen, and C. J. Yong, “Restricted Boltzmann machine based algorithm for multi-objective optimization,” in *Proceedings of the IEEE congress on evolutionary computation*, pp. 1–8, IEEE, Barcelona, Spain, July 2010.
- [20] V. A. Shim, K. C. Tan, J. Y. Chia, and A. Al Mamun, “Multi-objective optimization with estimation of distribution algorithm in a noisy environment,” *Evolutionary Computation*, vol. 21, no. 1, pp. 149–177, 2013.
- [21] Y. Hong, Q. Ren, and J. Zeng, “Optimization of noisy fitness functions with univariate marginal distribution algorithm,” vol. 2, pp. 1410–1417, in *Proceedings of the 2005 IEEE congress on evolutionary computation*, vol. 2, pp. 1410–1417, IEEE, Edinburgh, UK, September 2005.
- [22] V. G. Asouti, S. A. Kyriacou, and K. C. Giannakoglou, *Pca-enhanced Metamodel-Assisted Evolutionary Algorithms for Aerodynamic Optimization*, Springer International Publishing, Manhattan, NY, USA, 2016.
- [23] T. Chugh, R. Allmendinger, V. Ojalehto, and K. Miettinen, “Surrogate-assisted evolutionary biobjective optimization for objectives with non-uniform latencies,” in *Proceedings of the The Genetic and Evolutionary Computation Conference (GECCO) 2018*, pp. 609–616, Kyoto, Japan, July 2018.
- [24] A. Lombardi, D. Ferrari, and L. Santos, “Aircraft air inlet design optimization via surrogate-assisted evolutionary computation,” in *Proceedings of the International Conference on Evolutionary Multi-Criterion Optimization*, pp. 420–434, Shenzhen, China, March 2021.
- [25] L. Vincenzi and P. Gambarelli, “A proper infill sampling strategy for improving the speed performance of a surrogate-assisted evolutionary algorithm,” *Computers & Structures*, vol. 18, no. 24, pp. 58–70, 2021.
- [26] Y. Liu, J. Liu, S. Tan, Y. Yang, and F. Li, “A bagging-based surrogate-assisted evolutionary algorithm for expensive multi-objective optimization,” *Neural Computing & Applications*, vol. 14, no. 3, pp. 319–377, 2022.
- [27] R. Liu, Y. Li, H. Wang, and J. Liu, “A noisy multi-objective optimization algorithm based on mean and wiener filters,” *Knowledge-Based Systems*, vol. 228, Article ID 107215, 2021.
- [28] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, “On the surprising behavior of distance metrics in high dimensional space,” in *Proceedings of the International conference on database theory*, pp. 420–434, Springer, London, UK, January 2001.
- [29] P. Le Yu, “Cone convexity, cone extreme points, and non-dominated solutions in decision problems with multi-objectives,” *Journal of Optimization Theory and Applications*, vol. 14, no. 3, pp. 319–377, 1994.
- [30] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.
- [31] P. Huang, H. Wang, and W. Ma, “Stochastic ranking for offline data-driven evolutionary optimization using radial basis function networks with multiple kernels,” in *Proceedings of the 2019 IEEE symposium series on computational intelligence (SSCI)*, pp. 2050–2057, IEEE, Xiamen, China, December 2019.
- [32] W. Shen, X. Guo, C. Wu, and D. Wu, “Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm,” *Knowledge-Based Systems*, vol. 24, no. 3, pp. 378–385, 2011.
- [33] G. Fu, C. Sun, Y. Tan, G. Zhang, and Y. Jin, “A surrogate-assisted evolutionary algorithm with random feature selection for large-scale expensive problems,” in *Proceedings of the International Conference on Parallel Problem Solving from Nature*, pp. 259–267, Springer, Leiden, The Netherlands, December 2020.
- [34] X. Cai, L. Gao, and X. Li, “Efficient generalized surrogate-assisted evolutionary algorithm for high-dimensional expensive problems,” *IEEE Transactions on Evolutionary Computation*, no. 99, p. 1, 2019.
- [35] G. Chen, K. Zhang, X. Xue, L. Zhang, and Y. Yang, “Surrogate-assisted evolutionary algorithm with dimensionality reduction method for water flooding production optimization,” *Journal of Petroleum Science and Engineering*, vol. 42, no. 1, pp. 55–61, 2019.
- [36] 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, 2014.
- [37] H. Wang, Y. Jin, and X. Yao, “Diversity assessment in many-objective optimization,” *IEEE Transactions on Cybernetics*, vol. 47, no. 6, pp. 1510–1522, 2016.
- [38] T. Wagner, N. Beume, and B. Naujoks, “Pareto-, aggregation-, and indicator-based methods in many-objective optimization,” in *Proceedings of the Int. Conf. Evol. Multi-Criterion Optim*, pp. 742–756, Matsushima, Japan, March 2007.
- [39] W. K. Mashwani, A. Salhi, M. Jan, R. Khanum, and M. Suliaman, “Evolutionary algorithms based on decomposition and indicator functions: state-of-the-art survey,” *Advances in Computer Science and its Applications*, vol. 7, no. 2, pp. 583–593, 2016.
- [40] W. K. Mashwani and A. Salhi, “Multiobjective memetic algorithm based on decomposition,” *Applied Soft Computing*, vol. 21, pp. 221–243, 2014.
- [41] W. K. Mashwani and A. Salhi, “Multiobjective evolutionary algorithm based on multimethod with dynamic resources allocation,” *Applied Soft Computing*, vol. 39, pp. 292–309, 2016.
- [42] W. K. Mashwani, S. N. A. Shah, S. B. Belhaouari, and A. Hamdi, “Ameliorated ensemble strategy-based evolutionary algorithm with dynamic resources allocations,” *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 412–437, 2020.
- [43] A. Acan and J. Tamouk, “A dynamic metaheuristic network for numerical multi-objective optimization,” *International Journal on Artificial Intelligence Tools*, vol. 7, 2021.

- [44] M. V. Patil and A. J. Kulkarni, "Pareto dominance based multi-objective cohort intelligence algorithm," *Information Sciences*, vol. 38, no. 1, pp. 125–138, 2020.
- [45] I. P. Kougias and N. P. Theodossiou, "Multiobjective pump scheduling optimization using harmony search algorithm (hsa) and polyphonic HSA," *Water Resources Management*, vol. 27, no. 5, pp. 1249–1261, 2013.
- [46] M. Zhang, L. Wang, Z. Cui, J. Liu, D. Du, and W. Guo, "Fast nondominated sorting genetic algorithm II with lévy distribution for network topology optimization," *Mathematical Problems in Engineering*, vol. 2020, no. 2, pp. 1–12, 2020.
- [47] H. Wang, Y. Jin, and J. Doherty, "Committee-based active learning for surrogate-assisted particle swarm optimization of expensive problems," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2664–2677, 2017.

Research Article

A Novel AI-Based Visual Stimuli Generation Approach for Environment Concept Design

Yingjing Duan and Jie Zhang 

School of Design, Jiangnan University, Wuxi, China

Correspondence should be addressed to Jie Zhang; zhangjie@jiangnan.edu.cn

Received 12 May 2022; Revised 13 June 2022; Accepted 14 June 2022; Published 24 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yingjing Duan and Jie 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 nature of environment concept design is a visual-based issue, where designers need to find loads of visual stimuli to create the high-quality concept. This paper aims to introduce a novel AI-based method to automatically generate images as design inspiration for environment concept design. Through interviewing eight professionals, we discovered that acquiring the design stimulus with inspired “composition” in a short time is the prioritized need of designers. This paper takes spectacular ambience for example. Through testing six classic GAN model variants trained by a self-made data set, we selected qualified four models to generate black and white thumbnails as stimuli with spectacular ambience. Moreover, we conducted a qualitative study of the outputs of the four models in a manner that invite eight designers to discuss. Finally, we summarized five key factors that influence designers’ satisfaction with generated visual stimuli and discuss future directions that are worth studying.

1. Introduction

Conceptual design is an early stage of the design process, which involves information gathering, idea generation, and concept evaluation via a collection of sketches, images, and written statements to define design direction, as well as plays an instructive role in the following process of the produced objects design [1]. With the development of digital technology, the entertainment industry system, such as games and film, tends to mature, and the concept design that is originally widely used in product and architectural design begins to appear in the early stages of film and TV shows production, and its role is to visually present various settings in the planning and provide guidance during mid- and post-design phases, such as modeling, scene construction, or prop design. Environment concept design is a key part of digital industrial production. It requires designers to visually present scenes that conform to the worldview or plot development through the study of style, color, ambience, composition, and other knowledge about design and aesthetics, whose essence is visually based design. As shown by Gabriela Goldschmidt research [2], when designers are

required to solve design problems at a conceptual level, especially visual-based design issues, the presence of visual stimuli has an effect on the qualities of the solutions they arrive at. Therefore, in the early stages of concept scene design, designers need to look for a large number of visual stimuli to support concept expansion.

With the advancement of artificial intelligence, its techniques have been applied in many domains and gradually become significant assistance to help the development of other fields. In design, artificial intelligence boosts the efficiency of designers substantially via its powerful capabilities of data mining and analysis [3]. In design ideation, a variety of AI-based methods are proposed, which not only can create a larger number of cross-domain idea associations but also can advance the ideation process quickly and easily in terms of quantity and novelty [4]. Although the creativity supporting method for design ideation has yielded a wealth of research results, these studies are mostly in the areas of product, apparel, or graphic design, but not in the area of environment concept design. Therefore, the purpose of this paper is to present an AI-based method, which can generate loads of visual stimuli randomly for expansion of design

solution space to increase the quantity and diversity of generated design concepts in the ideation of environment concept design.

The rest of the paper is structured as follows. The next section reviews the literature regarding the influence of visual stimuli on design ideation and the proposed AI-based methods for design ideation. In the following section, we interviewed eight design professionals to identify the challenges in ideation as well as the potential solutions. Then, we introduce our method in detail and discuss the results. Recommendations for further studies and concluding remarks are finally made.

2. Related Work

2.1. The Influence of Visual Stimuli on Design Ideation.

Ideation is an essential step in the integral design process to generate, develop, and communicate ideas, where the idea is a basic element of thought that can be either visual, concrete, or abstract [5]. In visual-based design, image materials as design stimuli have a positive effect on the quality, novelty, and diversity of design outputs. Many studies have been carried out to understand the effects of image inspiration in design ideation.

In terms of the richness of visual stimuli, an empirical study has been conducted by Goldschmidt and Smolkov [6] who have revealed that the presence of visual stimuli of different kinds can affect performance. In this study, the designers were exposed to different situations (e.g., no visual stimuli, sketches as stimuli, and rich and diverse stimuli), whose performances were measured in terms of practicality, originality, and creativity scores in light of outputs. Besides, this study also has indicated that designers tend to be particularly sensitive to various types of external stimuli, especially surrounding visual displays. Cardoso and Schaub [7] have carried out a research study to expand on the type of visual stimuli (e.g., line-drawing and photographic). The result shows only one type of representation of external material could induce high levels of attribute repetition, and diverse and rich visual stimuli could reduce design fixation.

Regarding ambiguity, Tseng [8] has conducted an experiment to explore the influence of visual ambiguity on design ideation, where the ambiguous images used as stimulus were classified into high, moderate, and low ambiguity and subjects were asked to use the ideas suggested by the visual cues to design a novel table. The results have indicated with higher ambiguity of visual display can provide the designer with greater freedom to search for more diverse ways to resolve presented uncertainties, thereby increasing novelty during idea generation. Through an experiment to investigate the effect of ambiguous visual stimuli on creativity in design ideation, Jang et al. [9] have presented that ambiguous stimuli from dissimilar concept pairs fostered higher elaboration in individual working.

In addition, Laing and Masoodian [10] have carried out a study with 18 graphic design students, who were provided with images related to the aesthetic tastes of clients and their market competitors. The findings demonstrate that the differences in creativity are minimal measurably between the

two types of images, but the interview data from participants have shown that pictorial inspiration has a positive effect on the way to approach the tasks and increases their satisfaction with the textual task descriptions provided.

In summary, the studies motioned above have three main effects on design ideation:

- (1) Rich and diverse visual stimuli have a positive effect on the phase of conceptual expansion in design
- (2) Availability of visual display can provide designers with benefits to their experience during the ideation phase to improve efficiency
- (3) Higher ambiguity of figures used as design cues could expand the freedom of designers for discovering more innovative ways to increase the novelty of generated ideas and reduce fixation

Our study is based on the conclusions the existing literature has suggested and then to obtain more precise needs about concept generation in environment concept design by interviewing design professionals. We set the goals to be achieved by our method eventually.

2.2. AI-Based Approaches for Design Ideation.

As a brunch of computational creativity, a variety of AI-based methods for promoting creative idea generation for design have been proposed with the advancement of computational creativity, which can be categorized into stimuli retrieval and stimuli generation accordingly to the ways they are produced [11].

In terms of retrieval, Hao et al. [12] have introduced an evolutionary computation method, which automatically generates language terms as design stimuli through retrieving from a vocabulary base based on 50,000 granted patents. Shi et al. [13] have proposed an unsupervised learning ontology network for design information retrieval, focusing on the associations between design and engineering by data-driven text mining and semantic network analysis. Sarica et al. [14], through employing the latest natural language processing techniques to extract semantic-level knowledge in all technology fields from six million US patent documents, have introduced a methodology named TechNet to retrieve the engineering knowledge in a field and explore engineering concepts around the field for future design considerations and innovation.

With the development of AI image generation technique, the computer can generate imagery material directly for creativity support. Chen et al. [4] have proposed a visual stimulus generation model, whose generated images visually captured partial elements of two distinct concepts in two different domains. Li et al. [15] have introduced a product concept generation method, which was formed by an affective recognition model to mark the affective preferences of users and product design GAN model (PD-GAN) to generate conceptual images with affective preferences via deep learning techniques and Kansei engineering. Compared to retrieval, there are relatively few studies about a generation now. Empirically, as the generation model can have a variety of new images, it has the infinite potentiality to be employed during the ideation phase in design.

The above literature review shows the crucial technique employed for approaches to AI-based design ideation is semantic network. Research studies start off by constructing a design knowledge base, which is formed by an enormous number of existing excellent design cases or patents, then retrieve the knowledge base through AI techniques, and generate semantic-level stimuli by analogizing, filtering, combining, and classifying concepts in the knowledge base. As to visual representation, the process of stimulus generation requires semantic understanding first before visual display generation. This type of method is suitable for well-defined design problems with concrete requirements, but not for ill-defined ones with abstract requirements, such as issues related to environment conceptual design. Therefore, this paper intends to develop an AI-based approach to solve this problem.

3. Requirements Research

In this section, we will interview eight design professionals in the domain. By analysis of the results of the interview, we then set the goals of our approach.

3.1. Interviews with Design Professionals. Design professionals (five males and three females), who are invited to participate in the interview, are all majored in environment concept design; four of them are graduate students; and the other four work in the company, who all have over five years of experience in this domain. The interviews were conducted online via screen-sharable video software. Each interview took about 60 minutes. Two researchers as interviewers carried out and recorded the interviews. Besides, for later analysis, we video-recorded the interviews.

In the interview, we mainly asked about: (1) their ideation process in environment concept design practice, (2) barriers and challenges that trouble them during the phase of concept generation, and (3) without warring about limitations, their ideal solutions to address these challenges. During the interviews, the researchers pursued further questions about the ambiguous answers given by the interviewees so that they could get clearer and more enlightening answers.

3.2. Results. Afterwards, we arranged the feedback from the professionals and summarized them in three aspects: ideation process, challenges in ideation, and ideal solutions as follows.

3.2.1. Ideation Process. We identified three steps of the environment concept design ideation process:

(1) *Ambience Defining.* Designers analyze the design requirement from the scenario to define the ambience in the conceptual environment. Ambience refers to the feeling and emotion that the scene wants to convey, which is usually defined by the semantics of composition color and elements on the visual guiding line in a picture. The definition of

ambience requires the designer to analyze the design requirements given by the plan, extract, or derive the description related to the ambience, which is the key part in the early stage of ideation and provides guidance for the next steps.

(2) *Composition Setting.* The expression of the ambience is entwined with the composition in environment design. Composition is combinations and arrangements of different elements in a conceptual environment graphic. Generally, during this phase, designers first select the composition category; then determine the position of subjects, such as landmark buildings or other objects in the visual center of the picture; and finally arrange or combine the other elements by designing a visual guideline, which is used to direct the viewer's gaze to the center of vision. Composition is the beginning of the concept visualization. In this stage, experienced designers usually analyze the planning scheme, combine the experience and aesthetic principles to create mental images, and then look for visual materials that can be employed as stimuli to find the optimal solution according to the mental images, while novices usually need to collect a large number of design stimuli before they produce creative imagery.

(3) *Hue Setting.* Color is another essential element that embodies the ambience. Color refers to the large blocks of color in a picture. The hue is composed of two main parts: color matching and color proportion. Color matching usually depends on emotions, the climate, time period, and geographical location in environment concept design, and color promotion is mainly affected by the composition. Therefore, the hue setting depends not only on the ambience but also on the composition. In this part, as with the composition setting, designers also need to look for a large number of visual stimuli to support the idea generation of hue.

Overall, the environment concept design is mainly composed of three stages, and its process is shown in Figure 1. We find that the ideation of composition is a key part of the phase of ideation in environment concept design, which plays a role in carrying forward the upper and lower levels. The setting of composition is not only a preliminary presentation of the ambience in a conceptual environment but also influence the setting of the hue. Therefore, our supporting creation method focuses on the assistance of the composition. The reporting and analysis of the interview results of the following two aspects will also focus on the composition.

3.2.2. Challenges in Ideation. We also identified design professionals' thoughts on hinders that interfere with concept expansion in design ideation. Through interviews, we learned that compositional creative thinking is primarily by two aspects, namely composition principles and the designer's experience. Composition principles here mainly refer to the correspondence between composition and ambience; a certain type of ambience is most appropriately

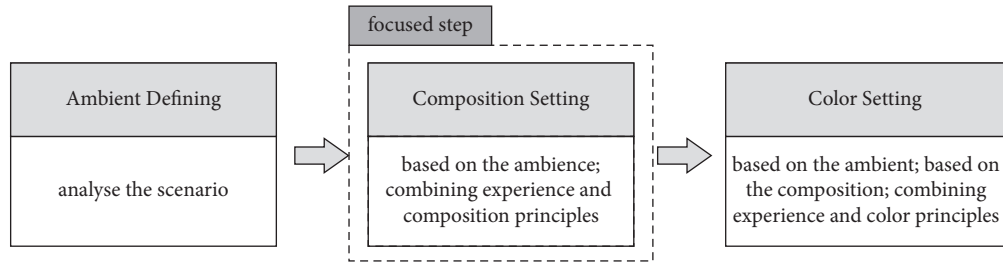


FIGURE 1: The process of ideation.

presented by corresponding compositions. The nature of composition principles is knowledge, which is obtained by the designer during their learning period. Another impact is previous experience, involving not only designers' own experience but also the experience of others. The main presentation of previous experience is visual material, such as completed graphics, photo references, and video clips, in which ambience is similar to the design output required to present.

Regarding the hints encountered in the ideation process, designers generally think that there are two types of challenges. One barrier is that the incomplete mastery of composition principles leads to barriers in the process of generating mental imagery in the early stage of conception ideation, often when the designer is a novice. Another one is that designers cannot quickly and accurately find materials that meet their needs to stimulate their creation effectively, through existing methods. Designers often use the engine to search for keywords to collect visual material. However, such an approach is usually suitable for the retrieval of elements with specific semantics. As the composition is an abstract word, the search engine cannot give the designer a satisfactory stimulus in the semantic-based search process, which makes it difficult for designers to collect the visual material closely related to the composition.

3.2.3. Ideal Solution. In response to the two challenges, the designers explain their own approach to overcoming these challenges and imagined the ideal solution without considering the constraints. For the first challenge about mental imagery, designers believe that in addition to a more thorough grasp of the aesthetic knowledge related to composition, they tend to also collect a large number of visual displays with similar features to mental imagery, such as black and white thumbnails or digital speed paintings. These materials are visual presentations of others' early ideas, although they do not contain rich details, but their general structures are clear, and their ambiguity provides designers with space for exploration. Therefore, when it comes to the desired solution, designers said that they hope that the computer can automatically generate mental image pictures based on composition principles to help them successfully pass through the early stage of ideation.

For another challenge related to collecting effective visual stimuli, designers explained that when they have a clear mental image, they will try to summarize keywords that accurately describe intentions and search for them. Usually,

designers expand the scope of keywords to expand the search scope of search engines. In terms of the ideal solution, the following two ideal solutions have been repeatedly proposed by more than half of the designers, one is to hope that by inputting the composition sketch, the computer can automatically match the complete design case similar to the composition of inputs; the other is to hope that by inputting some composition types, the computer can output materials diversely with similar composition.

Overall, designers have both semantic thinking about design issues and aesthetic experience based on visual thinking during the phase of ideation in environment concept design. Effective visual stimulation not only assists the designer to produce the appropriate mental imagery but also promotes the design process after designers create creative imagery, which leads designers to find the optimal design solution.

4. The AI-Based Approach

Based on the interview results in the last section and research findings in related work, we proposed three main goals of the AI-based method we introduced:

Goal 1: this method can quickly generate visual stimuli for the composition according to the needs of designers, which are diverse and numerous; goal 2: the visual stimuli generated by this method need to follow composition principles; and goal 3: the visual stimuli generated by this method need to be similar to a mental image. The framework of our approach is shown in Figure 2.

In goal 1, the generator is responsible for stimulus outputs, and its essence is image generation. Compared to other image generation methods, GAN [16] is able to generate images quickly with high quality. Thus, we will select the appropriate GAN variant models. Our requirement for the generation model is that the trained model can generate a great number of diverse images as visual stimuli in a short time. For the achievement of goal 2, we will train the model on a training set that follows the composition principles and had a single ambience feature to generate visual stimulus. In goal 3, to make the generated images similar to the mental imagery, which is ambiguous but clear in structure in general, we set the picture type of the data set to black and white thumbnail sketches. We then take the generation of visual materials with spectacular ambience as an example to introduce our method in respect of data collection and generator implementation in detail.

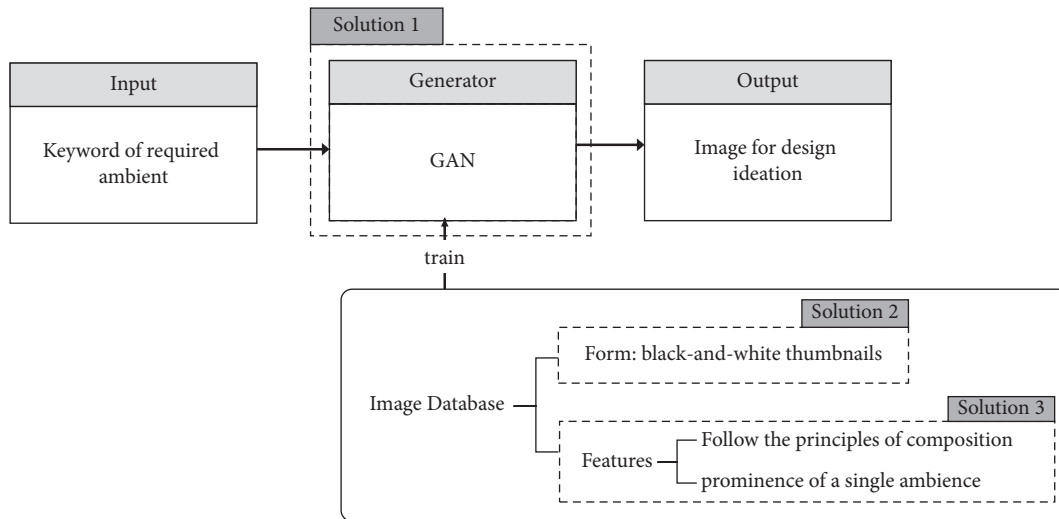


FIGURE 2: The framework of our approach.

4.1. Data Collection. The data used to train the model are from the Google Image search engine, and the training set consists of 153 black and white thumbnails. The reason why we chose thumbnails is that through interviews, thumbnails are generally considered by designers to be a visual presentation of mental images and the characteristics of overall structure clarity that allow designers to quickly obtain effective information from it. Meanwhile, due to the ambiguity brought about by its lack of details, designers are provided with a vast creative idea space, so this is an effective form of design stimulation. In addition, the black and white feature is to exclude the interference of color, as color is another major factor affecting the ambience, and the computer improves the quality of the generated images by extracting a single data feature. Since taking spectacular ambience as an example, the images that make up the data set need to have prominent features of spectacular ambience.

We summarized the characteristics of the composition: the cases that embody this ambience are mostly composed in the rule of thirds by learning the methods of environment concept design and studying the excellent cases recommended by designers with spectacular ambience. The guideline proposes that an image should be imagined as divided into nine equal parts by two equally spaced horizontal lines and two equally spaced vertical lines and that important compositional elements should be placed central grid or intersections of golden section lines [17], as shown in Figure 3. Generally, this composition not only can highlight the theme elements but also can increase the depth of the scene space through the relative blank picture outside the central grid to create a spectacular view. In addition, the themes of design cases with the ambience are mostly the surroundings of large buildings such as palaces and castles or natural landscapes such as mountain views.

After the study, we start to make a data set based on the compositional features. First of all, we searched on the Google image engine by inputting the keyword—conceptual environmental thumbnail sketches—and then browsed the images retrieved by the engine one by one, analyzed the

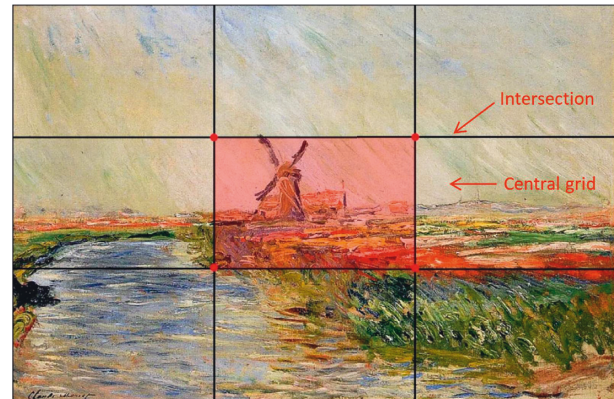


FIGURE 3: Composition of the rule of thirds.

composition method, and downloaded the cases that follow the rule of thirds. We obtained a total of 209 images during the period. Subsequently, we invited 3 professional designers to identify the ambience of these cases with us to form the data set for training. The process is shown in Figure 4.

4.2. Implementation of Generator. This section introduces the process of selecting a generator in detail and discusses the results of the outputs. In order to find suitable GAN models, this paper uses the prepared data set to train different GAN models, and then the researchers test the results from both duration for generation and diversity of outputs. The results of the test will determine whether the model can enter the next qualitative research. The qualitative study is conducted in a group discussion with design experts invited to participate in the interview in the third part. We analyze the results of the interview and draw a conclusion eventually.

4.3. Training and Testing of Different Models. During the next phase, we tested both the generation duration and the diversity of outputs of each model.

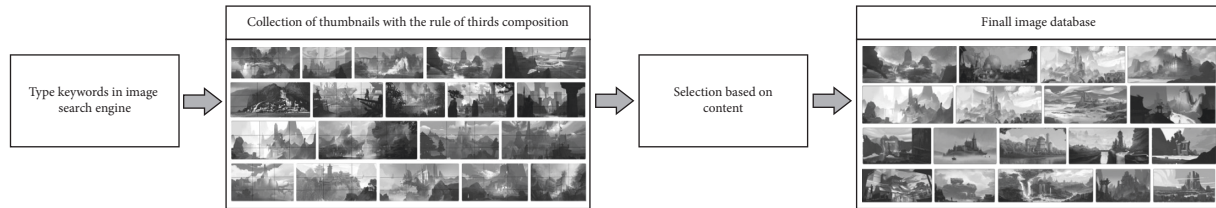


FIGURE 4: The process of making the data set.

A generative adversarial network (GAN) is a kind of method that can learn the deep representation of data without a large amount of annotation training data sets. By updating two networks through a backpropagation algorithm, the data with established characteristics can be generated in a competitive way. Although GAN has obtained remarkable achievements in image generation tasks, the classic GAN network still has problems such as unstable training and less quality and diversity of generated images. Therefore, how to obtain a stable and high-quality generative adversarial model has always been the focus of research. Several representative networks are follows.

Compared with traditional GAN networks, DCGAN [18] networks put forward a series of architectures to combine CNN with the original GAN network effectively, which improved the depth of the feature extraction network and the quality of generated images. WGAN [19] proposed a generative adversarial model based on Wasserstein distance, which not only solved the problem of training instability but also provided a reliable indicator of training to improve the quality of generated samples. LSGAN [20] adopted the same idea as WGAN. Specifically, the loss function of LSGAN was changed from cross-entropy loss to the least square loss function to obtain images with higher quality and diversity. WGAN-gp [21] focused on the training instability caused by the weight clipping strategy in WGAN. A method of gradient penalty was proposed to solve the above problem and make the training of WGAN more stable. For the effect of the local equilibrium state in training, DRAGAN [22] proposed a novel gradient penalty scheme, which effectively improved the stability of GAN. Besides, a novel progressive GAN training model is proposed in PGGAN [23], which solved the problem of high cost in generating high-resolution images under the original GAN framework and greatly improved the quality of generated high-resolution images.

In summary, based on the self-made data sets, we have tested the above six GAN variant networks with certain advantages in model stability, generated image quality, and diversity, including DCGAN, WGAN, LSGAN, WGAN-GP, DRAGAN, and PGGAN. Then we analyzed the generated results in detail. The characteristics, corresponding advantages, and training parameters of the six models are shown in Table 1.

4.3.1. Generation Duration Test. We ask each of the six trained models to generate 100 images, and some of the outputs are shown in Figure 5. The mean and standard

deviation (in second) of each model are shown in Table 2. Among them, the average generation duration of WGAN models is the shortest and most stable. The passing criteria for this testing are as follows: the generation duration of a model is less than or equal to 10 seconds. Otherwise, it will be eliminated because it does not meet the requirements of goal 1—quick generation of numerous visual stimuli. The results reveal that the generation duration of each six models is less than 10 seconds, so they all pass the test.

4.3.2. Diversity Test. In this test, we calculate the diversity rate for 100 outputs of each model in the generation duration test, formula as follows:

$$\text{Diversity rate} = \frac{\text{The number of filtered images}}{\text{The total number of image}}. \quad (1)$$

In the process of filtering images, we regard two or more pictures that are similar or identical to each other as one and delete the extra pictures to finally get the total number of pictures after filtering. The criterion for judging whether pictures are similar is that the black and white blocks of two or more pictures are distributed consistently together with the same spatial level. The criterion for passing the test is that the diversity rate is higher than or equal to 60%; otherwise, it will be eliminated as it does not meet the requirement in goal 1, generation of diverse visual stimuli. The diversity test results are shown in Figure 6, from which it can be seen that the diversity of LSGAN and DRAGAN is less than 60%, so they are eliminated and cannot enter the quantitative analysis stage. As for the causes for the low diversity, we speculate that it may be because the data set is not rigorously de-reprocessed, where the images with high similarity are removed, resulting in the overfitting of the models.

4.3.3. Qualitative Research. The qualitative study is conducted in a group discussion with eight design professionals who participated in the expert interviews before. The reason why the format of the group discussion is because we hope that designers have sufficient freedom to express their views under the given topics, which can help us obtain the advantages and disadvantages of the designers' analysis of generated images of the four models, so as to summarize the designers' concerns and requirements for AI-generated visual stimuli.

The interview process is as follows: we describe the design requirement—designing a spectacular environment for the surroundings of a large building, to the participants in detail and give them five minutes for composition

TABLE 1: The characteristics, corresponding advantages, and training parameters of the six models.

Model	Feature of model	Corresponding advantages	Training parameter			
			Epoch	Learning rate	Optimizer	Total parameter (M)
DCGAN	Effective combination of GAN and CNN	(1) Improvement of the depth of the feature extraction network; (2) enhancement of image resolution	8×10^3	2×10^{-3}	Adam	12.33
WGAN	Wasserstein-distance-based GAN	(1) Improvement of training stability; (2) improvement of generated image quality	8×10^3	2×10^{-4}	Adam	23.36
LSGAN	Replacement from cross-entropy loss to least squares loss	(1) Mitigation of gradient disappearance problem; (2) improvement of generated image quality	8×10^3	2×10^{-4}	Adam	23.36
WGAN_gp	Introduction of the gradient penalty	Improvement of training stability	8×10^3	2×10^{-4}	Adam	23.36
DRAGAN	Introduction of a novel gradient penalty scheme	Improvement of stability	8×10^3	2×10^{-4}	Adam	23.36
PGGAN	Introduction of a novel progressive GAN training model	Solution of high loss problem for generating high-resolution images	8×10^5	2×10^{-3}	Adam	46.14



FIGURE 5: Some outputs of six models.

TABLE 2: Mean and standard deviation of generation duration.

Model	Mean	±	Standard deviation
WGAN	1.8132	±	0.0224
DRAGAN	1.8728	±	0.0614
LSGAN	1.9011	±	0.0675
WGAN_gp	2.0201	±	0.0482
DCGAN	2.0358	±	0.5322
PGGAN	5.3076	±	0.0364

ideation. Then, we present the 100 resized images as design stimulus generated by the four models in advance to the designers, and some of the images are shown in Figure 7.

There are two topics in the discussion: (1) analysis of advantages and disadvantages of the generation results of each four models and (2) horizontal comparison of the generation

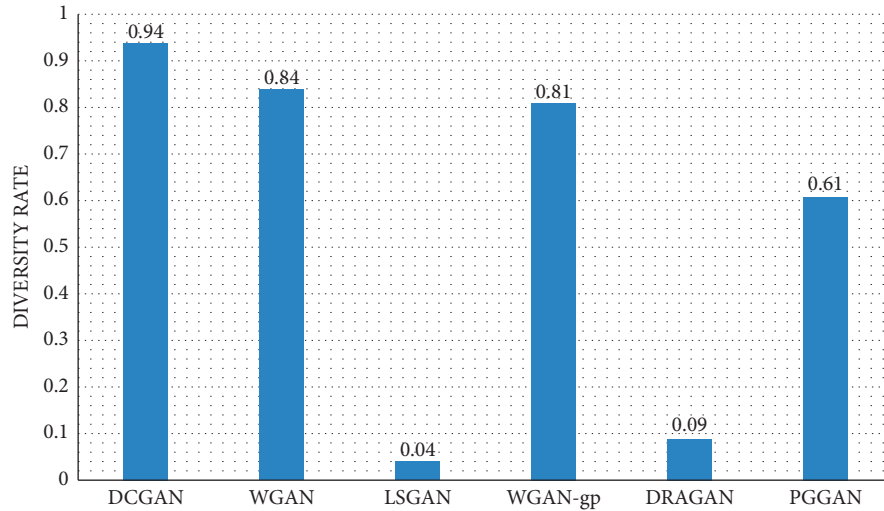


FIGURE 6: Diversity rate of models.

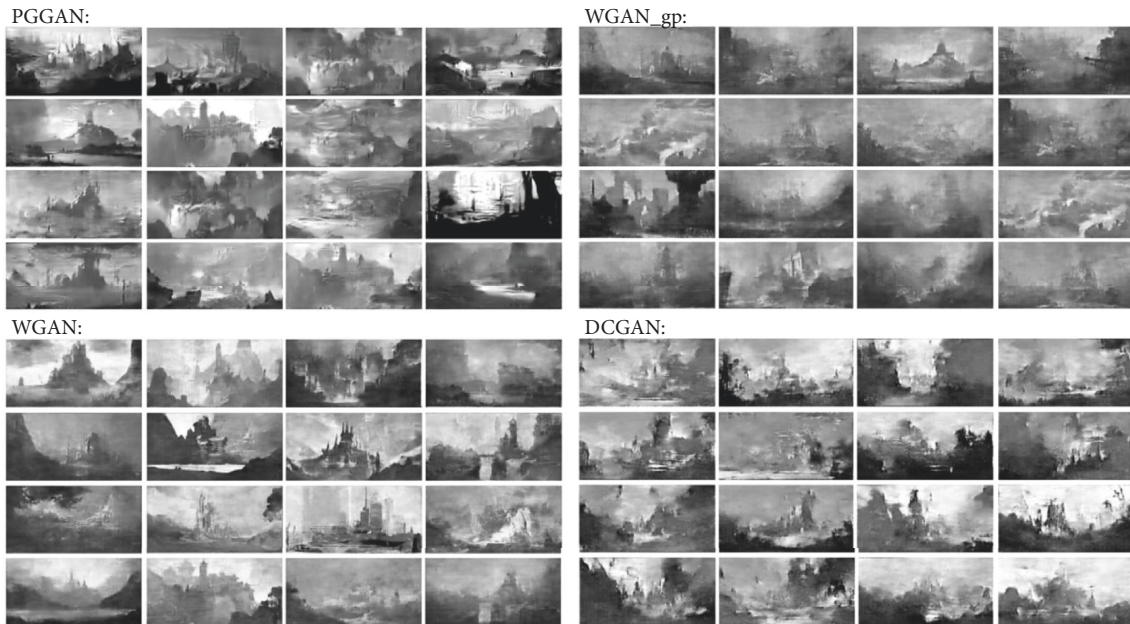


FIGURE 7: Some of the generated images as stimuli.

TABLE 3: The advantages and disadvantages of the four model generation results.

Model	Strengths of model	Weakness of model
PGGAN	The spatial level of the picture is rich; the clarity is high; and the position of the center in the vision basically follows the principle of composition.	The contour line of the visual center is not clear, and its small distinction from the surrounding elements is easily misleading.
WGAN	The pictures have rich spatial levels, high definition, and clear contour lines of the subjects whose position follows the principles of composition so that it is easy to distinguish from other elements.	The distribution of elements in the picture is monotonous, and the diversity is inadequate.
WGAN_gp	The completeness of the pictures is high, and the position of the main objects basically follows the principle of composition.	There is lack of spatial level, small difference between the main object and other elements, as well as lack of clarity in the pictures.
DCGAN	The completeness of the pictures is high with rich spatial levels.	The elements of the picture are ambiguous and illegible, which are arranged in a haphazard manner that does not follow the principles of composition.

results of each four models, and finally, they have to form a unified opinion. Discussions take place online, using online conferencing software with video and audio access for an unlimited period of time. The discussion is recorded by two researchers, and at the end of the discussion, the researchers confirm the advantages and disadvantages of the four models' outputs and the uniform results of the satisfaction ranking with designers and then obtain the final results.

The advantages and disadvantages of the outputs are shown in Table 3.

The designers' satisfaction with the outputs of the four models is sorted from highest to lowest: WGAN and PGGAN, WGAN-pg, and DCGAN. In general, in terms of coherence and completeness, there is no fragmented or split picture in the results of the four models, so it is basically in line with expectations. Besides, based on the results of the discussion, we can find that satisfaction with the generated stimuli of different models is mainly affected by spatial levels, the recognizability of main objects, the diversity of the arrangement of elements, and their responsiveness to composition principles. These factors are weighted equally by designers, so satisfaction with the results of model generation depends on how well the model responds positively to these factors. In addition, the designers claim that the visual stimuli generated by the four models are basically a response to the design requirement.

4.3.4. Results Analysis. We summarize five factors mentioned above that affect the quality of generated visual stimuli—completeness, spatial levels, recognizability of main objects, diversity of the arrangement of elements, and responsiveness to composition principles, which can also be regarded as attributes of generated visual stimuli that are the focus of designers. This paper will explain these five as follows:

- (i) **Completeness:** In environment concept design, the composition emphasizes integrity; it is the preliminary conception of the layout and organization relationship of elements. Therefore, completeness directly affects the designer's initial impression of the visual stimulus, and the highly complete image can clearly present the distribution of the elements, attracting the designer to further observe.
- (ii) **Spatial levels:** The essence of environment design is the design of space, so designers need to consider not only the elements in the conception but also the way the elements are presented in space, which is a process of space planning. In general, an excellent scene design presents at least three levels of space, namely close-up, medium-shot, and far-range. In the stimulus generated by the model used in this paper, the spatial level is mainly reflected in the difference between the black, white, and gray levels. WGAN-pg is because the gap between black, white, and gray levels is not large, resulting in the space levels are not rich enough, only two layers or even one layer, thus becoming the object of criticism by designers.

- (iii) **Recognizability of main objects:** The main object occupies the visual center of the picture, which is the most eye-catching part and the area with the most detail in an image. Therefore, designers tend to define the position of elements initially, and then design subjects. In the generated visual stimulus, the recognition of main objects depends on the details of the outline of the subject and the difference with the surrounding elements in the black and white level. Thus, high recognizability can help designers successfully generate ideas about the subjects.
- (iv) **Diversity of the arrangement of elements:** Diversity refers to the diversity of the arrangement and combination of elements under the premise of following the principle of composition. In this paper, the generated visual stimuli should follow the principle of nine-square grid composition, which stipulates the position and proportion of main objects. But the proportion, arrangement, and combination of other elements are free, which is also the embodiment of the diversity of visual stimulus composition.
- (v) **Responsiveness to composition principles:** It refers to whether the picture follows the composition principle or to what extent it follows the composition principle, which is an important criterion for designers to judge whether the visual stimulus meets its aesthetic standards.

In addition, we also find that the ambiguity of the elements needs to be controlled within a certain range. However, existing studies of visual stimuli have shown that the ambiguity of stimuli can stimulate designers' creativity by expanding the space for the exploration of design solutions. However, in the environment concept design, excessive ambiguity not only cannot stimulate the designer's creative thinking but also will become an obstacle in the ideation process. For example, the images generated by DCGAN because of the high degree of ambiguity of the elements lead designers to claim that the model is flawed.

Overall, this paper realizes goals 2 and 3 by making a model training data set that meets the requirements of designers and achieves goal 1 by testing different GAN variants to select appropriate ones as the generator. In the qualitative analysis of the results of GAN variant generation, WGAN and PGGAN were considered by the designers to be the most suitable models for generating visual stimuli. Although the method proposed in this paper has successfully achieved the stated goals, there are still shortcomings. The problem is that the composition form of visual stimulation under a certain ambience is relatively single, which is easy to trigger design fixation. The main reason for this problem is that this paper employs a "one-to-one" method when making the data set, which means one ambience corresponds to a fixed composition, such as the spectacular ambience corresponds to the composition in the rule of thirds. Therefore, the process to make the

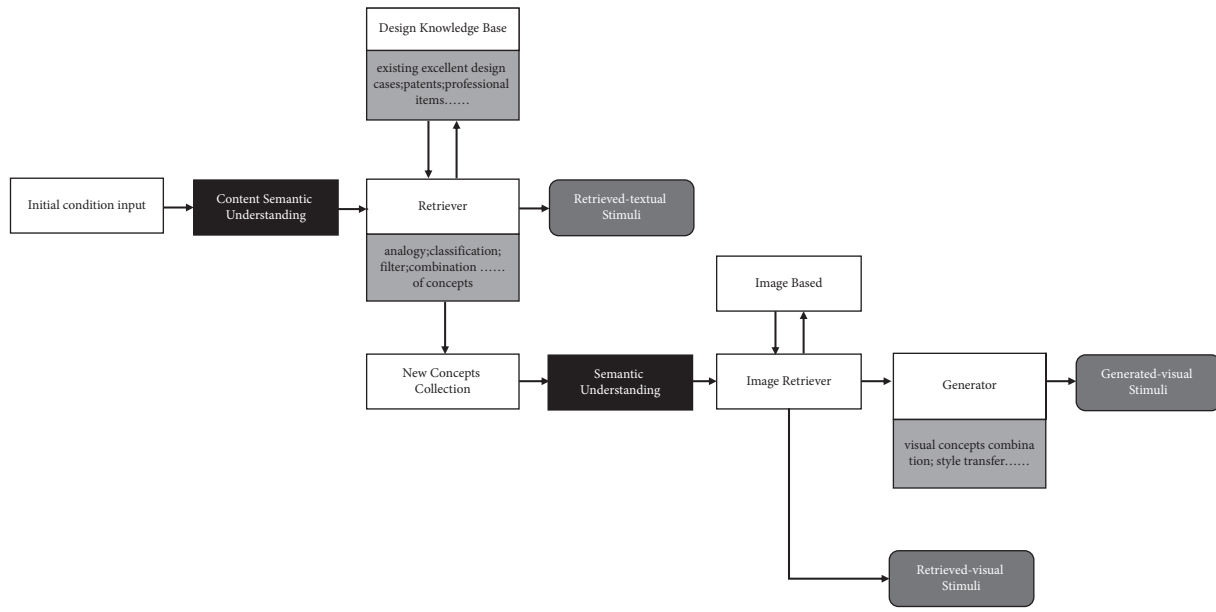


FIGURE 8: The general process of the existing method for producing design stimulus.

data set is to first filter the composition and then check whether it corresponds to the ambience. However, when actually solving the problem of environment concept design, there is a diverse correspondence between ambience and composition principles. In the case used in this paper, for example, the spectacular ambience can also be expressed through other compositions or atypical compositions in the rule of thirds. Limiting the presentation of the ambience to an exclusive composition is bound to reduce the diversity of composition. Therefore, exploring the correspondence between ambience and composition principles and finding a more suitable data set production process are the problems that need to be further solved in this paper.

5. Conclusion

This section will discuss the expansion of AI-based approaches for design ideation in two aspects, forms and the process of methods, as well as make concluding remarks.

5.1. Forms. As far as the existing method is concerned, according to the form of the stimulus produced, it can be divided into textual materials and visual materials. Based on the process of producing stimuli, it can be categorized into retrieval stimuli and generative stimuli. Thus, there should be four types of stimuli, namely retrieval textual stimuli, retrieval visual stimuli, generated textual stimuli, and generated visual stimuli. Producing retrieval stimuli has always been researching hotspot, while there are relatively few methods for generated stimuli. However, with the continuous development of the generative model, the efficiency of the generative model is getting higher and higher, and the tasks that can be finished by the generative model are becoming more and more complicated. Therefore, it is worth

exploring the potential of generative models and making them do their best to support the creative thinking of designers. Additionally, producing more novel forms of design stimuli by AI, such as sound, is also a direction in further study.

In addition, both retrieval stimuli and generated stimuli have their own advantages and disadvantages. The advantage of the retrieval stimulus is that the quality is stable; the disadvantage is that retrieval stimulus usually roots from existing cases and cannot be adjusted according to the requirements of the designer, so the plasticity is not strong. On the contrary, the advantage of generative stimuli is that they are highly malleable, and designers can adjust the parameters of generative stimuli according to their own needs, but the quality of generative stimuli is unstable, and low-quality design stimuli will become an obstacle in the design ideation. Therefore, a future study should focus on how to integrate the two types of design stimuli in a way that fully exploits their respective advantages while avoiding the negative impacts of doing so.

5.2. The Process of Methods. This paper summarizes the production methods of existing stimuli, as shown in Figure 8. It can be seen that producing the four forms of stimuli must be supported by semantic-level understanding and semantic knowledge base in the existing methods. This is because the semantic area in the brain plays an important role in creative thinking. Especially, when engaged in conceptual expansion, the brain's semantic processing network operates on overdrive, particularly in the higher-order regions that mediate lexical selection, controlled retrieval, combination, and integration processes [24]. In the field of artificial intelligence, with the development and improvement of NLP and computing creativity, computers can simulate the function of the brain's semantic network

through algorithms, assisting designers in concept expansion. This is the reason why existing methods are mostly based on semantic networks.

However, conceptual extensions guided by visual thinking are also a key part of solving design problems. After the qualitative research of the fourth part, we find that in the composition ideation, the five factors that affect the effectiveness of generative visual stimulus are derived from the aesthetic evaluation of the picture in the field of visual art. Thus, we claim that when visual-based design problems, effective design stimulus, in addition to semantically supporting the expansion of concepts, also needs to bring a good aesthetic experience to designers visually. It can be seen that in the method development, researchers should conduct a comprehensive analysis of the object served by the method and the problems it needs to solve, and the method of inspiring design ideas can be based on aesthetic knowledge in addition to the semantic level. Creative stimulation can solely effectively serve the designer's cognitive process in order to achieve the true sense of supporting design ideation.

In conclusion, conceptual expansion is a crucial part of design ideation, in which effective stimulus can inspire designers to find the optimal solution. In AI-based approaches to design ideation, AI can improve the effectiveness of design stimuli by producing materials that are more in line with the designer's needs. This paper starts from the visual-based design problem of environment concept design, identifies the challenges, and obtains ideal solutions of hints through expert interviews. Through the summary and induction of the interview results, this paper summarizes the three goals of the new method. Subsequently, this paper achieves the stated goals by making its own data sets and selecting the appropriate generation model. Subsequently, we quantitatively analyzed the outputs of the selected generative model through expert discussions and summarized the five main factors affecting the designer's satisfaction with the stimulus generated by AI. The key to supporting designers in their creative thinking by exposing them to stimuli is to improve the effectiveness of design stimuli, and the reasons that affect effectiveness are complex and changeable. Therefore, in order to improve the effectiveness of design stimulation, it is necessary to have a deep understanding of the process of designers to design and conceive, find the difficulties encountered by designers in design ideation, and find the entry point of artificial intelligence to solve the difficulties, in order to achieve the true sense of auxiliary creative ideas.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported in part by the Science and Technology Demonstration Project of Social Development of Jiangsu Province, under Grant no. BE2019631.

References

- [1] W. Hsu and B. Liu, "Conceptual design: issues and challenges," *Computer-Aided Design*, vol. 32, no. 14, pp. 849-850, 2000.
- [2] H. Casakin and G. Goldschmidt, "Expertise and the use of visual analogy: implications for design education," *Design Studies*, vol. 20, no. 2, pp. 153-175, 1999.
- [3] Y. Duan, J. Zhang, and X. Gu, "A novel paradigm to design personalized derived images of art paintings using an intelligent emotional analysis model," *Frontiers in Psychology*, vol. 12, 2021.
- [4] L. Chen, P. Wang, H. Dong et al., "An artificial intelligence based data-driven approach for design ideation," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 10-22, 2019.
- [5] B. Jonson, "Design ideation: the conceptual sketch in the digital age," *Design Studies*, vol. 26, no. 6, pp. 613-624, 2005.
- [6] G. Goldschmidt and M. Smolkov, "Variances in the impact of visual stimuli on design problem solving performance," *Design Studies*, vol. 27, no. 5, pp. 549-569, 2006.
- [7] C. Cardoso and P. B. schaub, "The influence of different pictorial representations during idea generation," *Journal of Creative Behavior*, vol. 45, no. 2, pp. 130-146, 2011.
- [8] W. S. W. Tseng, "Can visual ambiguity facilitate design ideation?" *International Journal of Technology and Design Education*, vol. 28, no. 2, pp. 523-551, 2018.
- [9] S. H. Jang, B. Oh, S. Hong, and J. Kim, "The effect of ambiguous visual stimuli on creativity in design idea generation," *International Journal of Design Creativity and Innovation*, vol. 7, no. 1-2, pp. 70-98, 2019.
- [10] S. Laing and M. Masoodian, "A study of the influence of visual imagery on graphic design ideation," *Design Studies*, vol. 45, pp. 187-209, 2016.
- [11] Y. C. Tang, J. J. Huang, M. T. Yao et al., "A review of design intelligence: progress, problems, and challenges," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 12, pp. 1595-1617, 2019.
- [12] J. Hao, Y. Zhou, Q. Zhao, and Q. Xue, "An evolutionary computation based method for creative design inspiration generation," *Journal of Intelligent Manufacturing*, vol. 30, no. 4, pp. 1673-1691, 2019.
- [13] F. Shi, L. Chen, J. Han, and P. Childs, "A data-driven text mining and semantic network analysis for design information retrieval," *Journal of Mechanical Design*, vol. 139, no. 11, 2017.
- [14] S. Sarica, B. Song, J. Luo, and K. L. Wood, "Idea generation with technology semantic network," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 35, no. 3, pp. 265-283, 2021.
- [15] X. Li, J. Su, Z. Zhang, and R. Bai, "Product innovation concept generation based on deep learning and Kansei engineering," *Journal of Engineering Design*, vol. 32, no. 10, pp. 559-589, 2021.
- [16] I. Goodfellow, J. P. Abadie, M. Mirza et al., "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139-144, 2014.

- [17] S. A. Amirshahi, G. U. H. Leichsenring, J. Denzler, and C. Redies, "Evaluating the rule of thirds in photographs and paintings," *Art & Perception*, vol. 2, no. 1-2, pp. 163–182, 2014.
- [18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Computer ence*, 2015.
- [19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, <https://arxiv.org/abs/1701.07875>.
- [20] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2794–2802, Venice, Italy, October 2017.
- [21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [22] N. Kodali, J. Abernethy, J. Hays, and Z. Kira, "On Convergence and Stability of GANs," 2017, <https://arxiv.org/abs/1705.07215>.
- [23] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," 2017, <https://arxiv.org/abs/1710.10196>.
- [24] A. Abraham, "Creative thinking as orchestrated by semantic processing vs. cognitive control brain networks," *Frontiers in Human Neuroscience*, vol. 8, 2014.

Research Article

Quality Risk Management Algorithm for Cold Storage Construction Based on Bayesian Networks

Yaping Song  and Zhanguo Wei 

School of Logistics and Transportation, Central South University of Forestry and Technology, Changsha 410004, China

Correspondence should be addressed to Zhanguo Wei; t20110778@csuft.edu.cn

Received 4 May 2022; Revised 2 June 2022; Accepted 4 June 2022; Published 24 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yaping Song and Zhanguo Wei. 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 cold storage construction project, only by controlling the quality risk of the project can ensure that the cold storage can meet the expected use function and achieve the expected economic benefits after the completion of the cold storage. In order to effectively ensure the key pivot role of cold storage in cold chain logistics, a cold storage construction quality risk management system is constructed to identify and analyze quality risk factors from three dimensions: construction procedures, participating units, and work processes, construct a cold storage construction quality risk evaluation model based on Bayesian network, and through reverse reasoning analysis and sensitivity analysis, key quality risk factors are derived: inadequate quality assurance system, technical delivery is not in place, mismatch of building materials and equipment, inadequate training of skilled workers, completion acceptance is not careful or acceptance standards are unreasonable, and duration does not meet the requirements. Finally, in view of the above quality risks, suggestions and measures are put forward from five aspects: man, material, machine, method, and environment.

1. Introduction

In recent years, with the continuous improvement of people's living standards and the rapid development of fresh e-commerce, the demand for fruit and vegetable agricultural products continues to increase. Cold storage is the central link and necessary infrastructure of cold chain logistics, and its construction demand also shows a clear upward trend under the dual promotion of the market and policy. The quality requirements act on the project process at the same time, bringing quality risks into the execution. In addition, cold storage is a complex system facing various unknown risks, so there are more specific and specialized quality risks in the cold storage construction process, which need to be targeted for quality risk management. The external environment of the project is highly variable and unpredictable, coupled with the limitations of human understanding and forecasting ability, which may lead to incomplete cognition and lack of data information in quality risk analysis. The use of Bayesian networks, in this case, is very appropriate

because the Bayesian network approach reflects the probabilistic relationship model between the data in the entire database, the lack of a data variable can still build an accurate model. Moreover, the Bayesian network can update the estimated values of all other unknown nodes in the whole network based on the observed node values at any time. For a research object like a cold storage construction project, which is full of uncertainty risks, applying a Bayesian network for quality risk management is in line with the requirements.

After years of development, Bayesian networks have proven their value in many fields, including medical diagnosis, credit evaluation, risk assessment, reliability analysis, prediction, and troubleshooting. A part of experts and scholars used Bayesian networks for various aspects of risk assessment [1–9]. Also, the Delphi method has been utilized along with Bayesian networks to implement risk assessment studies [10, 11]. Some other researchers have combined Bayesian networks with fuzzy set theory in order to perform risk evaluation analysis more easily and quickly [12, 13].

TABLE 1: Quality risk factors of cold storage construction procedures.

Construction procedures	Quality risk factors
Decision-making stage	Project proposals and feasibility studies are not reasonable
Design work stage	Incomplete design content, defective design, errors and omissions, inappropriate specifications, failure to consider geological conditions, failure to consider construction possibilities, etc.
Engineering construction stage	Backward construction techniques, unreasonable construction techniques, and solutions, improper construction safety measures, failure to apply new technologies and solutions, failure to consider site conditions, etc.
Completion acceptance and delivery stage	Operation and maintenance risks
Full stage	Improper coordination of the parties involved

Some experts and scholars used the combination of Bayesian networks and fuzzy hierarchical analysis to achieve their research objectives [14, 15]. Jianxing [16] proposed an intuitive fuzzy Bayesian network-based fault assessment method for process systems in response to imprecise and inadequate historical data. Han [17] constructed a disaster chain hazard assessment model combining the Bayesian network model and ArcGIS program software, then determined hazard chain probabilities and hazard intensities of seismic events using probabilities obtained from Bayesian networks, and produced disaster chain hazard maps using ArcGIS. Li [18] predicted the impacts of land use and climate change on riverine macroinvertebrates based on the linkage of structural equation models with Bayesian networks. Chen and Huang [19] introduced Bayesian networks for flight crew performance assessment to provide data support for interventions in human error management for aviation safety. Mendes [20] used Bayesian networks in context of value-based software engineering to estimate the value of decisions. As for cold storage, Chen [21] prioritized alternative points by analyzing the distribution of cold storage in the northwest region of Zhengming Modern Logistics Company using an improved gray correlation model. Kuźmicki [22] introduced the engineering analysis of cold storage for aquatic products. Chukwu and Adibe [23] performed a quality assessment of cold chain storage facilities for regulatory and quality management compliance in developing countries. Miao and Zhang [24] studied the energy consumption evaluation method and energy-saving operation technology of cold storage.

Not much research has been done on cold storage construction, and in view of this, this paper applies Bayesian networks to develop a study on the quality risk of cold storage construction.

The main contributions of this paper are as follows:

- (1) Based on the Bayesian network, this paper identifies and corrects the quality risk factor indicators and enriches the theoretical system of quality risk management for cold storage construction projects.
- (2) This paper gradually changes the quality risk management work from subjective qualitative analysis to visual, quantitative analysis, which is conducive to improving the visualization of quality risk management.
- (3) It provides a reference for the quality management of cold storage construction projects, which is conducive to improving the construction quality of cold storage.

This paper is organized as follows: Section 2 introduces the quality risk management system for cold storage construction, Section 3 describes the proposed algorithm, Section 4 covers the simulation study, Section 5 presents the experiments and results, and Section 6 concludes the paper.

2. Cold Storage Construction Quality Risk Management System

In this paper, we will study the quality risks in the construction process of cold storage engineering projects from three dimensions: construction procedures, participating units, and work processes. Firstly, we identify and analyze the quality risk factors of which each stage and each participating unit needs to bear the corresponding responsibility, use Bayesian networks for structure learning and parameter learning to determine the probability of occurrence of each quality risk, and find out the key quality risk factors through reverse reasoning and sensitivity analysis, and finally propose effective quality risk control measures, in order to be able to provide a reference for the quality management of cold storage construction and to help achieve the goal of safe production, so as to give full play to the role of the cold storage itself and ensure the normal operation of the storage and preservation function.

2.1. Construction Procedures. Considering the complexity of the research object of this paper and taking into account the actual situation, this paper will carry out the research process in four stages: decision-making stage, design work stage, engineering construction stage, and completion acceptance and delivery stage, and combine the four stages to analyze the quality risk factors from the whole process. By reading the relevant literature, it can be concluded that the main quality risks existing in each stage are shown in Table 1.

2.2. Participating Units. Quality risk management of engineering construction projects reflects the characteristics of diversified subjects. Construction units, design units, building units, supervision units, etc., are responsible for the

TABLE 2: Quality risk factors of the units involved in cold storage construction.

Participating units	Quality risk factors
Construction units	Poor investor decisions
Design units	The professional quality of the design party is not high
Building units	Construction level is not in accordance with the specifications
Supervision units	Inadequate supervision by the supervisor

TABLE 3: Quality risk factors in cold storage construction work processes.

Work processes	Quality risk factors
Preparation of technical data and documents	Inadequate survey information Unqualified construction organization design document preparation
Design delivery and drawing review	Inadequate design delivery and drawing review
Procurement and subcontracting	Material procurement does not meet specifications Inadequate control of subcontracting services
Technical presentation	Inadequate technical presentation
Engineering surveying	Deviations in measurement results
Completion and acceptance	Unreasonable acceptance criteria

quality risk management of engineering construction projects. Therefore, this paper delineates the participating units in cold storage construction as construction units, design units, building units, and supervision units and studies the quality risk factors that exist in terms of the dimensions of the participating units. A review of the literature shows that the main quality risk factors that may exist in the participating units in cold storage construction are shown in Table 2.

2.3. Work Processes. From the perspective of the work content involved in the cold storage construction project, there are a series of interrelated and mutually influencing workflows that require quality risk management from the decision-making stage to the completion acceptance and delivery stage. In order to ensure the smooth progress of each work process, it is necessary to analyze the quality risk factors that may exist in the process. According to relevant references, the quality risk factors existing in the cold storage construction work processes are shown in Table 3.

3. Proposed Algorithm

The seminal work on Bayesian networks is the book written by British mathematician Thomas Bayes: *An Essay toward Solving a Problem in the Doctrine of Chances*. In 1958, the British statistical journal *Biometrika* republished the Bayesian paper in full. In the 1950s, the combination of empirical Bayesian methods with classical methods in the estimation of econometric models, represented by H. Robbins, attracted widespread attention. In 1985, Judea Pearl first proposed the Bayesian network model and applied it to artificial intelligence aspects for probabilistic inference.

3.1. Bayesian Theory. The process of Bayes' theorem can be summarized as follows: "past experience" plus "new evidence" yields "modified judgment." It provides an objective method of combining newly observed evidence with existing experience to make inferences.

Suppose there are random events A and B , their conditional probability relationship can be expressed by the following mathematical formula:

$$P(B|A) = P(B) \frac{P(A|B)}{P(A)}. \quad (1)$$

The event B is the target event to be examined and $P(B)$ is the initial probability of the event B , called the prior probability. A is a newly emerged event that will affect event B , $P(A)$ denotes the probability of event A occurring. $P(A|B)$ denotes the probability of A when B occurs, which is a conditional probability, and $P(B|A)$ denotes the probability of B when A occurs, which is both a conditional and a posteriori probability.

According to the Bayesian formula, the prior probability is generally the probability data obtained from previous data analysis or statistics. The posterior probability is the probability of occurrence under certain conditions and is the probability that is re-corrected after the information is obtained. In other words, the posterior probability can be corrected and obtained on the basis of the prior probability.

Suppose B_1, B_2, \dots, B_i is a division of the sample space Ω , then for any event A ($P(A) > 0$), there is the following:

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}. \quad (2)$$

The above formula is the Bayes formula, B_i is often regarded as the cause of the test result A , $P(B_i)$ ($i = 1, 2, \dots$) indicates the probability of various causes, so it is called the prior probability; $P(B_i|A)$ reflects the new understanding of the probabilities of various causes after the test has produced results, so it is called the posterior probability.

3.2. Definition. A Bayesian network is a causal network graph, which mainly consists of two parts: a directed acyclic graph and a conditional probability table. A directed acyclic graph is a non-closed relational graph with directionality, consisting of several nodes and directed arcs between nodes.

The details are shown in Figure 1. The nodes represent random variables in the knowledge domain, and the directed arcs represent the interdependencies between the nodes, which are directed from the parent nodes to the child nodes. The conditional probability table is a quantitative description of the dependencies between the node variables. Its expression is as follows:

$$F_{BN} = \langle G, p(x) \rangle. \quad (3)$$

$G = (I, E)$ denotes a directed acyclic graph, where I represents the set of all points in the graph and E represents the set of directed connected line segments. And let $X = (X_i)_{i \in I}$ be the random variable represented by a node i in its directed acyclic graph, if the joint probability distribution of the variable X can be denoted as $p(x) = \{p(X_i | X_{pa(i)})\}$, namely the conditional probability table is $p(x)$, where $pa(i)$ denotes the cause of node i , or $pa(i)$ is the parents of i .

Furthermore, for any random variable, its joint probability can be obtained by multiplying the respective local conditional probability distributions:

$$p(x_1, \dots, x_i) = p(x_i | x_1, \dots, x_{i-1}) \cdots p(x_2 | x_1) p(x_1). \quad (4)$$

3.3. Characteristics. The Bayesian network can use the knowledge obtained under uncertain and complex conditions to infer logical results and has the ability of system modeling, reasoning, and diagnosis. It is the most effective way to deal with event polymorphism and uncertainty in data mining technology. As an ideal analysis tool, the Bayesian network based on probabilistic reasoning has great advantages in representing and solving decision-making problems with uncertain factors and can solve the faults caused by the uncertainty and correlation of complex equipment, so it can be used in multiple widely used in the field.

The Bayesian network has the following advantages: the estimated values of all other unknown nodes in the entire network can be updated at any time according to the observed node values; the causal relationship between nodes can be displayed intuitively, and modeling can be based on expert knowledge in the case of insufficient data; two-way reasoning can be performed, either from the cause to the result or from the result to the cause. Such reverse reasoning ability cannot be achieved by other classical probabilistic reasoning methods; reasoning can be done in the case of incomplete data because the Bayesian network method reflects the probability relationship model between the data in the entire database; it can combine multiple types of data, such as subjective empirical data and objective data; the results of quality risk analysis are given in probabilistic form, which is more intuitive and reliable.

3.4. Bayesian Network Structure Learning. Bayesian network structure learning refers to the determination of a structural model, namely a directed acyclic graph, by experts in the relevant field based on the relationships between things. The selection of a suitable and appropriate structure learning method helps to construct the optimal Bayesian network

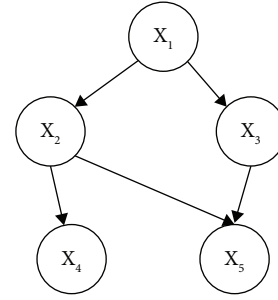


FIGURE 1: Directed acyclic diagram.

structure model. There are three general Bayesian network structure learning methods: First, based on expert knowledge construction, using the empirical knowledge of experts in related fields to analyze the connection relationship between nodes and determine a Bayesian network structure that meets reality. Second, based on data learning, the Bayesian network topology is obtained by training on previous sample data. Third, the sequence of nodes is determined based on expert knowledge and experience, and the structure of the Bayesian network is determined with a suitable algorithm. Due to the specificity of the research object and the restriction of the sample size, this paper chooses to construct the Bayesian network topology based on expert knowledge and experience combined with the questionnaire survey.

3.5. Bayesian Network Parameter Learning. Bayesian network parameter learning refers to determining the conditional probability distribution at each node of a Bayesian network model for given Bayesian network topology, using the expert's prior knowledge combined with the actual situation. For Bayesian network parameter learning, there are two common methods: the maximum likelihood estimation method and the Bayesian statistics method. Both methods need to meet the sample data independently and identically distributed, and the biggest difference is that the estimated parameters are different. Among them, the maximum likelihood estimation method regards the parameter to be estimated as a fixed form of an unknown variable, and the solution process is not constrained by previous knowledge and experience. In contrast, the Bayesian statistical method regards the parameter as a random variable with some known prior distribution and can refer to previous knowledge and experience. Given that the research object of this paper is a cold storage construction project, we need to fully cross-reference similar engineering projects and draw on the relevant experience of engineers, so we choose Bayesian statistics for parameter learning of Bayesian networks in this paper.

4. Simulation Study

4.1. Risk Identification. In this paper, a questionnaire survey was issued to relevant professionals in the industry to achieve the identification of quality risk factors in cold storage construction projects, the partial list of which is shown in Table 4. Practitioners include project managers,

TABLE 4: Partial list of quality risk factors of cold storage construction.

Construction procedures	Participating units	Work processes
A-decision-making stage	Construction units	A1-unreasonable preparation of project proposal A3-insufficient financing

B-design work stage	Design units	B1-construction drawings are not designed to meet construction and billing requirements and do not take into account construction possibilities B3-frequent design changes

C-engineering construction stage	Building units	C1-inadequate construction preparation C2-construction does not meet completion standards

D-completion acceptance and delivery stage	Construction units, design units, building units, supervision units	D1-completion acceptance is not careful or acceptance standards are unreasonable Acceptance standards are unreasonable D2-the technical information is not complete
	Building units	...

E-full stage	Supervision units	E5-supervision is not effective

designers, construction technicians, supervision engineers, etc., who have rich experience in cold storage construction. The main content of the questionnaire is divided into the probability of occurrence of quality risks in cold storage construction projects and the degree of impact of the risks, which are divided into five levels.

4.2. Risk Analysis and Assessment. Risk analysis and assessment is the process of measuring the risks of the identified project processes and project products. The main components include determining the probability of risk occurrence and the severity of impact on project objectives, evaluating the potential impact of all risks, and thus obtaining the values of risk decision variables for the project, which are used as an important basis for project decision making. Each risk can be measured in terms of its probability of occurrence and potential loss value and can also be analyzed with the help of a risk level matrix, as shown in Figure 2, where different grids represent different amounts of risk with R_1 representing low risk, R_2 representing medium risk, and R_3 representing high risk. The horizontal axis indicates the probability of risk occurrence, and the vertical axis indicates the degree of risk loss, both with five levels. The expected value of risk is equal to the probability of risk occurrence multiplied by the risk loss level. The risk classification table for cold storage construction quality is shown in Table 5.

Statistically, 72 questionnaires were returned, of which 58 were valid. The data in the questionnaires were processed and the quality risk factors with R_3 ratio of 15% and above were selected as key indicators. The statistical results are shown in Table 6.

5. Experiments and Results

5.1. Bayesian Network Model. In this paper, Netica is used for simulation experiments. Netica is currently the most widely used Bayesian network analysis software in the world,

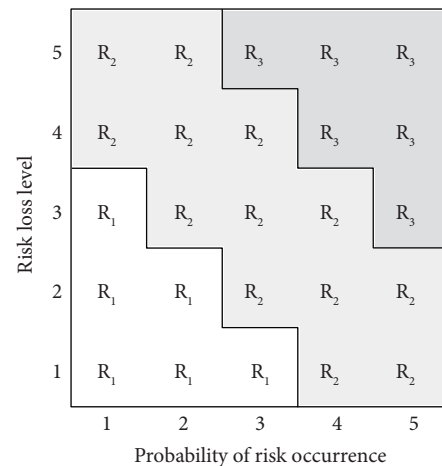


FIGURE 2: Risk level matrix.

featuring simplicity, reliability, and efficiency. Netica Application is a comprehensive tool for working with Bayesian belief nets and decision nets (influence diagrams). It can build, learn, modify, transform and store nets, as well as answer queries or find optimal solutions using its powerful inference engine. As a decision-making tool, it is widely used in business, engineering, medical and ecological analysis.

By consulting relevant materials and asking experts and scholars for their opinions, the following quality risk causality chains are drawn:

- ① Improper coordination between the construction unit and the design, building, and supervision parties (E1) → Insufficient financing (A3) → Unqualified quality of building materials and equipment (C15) → Cold storage technology is not applicable (C12)
- ② Improper coordination between the construction unit and the design, building, and supervision parties (E1) → Frequent design changes (B3) →

TABLE 5: Classification of quality risk levels of cold storage construction.

Risk level	Nature of risk	Measures
R_1	Low risk	Monitoring is required and no action is required
R_2	Medium risk	Efforts are made to reduce risks and monitoring is required
R_3	High risk	Monitoring is required and work will not start until the risk has been reduced

TABLE 6: Key quality risk factors of cold storage construction.

Construction procedures	Participating units	Work processes	Statistical result of risk level (%)		
			R_1	R_2	R_3
A-decision-making stage	Construction units	A2-the preparation of the feasibility study report is unreasonable	10	75	15
		A3-insufficient financing	26	45	29
B-design work stage	design units	B1-construction drawings are not designed to meet construction and billing requirements and do not take into account construction possibilities	36	26	38
		B3-frequent design changes	37	35	28
		C1-inadequate construction preparation	42	27	31
		C5-duration does not meet the requirements	6	61	33
C-engineering construction stage	Building units	C8-technical delivery is not in place	11	73	16
		C10-backwardness of the construction process	32	42	26
		C11-construction technology is difficult	24	32	44
		C12-cold storage technology is not applicable	31	42	27
		C13-inadequate training of skilled workers	10	71	19
		C14-impact of force majeure	23	31	46
D-completion acceptance and delivery stage	Construction units, design units, building units, supervision units	C15-unqualified quality of building materials and equipment	5	74	21
		C16-mismatch of building materials and equipment	23	52	25
E-full stage	Construction units, design units, building units, supervision units	D1-completion acceptance is not careful or acceptance standards are unreasonable	21	62	17
		E1-improper coordination between the construction unit and the design, building and supervision parties	16	57	27
		E2-inadequate quality assurance system	16	54	30
	Supervision units	E4-unfavorable project environment	33	36	31
		E5-supervision is not effective	47	38	15

Construction technology is difficult (C11) → Duration does not meet the requirements (C5)

- ③ Inadequate quality assurance system (E2) → Technical delivery is not in place (C8) → Inadequate training of skilled workers (C13) → Completion acceptance is not careful or acceptance standards are unreasonable (D1)
- ④ Inadequate quality assurance system (E2) → Mismatch of building materials and equipment (C16) → Cold storage technology is not applicable (C12)
- ⑤ Inadequate quality assurance system (E2) → Supervision is not effective (E5) → Completion acceptance is not careful or acceptance standards are unreasonable (D1)
- ⑥ Construction drawings are not designed to meet construction and billing requirements and do not

take into account construction possibilities (B1) → The preparation of the feasibility study report is unreasonable (A2) → Backwardness of the construction process (C10)

- ⑦ Construction drawings are not designed to meet construction and billing requirements and do not take into account construction possibilities (B1) → Impact of force majeure (C14) → Construction technology is difficult (C11) → Duration does not meet the requirements (C5)
- ⑧ Inadequate construction preparation (C1) → Unfavorable project environment (E4) → Construction technology is difficult (C11) → Duration does not meet the requirements (C5)

According to the above causal relationship chain, a Bayesian network topology is constructed in Netica software, the questionnaire data is made into a table, and the

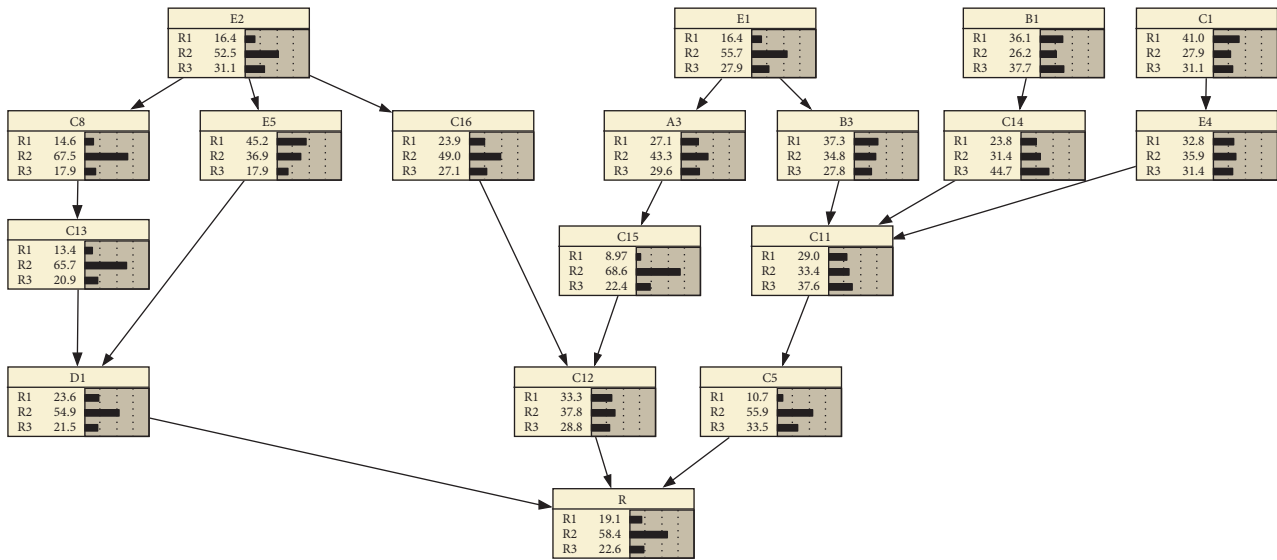


FIGURE 3: Bayesian network of cold storage construction quality risks.

TABLE 7: Quality risk levels of cold storage construction.

Risk level	Probability (%)
R_1	19.1
R_2	58.4
R_3	22.6

table data is imported for parameter learning, and finally a complete Bayesian network model is presented, as shown in Figure 3.

Finally, through the established Bayesian network of quality risk of cold storage construction, it can be concluded that the overall quality risk level of cold storage construction is shown in Table 7 below.

5.2. Bayesian Network Diagnostic Inference Analysis

5.2.1. Reverse Reasoning. Reverse reasoning refers to assuming that the probability of occurrence of a child node is 100%, and reversely infers the probability of occurrence of the corresponding parent node. Assuming that the occurrence level of the total quality risk of cold storage construction is R_1 ; that is, the probability of occurrence of R_1 is 100%, the following Bayesian network can be obtained as shown in Figure 4.

The comparative analysis shows that, compared with the original Bayesian network, when the overall quality risk level of the cold storage construction is assumed to be R_1 , the risk occurrence probability of $D1$ and $C5$ has changed significantly, the probability of R_2 is reduced, and the probability of R_1 and R_3 increases. This shows that when the overall quality risk level of cold storage construction is R_1 , the most affected factors are duration does not meet the requirements and completion acceptance is not careful or acceptance standards are unreasonable.

Assuming that the occurrence level of the total quality risk of cold storage construction is R_2 ; that is, the probability

of occurrence of R_2 is 100%, the following Bayesian network can be obtained as shown in Figure 5.

The comparative analysis shows that compared with the original Bayesian network, the probability of risk occurrence of $D1$ and $C5$ appears to have more obvious changes when assuming that the level of the overall quality risk of cold storage construction is R_2 , showing an increase in the probability of R_2 and a decrease in the probability of occurrence of R_1 and R_3 .

Assuming that the occurrence level of the total quality risk of cold storage construction is R_3 ; that is, the probability of occurrence of R_3 is 100%, the following Bayesian network can be obtained as shown in Figure 6.

The comparative analysis shows that compared with the original Bayesian network, the probability of risk occurrence of $D1$ and $C5$ appears to have more obvious changes when assuming that the level of the overall quality risk of cold storage construction is R_3 , showing a decrease in the probability of R_2 and an increase in the probability of occurrence of R_1 and R_3 .

In summary, when different levels of quality risks occur in cold storage construction, the corresponding risk levels of other nodes will increase, with the probability of $C5$ (Duration does not meet the requirements) and $D1$ (Completion acceptance is not careful or acceptance standards are unreasonable) changing more significantly, indicating that these two factors have a greater impact on the overall quality risk of cold storage construction.

5.2.2. Sensitivity Analysis. Sensitivity analysis refers to finding out the key nodes that are greatly affected by other nodes by observing the degree of influence of each node by other nodes and using the correlation effect between nodes. Select the “Sensitivity to Findings” operation in the Netica software, and then check the influence of other discovery nodes in the network on selecting different target nodes. The sensitive factors that are most affected are listed below. They

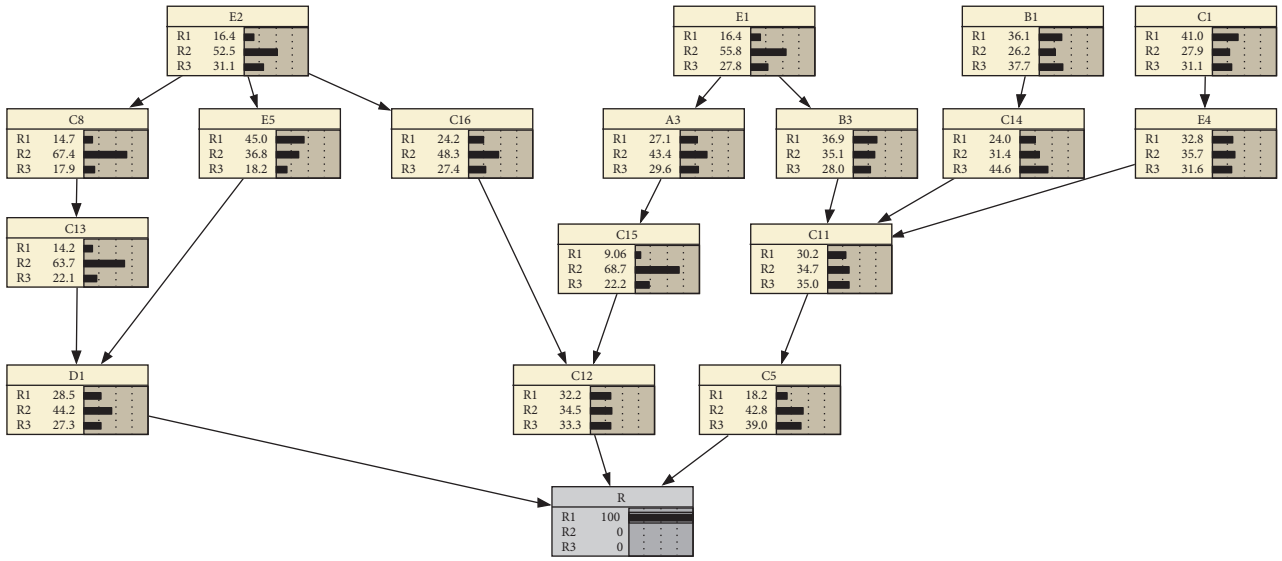


FIGURE 4: Bayesian network reverse reasoning of cold storage construction quality risks (R_1).

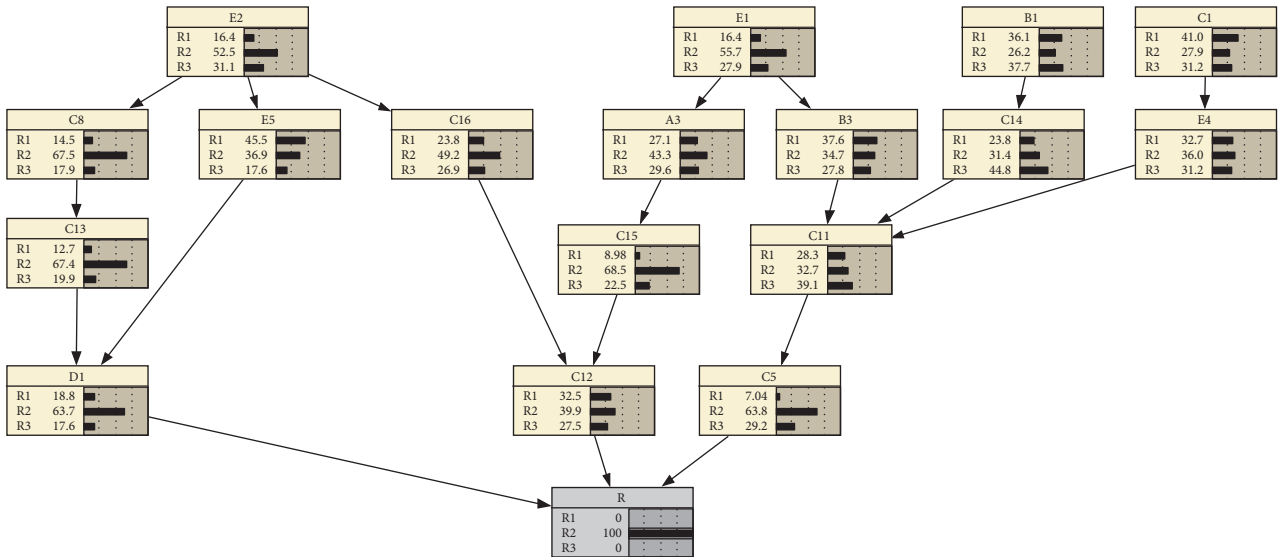


FIGURE 5: Bayesian network reverse reasoning of cold storage construction quality risks (R_2).

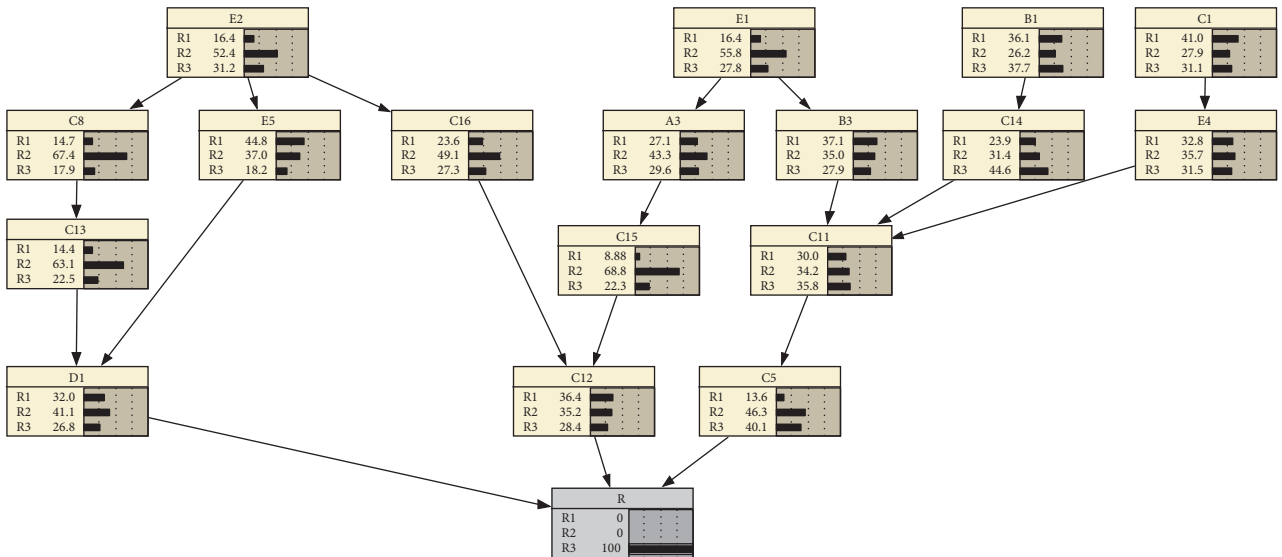


FIGURE 6: Bayesian network reverse reasoning of cold storage construction quality risks (R_3).

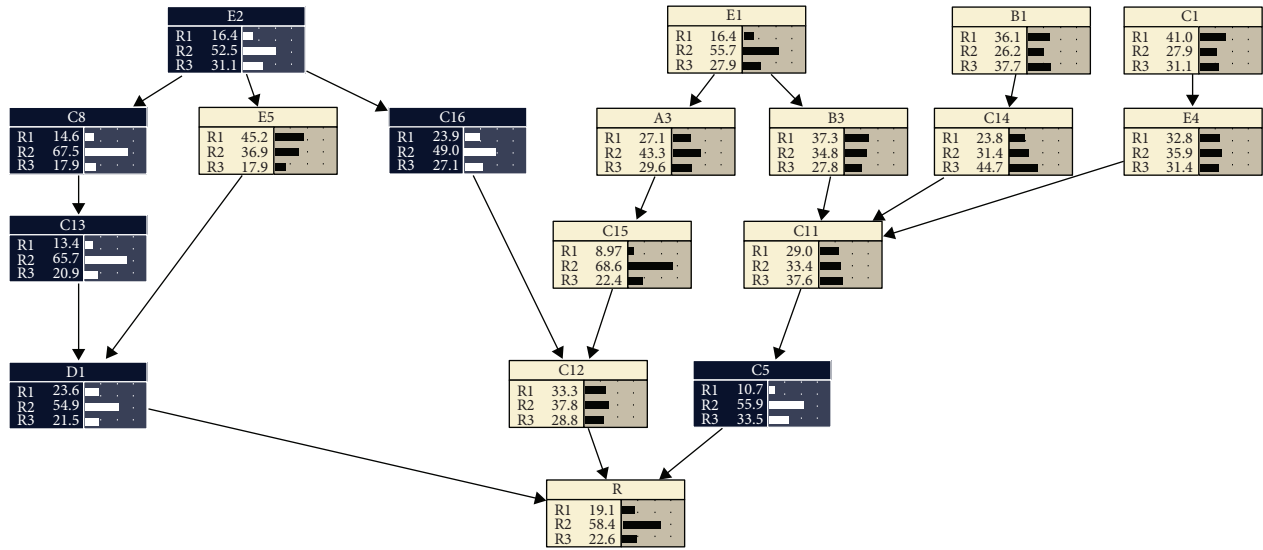


FIGURE 7: Bayesian network sensitivity analysis of cold storage construction quality risks.

TABLE 8: Sensitivity of E2 to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
E2	1.44008	100	0.3808160
E5	0.11391	7.91	0.0248665
C16	0.06810	4.73	0.0064580
C8	0.03475	2.41	0.0038769

TABLE 9: Sensitivity of C8 to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
C8	1.23280	100	0.2986122
C13	0.03734	3.03	0.0068553
E2	0.03475	2.82	0.0052657

are E2 (Inadequate quality assurance system), C8 (Technical delivery is not in place), C16 (Mismatch of building materials and equipment), C13 (Inadequate training of skilled workers), D1 (Completion acceptance is not careful or acceptance standards are unreasonable), and C5 (Duration does not meet the requirements). The specific nodes are shown in Figure 7.

The sensitivity analysis of E2 is shown in Table 8. The main influencing factors of E2 (Inadequate quality assurance system) are E5 (Supervision is not effective), C16 (Mismatch of building materials and equipment), and C8 (Technical delivery is not in place).

The sensitivity analysis of C8 is shown in Table 9. It can be seen that C8 (Technical delivery is not in place) is mainly affected by C13 (Inadequate training of skilled workers) and E2 (Inadequate quality assurance system).

The sensitivity analysis of C16 is shown in Table 10. C16 (Mismatch of building materials and equipment) is mainly affected by E2 (Inadequate quality assurance system) and C12 (Cold storage technology is not applicable).

The sensitivity analysis of C13 is shown in Table 11. It can be seen that C13 (Inadequate training of skilled workers) is mainly affected by C8 (Technical delivery is not in place) and D1 (Completion acceptance is not careful or acceptance standards are unreasonable).

The sensitivity analysis of D1 is shown in Table 12. It can be seen that D1 (Completion acceptance is not careful or acceptance standards are unreasonable) is mainly affected by C13 (Inadequate training of skilled workers).

TABLE 10: Sensitivity of C16 to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
C16	1.50800	100	0.4097652
E2	0.06810	4.52	0.0050632
C12	0.02476	1.64	0.0035677

TABLE 11: Sensitivity of C13 to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
C13	1.25863	100	0.3085228
C8	0.03734	2.97	0.0021347
D1	0.02978	2.37	0.0067237

The sensitivity analysis of C5 is shown in Table 13. It can be seen that C5 (Duration does not meet the requirements) is mainly affected by C11 (Construction technology is difficult).

The sensitivity analysis of R is shown in Table 14. The main influencing factors of R (Cold storage construction quality risk) are D1 (Completion acceptance is not careful or acceptance standards are unreasonable) and C5 (Duration does not meet the requirements).

A quality assurance system is a key part of quality management and risk management, which is sensitive to the construction of cold storage; technical delivery is the preparation work before construction so that the construction personnel understands all aspects of the project in detail so as to facilitate the scientific organization of construction and avoid accidents, so this link is also a sensitive

TABLE 12: Sensitivity of $D1$ to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
$D1$	1.44373	100	0.3820662
R	0.03218	2.23	0.0069764
$C13$	0.02978	2.06	0.0062238

TABLE 13: Sensitivity of $C5$ to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
$C5$	1.34205	100	0.3420480
R	0.03069	2.29	0.0059037
$C11$	0.02718	2.03	0.0068624

TABLE 14: Sensitivity of R to a finding at another node.

Node	Mutual info	Percent	Variance of beliefs
R	1.39406	100	0.3614260
$D1$	0.03218	2.31	0.0072761
$C5$	0.03069	2.2	0.0064688

factor; construction materials and equipment are indispensable in the construction process, and its quality is crucial to the successful completion of the project; skilled workers are the main implementers of the construction, and the quality of the operators will directly affect the construction of the cold storage; completion acceptance is the assessment and inspection of the project, and is the last line of defense to regulate the quality of the building. Duration is one of the important accounting indicators of construction enterprises, and the length of duration directly affects the economic benefits of construction enterprises. Reasonable arrangement of duration and organization of flow work can improve labor productivity and reduce engineering costs. In summary, the nodes identified by Bayesian network sensitivity analysis are all sensitive factors in the cold storage construction process, which need to be monitored by key prevention.

5.3. Quality Risk Management Measures. After the Bayesian network reverse reasoning analysis and sensitivity analysis, this article provides relevant quality risk management measures based on the above analysis results and add from after explains five aspects of man, material, machine, method, and environment.

5.3.1. Measures on the Man Aspect. Man is the main force of implementation. To ensure the smooth progress of construction projects, standardized management measures need to be taken for technical workers. First of all, focus on construction related training to ensure that workers are proficient in relevant basic operation steps, and organize technical workers to conduct technical intercourse, so that they can learn more about the entire project in order to effectively organize scientific construction and improve work efficiency. At the same time, regularly organize

technical assessment, screen qualified workers, and give workers a sense of urgency, pressure, and responsibility.

As a party on behalf of the construction unit, the supervision unit shall strengthen the professionalism of the supervision engineer and organize relevant training on a regular basis. It is necessary to ensure that the supervision engineer must be serious, professional, and responsible in the process of engineering construction to ensure the smooth completion of the cold storage.

5.3.2. Measures on the Material Aspect. In order to grasp the quality off and quantity off, materials and equipment need to be inspected and accepted according to the incoming material plan, acceptance specifications, etc., when they come in, and records should be made. When the materials are put into the warehouse, a ledger should be established, reasonably placed, and regularly inventoried to ensure that the accounts match. When receiving and issuing materials, it is necessary to receive and issue materials within the limit, to indicate the reason for using materials beyond the limit, and to establish a ledger to record the receiving and issuing situation. The person responsible for on-site material management should supervise the use of on-site materials to ensure that the materials are used reasonably according to the material plan. When the surplus material is recycled, the material will be returned in time and the subsequent use process will be reasonably arranged.

In view of the particularity of the cold storage, in order to realize the normal operation of the thermal insulation function of the cold storage, it is necessary to purchase qualified thermal insulation materials, to achieve good thermal insulation and moisture resistance, and to adapt to local conditions and make full use of them.

5.3.3. Measures on the Machine Aspect. The selection of construction machinery should follow the principles of relevance to needs, practical possibility, and economic reasonableness. When choosing construction machinery, first choose the leading project machinery, according to the characteristics of the project, to choose the most suitable type. At the same time, in order to give full play to the efficiency of leading machinery, the corresponding auxiliary facilities should be selected accordingly to coordinate their production capacity.

Construction machinery and equipment can only be admitted after acceptance, and the maintenance responsibility system is implemented during use, with dedicated maintenance and repair to maintain the good technical condition of the equipment, improve the reliability and safety of equipment operation, reduce losses, and improve economic efficiency. The operators of the equipment must be trained uniformly and can only take up their posts after passing the test. Regularly organize skill examinations to make technical personnel proficient in equipment operation skills and ensure that the equipment is used reasonably. Compile machinery and equipment use plans scientifically, and consider when using a certain construction process, the

use of what and how to use machinery and equipment is the most reasonable and efficient.

5.3.4. Measures on the Method Aspect. Whether the implementation plan is reasonable, the process operation is correct, and the technology is advanced, all of these can have a big impact on the project. The selection of construction machinery and construction methods should be unified and coordinated. When choosing the construction method, it should focus on the construction method of the component works that affect the construction of the whole project.

Cold storage should have a reasonable structure and good heat insulation to ensure the quality of food storage. Cold storage mainly consists of an enclosure structure and load-bearing structure, of which the enclosure structure should have good heat insulation and moisture-proof effect to resist the storm outside the storage, and the load-bearing structure should be able to support the weight of goods and loading and unloading equipment. As a productive power-using unit, energy-saving measures should be taken to reduce energy consumption in cold storage. The heat transfer of the enclosure structure accounts for 20%–35% of the total heat load of the cold storage, so the construction process needs to ensure the good performance of the cold storage enclosure structure.

Use project management software appropriately for time scheduling, resource management, cost management, project monitoring, and information sharing, etc. Information technology can propose a variety of options for decision makers in response to project characteristics, making project management work fast and thus saving a lot of human resources.

Improve the quality assurance system. Decompose the quality objectives into specific tasks and implement them in each department with clear responsibilities. Establish a quality management system and document it for implementation, maintenance, and continuous improvement. According to the characteristics of different projects, on the basis of the original quality management system, more targeted quality plans and quality procedures are formulated. During the operation of the system, regular supervision and assessment are carried out, and the effectiveness and suitability of the system are promoted through internal and external review.

5.3.5. Measures on the Environment Aspect. Reasonably arrange the progress of the project to avoid the influence of bad weather. Investigate the geological and hydrological conditions required by the cold storage project in advance. Improve the working environment of project operators, provide good office conditions, office facilities, communication conditions, and security to improve labor productivity. Implement the environmental protection target responsibility system and assign environmental protection responsibility to departments or personnel. Establish and operate an environmental management system efficiently, mobilize the enthusiasm of relevant on-site organizations, and implement environmental inspection and monitoring.

Formulate emergency measures to maintain a good working environment, sanitary conditions, and public order on-site, and make continuous improvements.

According to the principle of purpose, set up affairs according to the goals, set up institutions and staff according to the situation, set up positions according to the staff, set up systems, and delegate powers according to responsibilities; according to the principle of lean and efficient, try to simplify the organization as much as possible, so that people can make the best use of their talents. Through the establishment of a highly efficient organization that operates freely, a responsibility system and an information communication system are formed so as to achieve a reasonable division of labor and collaboration. The production activities of construction projects are staged, open-air and fluid, and management work and organizational structures need to be adjusted accordingly to adapt to changes in construction tasks.

6. Conclusion

This paper starts with the research on the quality risk management of cold storage construction projects from three aspects: construction procedures, participating units and work processes. First, analyze and identify the quality risk factors existing in the four stages of the decision-making stage, design work stage, engineering construction stage, completion acceptance and delivery stage, analyze the whole process by integrating the four stages, and correspond to the relevant responsible units, then carry out relevant research according to the risk management process, and build a Bayesian network model of cold storage construction quality risk. Through reverse reasoning analysis and sensitivity analysis, the key quality risk factors are obtained: inadequate quality assurance system, technical delivery is not in place, mismatch of building materials and equipment, inadequate training of skilled workers, completion acceptance is not careful, or acceptance standards are unreasonable, and duration does not meet the requirements. Finally, in view of the above-mentioned quality risks, suggestions and measures are put forward from five aspects: man, material, machine, method, and environment. The quality risk management system of cold storage construction in this paper makes full use of the advantages of the Bayesian network in quality risk assessment and realizes the application of the Bayesian network for the complex engineering project of cold storage construction so as to conduct related research more effectively.

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.

References

- [1] A.-A. Baksh, R. Abbassi, V. Garaniya, and F. Khan, "Marine transportation risk assessment using Bayesian Network: application to Arctic waters," *Ocean Engineering*, vol. 159, pp. 422–436, 2018.
- [2] N. E. Fenton, S. McLachlan, P. Lucas et al., *A Bayesian Network Model for Personalised COVID19 Risk Assessment and Contact Tracing*, MedRxiv, 2021.
- [3] C. Liu, Y. Wang, X. Li, Y. Li, F. Khan, and B. Cai, "Quantitative assessment of leakage orifices within gas pipelines using a Bayesian network," *Reliability Engineering & System Safety*, vol. 209, Article ID 107438, 2021.
- [4] Q. Xiao, F. Luo, and Y. Li, "Risk assessment of seaplane operation safety using Bayesian network," *Symmetry*, vol. 12, no. 6, p. 888, 2020.
- [5] G. Zhang, V. V. Thai, A. W. K. Law, K. F. Yuen, H. S. Loh, and Q. Zhou, "Quantitative risk assessment of seafarers' nonfatal injuries due to occupational accidents based on bayesian network modeling," *Risk Analysis*, vol. 40, no. 1, pp. 8–23, 2020.
- [6] A. Chan, F. Wong, C. Hon, and T. Choi, "A Bayesian network model for reducing accident rates of electrical and mechanical (E&M) work," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, p. 2496, 2018.
- [7] Y. Zhou, C. Li, C. Zhou, and H. Luo, "Using Bayesian network for safety risk analysis of diaphragm wall deflection based on field data," *Reliability Engineering & System Safety*, vol. 180, pp. 152–167, 2018.
- [8] S. Cao, K. Bryceson, and D. Hine, "An Ontology-based Bayesian network modelling for supply chain risk propagation," *Industrial Management & Data Systems*, vol. 119, no. 8, pp. 1691–1711, 2019.
- [9] X. Zhang, X. Hu, Y. Bai, and J. Wu, "Risk assessment of gas leakage from school laboratories based on the Bayesian network," *International Journal of Environmental Research and Public Health*, vol. 17, no. 2, p. 426, 2020.
- [10] J. Wu, Z. Hu, J. Chen, and Z. Li, "Risk assessment of underground subway stations to fire disasters using Bayesian network," *Sustainability*, vol. 10, no. 10, p. 3810, 2018.
- [11] J. Zhang, H. Bian, H. Zhao, X. Wang, L. Zhang, and Y. Bai, "Bayesian network-based risk assessment of single-phase grounding accidents of power transmission lines," *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, p. 1841, 2020.
- [12] H. Chen, Q. Zhang, J. Luo, X. Zhang, and G. Chen, "Interruption risk assessment and transmission of fresh cold chain network based on a fuzzy bayesian network," *Discrete Dynamics in Nature and Society*, vol. 2021, pp. 1–11, Article ID 9922569, 2021.
- [13] H. Zhang, H. Feng, Y. Cui, and Y. Wang, "A fuzzy Bayesian network model for quality control in O2O e-commerce," *International Journal of Computers, Communications & Control*, vol. 15, no. 1, 2020.
- [14] M. Li, H. Wang, D. Wang, Z. Shao, and S. He, "Risk assessment of gas explosion in coal mines based on fuzzy AHP and bayesian network," *Process Safety and Environmental Protection*, vol. 135, pp. 207–218, 2020.
- [15] M. Li, D. Wang, and H. Shan, "Risk assessment of mine ignition sources using fuzzy Bayesian network," *Process Safety and Environmental Protection*, vol. 125, pp. 297–306, 2019.
- [16] Y. Jianxing, W. Shibo, Y. Yang et al., "Process system failure evaluation method based on a Noisy-OR gate intuitionistic fuzzy Bayesian network in an uncertain environment," *Process Safety and Environmental Protection*, vol. 150, pp. 281–297, 2021.
- [17] L. Han, J. Zhang, Y. Zhang, Q. Ma, S. Alu, and Q. Lang, "Hazard assessment of earthquake disaster chains based on a Bayesian network model and ArcGIS," *ISPRS International Journal of Geo-Information*, vol. 8, no. 5, p. 210, 2019.
- [18] X. Li, Y. Zhang, F. Guo, X. Gao, and Y. Wang, "Predicting the effect of land use and climate change on stream macro-invertebrates based on the linkage between structural equation modeling and bayesian network," *Ecological Indicators*, vol. 85, pp. 820–831, 2018.
- [19] W. Chen and S. Huang, "Evaluating flight crew performance by a bayesian network model," *Entropy*, vol. 20, no. 3, p. 178, 2018.
- [20] E. Mendes, M. Perkusich, V. Freitas, and J. Nunes, "Using bayesian network to estimate the value of decisions within the context of value-based software engineering," in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*, pp. 90–100, Christchurch New Zealand, June 2018.
- [21] W. Chen, "Research on the construction of northwest cold storage based on grey relational degree," *Academic Journal of Engineering and Technology Science*, vol. 2, no. 3, 2019.
- [22] J. P. Kuźmicki, "Project analysis of a cold room for fishery products storage," *Doctoral dissertation, Instytut Techniki Ciepłej*, Warsaw, Poland, 2021.
- [23] O. A. Chukwu and M. Adibe, "Quality assessment of cold chain storage facilities for regulatory and quality management compliance in a developing country context," *The International Journal of Health Planning and Management*, vol. 37, no. 2, pp. 930–943, 2022.
- [24] H. Miao and X. Zhang, "Research on energy consumption evaluation method and energy saving operation technology of cold storage," in *Journal of Physics: Conference Series* vol. 2205, no. 1, IOP Publishing, Article ID 012003, 2022.

Research Article

Research on Embedded Multifunctional Data Mining Technology Based on Granular Computing

Juan Li ^{1,2} and Xianghong Tian^{1,2}

¹School of Computer Engineering, Jinling Institute of Technology, Nanjing, Jiangsu 211169, China

²Jiangsu Provincial Key Laboratory of Data Science and Intelligent Software, Nanjing, Jiangsu 211169, China

Correspondence should be addressed to Juan Li; iamlj6@jit.edu.cn

Received 21 March 2022; Revised 22 April 2022; Accepted 25 April 2022; Published 20 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Juan Li and Xianghong Tian. 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 influence and limitations of the multisourced, heterogeneous, and unbalanced characteristics of embedded multifunctional data, the application effect of the current data mining technology is not good, and the accuracy is low. To solve the above problems, an embedded multifunctional data mining technology based on granular computing was studied. According to the three characteristics of embedded multifunctional data, preprocessing such as data reduction, data standardization, and data balance were implemented. We implemented data granulation for the preprocessed data and calculated the data granulation characteristics, including offset, particle density, and intraparticle interval. Taking granular features as the input content, embedded multifunctional data mining was realized by using a neural network to complete the objectives of data classification, anomaly detection, fault identification, and so on. The experimental results showed that the anomaly mining results of each type of data mining were greater than 0.9, indicating that the accuracy of the mining technology is high.

1. Introduction

Data mining refers to mining hidden rules or features from massive data for decision analysis. For example, in fault identification, data mining can judge the collected data and perform user classification to help develop sales strategies [1]. Therefore, data mining is an important aspect of big data processing and is the focus of research in the current information networks. To date, although big data research has achieved great success and significant achievements have been made in big data mining [2], there are still some problems that need to be deeply studied, among which embedded multifunctional data mining is a difficult point. Embedded multifunctional data are a kind of data with multiple description functions stored in a number of heterogeneous sources. The typical characteristics of such data are that it is multisourced, heterogeneous, and imbalanced. The existence of these three features makes data mining face great difficulties, and its accuracy and efficiency are greatly limited. Due to the influence and limitations of the

multisourced, heterogeneous, and unbalanced characteristics of embedded multifunctional data, the current data mining technology has low accuracy and poor application effects in practical applications. Based on the above background, how to improve the accuracy and efficiency of embedded multifunctional data mining has become the focus of the current research.

At present, there are many studies on data mining techniques. For example, Wang Zhanping et al. [3] applied data mining technology to container shipping price prediction. In their research, based on the collected historical container shipping price time-series data, the container shipping price prediction model was established based on the GBDT algorithm to realize the data mining. Zhang Lili et al. [4] applied data mining technology to aviation customer classification. In the research, based on the passenger flight records of airlines, customer loss was predicted using the decision tree method, and the K-means clustering algorithm was used to classify customer categories and explore customer value, which provides a reliable basis for

formulating effective marketing strategies and improving the economic benefits of airlines. Ma Lili et al. [5] applied a data mining technology map to abnormal data detection in an optical fiber communication network. In their research, the operation data of the optical fiber communication network were first collected, and the data characteristics were extracted. Finally, the optimal value of the entropy target function was calculated by the sample attribute probability, and the optimal value was used to complete the anomalous data detection. Wang et al. [6] proposed data mining technology for Internet industry collaborative innovation platform research. Information technology has therefore been integrated into every corner of production and life. Considering the computing cost, the Internet of things, and intelligent service collaborative innovation as the research object, we studied the combination based on data mining technology of Internet of things and intelligent service collaborative innovation. For the development of intelligent service industry and the improvement of the Internet of things collaborative innovation, we provide valuable theoretical basis. Based on previous research experience, an embedded multifunctional data mining technology based on particle calculation was studied. According to the three characteristics of embedded multifunctional data, data reduction, data standardization, and data balance, the data were preprocessed. The processed data were granulated and analyzed for particle characteristics, including offset, particle density, and intragrain spacer. Furthermore, a particle feature-based neural network was used to classify the data, detect anomalies, and identify faults. Through the data mining of the proposed data mining technology and comparing the methods presented in the literature, the accuracy of the various data mining methods was above 0.9. The innovation point of the studied technology is the data granulation and data particle feature calculation. The embedded multifunctional data were preprocessed and then the embedded multifunctional data feature extraction was performed. The above extracted features were used as input, and the embedded multifunctional data mining was implemented to achieve data classification, anomaly detection, fault identification, and other goals.

2. Embedded Multifunctional Data Mining Technology Based on Particle Computing

2.1. General Framework. Embedded multifunctional data are multisourced, heterogeneous, and imbalanced, so mining embedded multifunctional data with widely used data mining technology cannot achieve good results. Facing this situation, it is of great practical significance to study a new data mining technology to deal with embedded multifunctional data. The key to the data mining technology studied here is particle computing, which refers to the division of massive data or information according to certain rules or relationships, thus forming particles. Based on this theory, a data mining technique can be designed for the effective classification of embedded multifunctional data. The general framework of the embedded multifunctional data mining technology is shown in Figure 1.

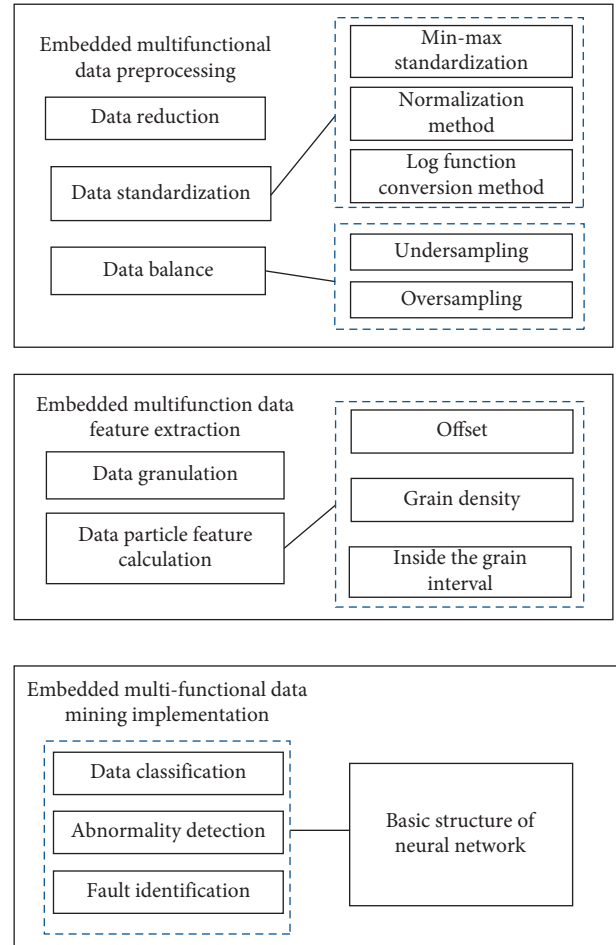


FIGURE 1: Overall framework of the embedded multifunctional data mining technology.

According to the content shown in Figure 1, embedded multifunctional data mining is a repeated process. If each link fails to achieve the expected results, it must return to the previous step for another adjustment and implementation. Comparing previous models, not all data mining efforts are required to be listed here; for example, data integration can be ignored when there are no multiple data sources in a job. For some multifunctional data, it is a very necessary process to conduct embedded multifunctional data preprocessing, embedded multifunctional data feature extraction based on particle calculation, and embedded multifunctional data mining.

2.2. Embedded Multifunctional Data Preprocessing. To realize the effective mining of the embedded multifunctional data, the embedded multifunctional data preprocessing is required first. Preprocessing can effectively reduce the multisourced, heterogeneous, and imbalanced data, improve the data quality, and facilitate mining [7]. Embedded multifunctional data preprocessing includes data reduction, data standardization, and data balancing. Specific analysis was performed for these three preprocessing steps.

2.2.1. Data Reduction. Embedded multifunctional data come from multiple different databases, and after pooling the data from multiple databases together, the embedded multifunctional data are formed. The embedded multifunctional data are therefore massive, and such data can be collectively referred to as redundant data [8]. The presence of redundant data will increase the computation and interfere with the data mining results, thus requiring data reduction, as shown in Figure 2.

2.2.2. Data Standardization. Embedded multifunctional data come from multiple different databases, and there is also some heterogeneity, which represents different data dimensions, leading to no synchronous processing between the data [9]. To this end, standardization of embedded multifunctional data is required. The methods for handling this are as follows:

- (1) Min-max standardization:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (1)$$

x represents the original embedded multifunctional data, x' represents the standardized embedded multifunctional data, and $\min(x)$ and $\max(x)$ represents the minimum and maximum values in the original embedded multifunctional data.

- (2) Normalization method:

$$x' = \frac{x - a}{b}, \quad (2)$$

where a and b represent the mean and standard deviation of the raw embedded multifunctional data.

- (3) Log function conversion method:

$$\tilde{x} = \frac{\log_{10}(x)}{\log_{10} \max(x)}. \quad (3)$$

The dimension of embedded multifunctional data is standardized to be unified [10].

2.2.3. Data Balancing. Imbalance is one of the major features of embedded multifunctional data, and the mining of unbalanced data will lead to mining accuracy distortion [11]. For this point, the unbalanced data need to be balanced with the data. Select the undersampling method or oversampling method based on the number of negative and positive samples in the data. The undersampling method is suitable for more negative samples and the oversampling method for more positive samples [12].

Undersampling. The undersampling principle refers to the removal of most redundant negative samples to balance with the positive samples [13]. The specific process is as follows:

- Step 1 : enter most class samples, that is, negative samples.

Step 2 : cluster the negative samples and divide the samples into subsamples of multiple categories.

Step 3 : calculate the similarity redundancy coefficient between each subsample with the following formula:

$$S_K = \sqrt{d_i^n \cdot D_{ij}}, \quad (i, j = 1, 2, \dots, n). \quad (4)$$

S_K represents the similarity redundancy coefficient, d_i^n represents the distance from the subsample i to its cluster center, and D_{ij} represents the Euclidean distance between the subsamples i, j .

Step 4 : make the calculated similarity redundancy coefficient into a matrix form.

Step 5 : delete one of the two subsamples of the minimum similarity redundancy coefficient in the matrix and the corresponding rows and columns in the matrix.

Step 6 : determine whether the sample deletion requirements are met. If achieved, remove most redundant negative samples and complete the undersampling; otherwise, return to the previous Step 5 until the end requirements are met.

Oversampling. The oversampling principle is to select negative samples and then calculate the distance between each Euclidean sample and all the Euclidean distances to determine the k nearest neighbors. Finally, the k nearest neighbors are selected according to the set sampling fold rate to generate new samples to compensate for the small number of negative samples and to achieve sample balance [14]. The principle formula is as follows:

$$Y_{\text{new}} = y_{k\text{近邻}} + \text{rand}(0, 1) \cdot (x^- - y_{k\text{近邻}}), \quad (5)$$

where Y_{new} represents a new sample formed after sampling, $y_{k\text{近邻}}$ represents k nearest neighbors, $\text{rand}(0, 1)$ represents a random number between $(0, 1)$, and x^- represents the original negative sample. The embedded multifunctional data preprocessing is completed to pave the way for the extraction of the embedded multifunctional data features based on particle calculation.

2.3. Embedded Multifunctional Data Feature Extraction Based on Particle Calculation. After finishing the embedded multifunctional data, then the embedded multifunctional data feature is extracted. The specific process includes two steps, namely, data granulation and data particle feature calculation [15]. Specific analysis of these two processes is described as follows.

2.3.1. Data Granulation. Data granulation refers to dividing embedded multifunctional data into one data block according to certain rules and relationships. A block of data is called a grain [16]. Through the granulation

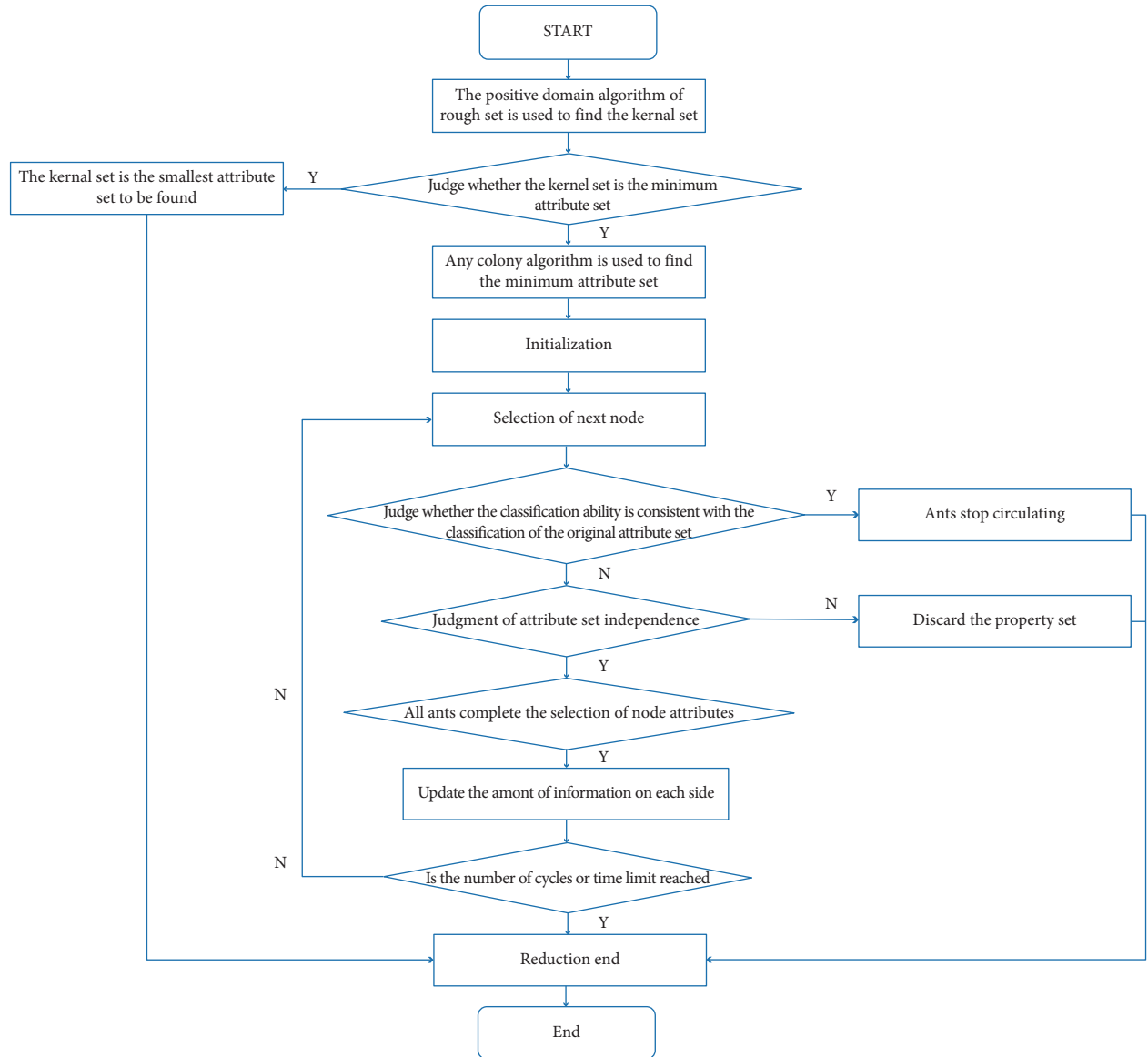


FIGURE 2: Data reduction process.

processing, it is easier to find the rules or characteristics between the data. The data granulation process is described as follows:

- Step 1 : enter the embedded multifunctional dataset, noted as $X = \{x_1, x_2, \dots, x_n\}$.
- Step 2 : select K data from $X = \{x_1, x_2, \dots, x_n\}$, as the initial category representative, which are recorded as $U^h = \{z_1^h, z_2^h, \dots, z_K^h\}$. Because it is the initial sample, so set $h = 0$.
- Step 3 : calculate the distance between all samples except the initial category sample and the initial category sample.
- Step 4 : according to the proximity principle, divide all the remaining samples into an initial sample category, and get a new cluster, recorded as P_j^{h+1} , $j = 1, 2, \dots, K$.

Step 5 : reslect the category representative from step 4 results, noted as p_j^{h+1} .

Step 6 : determine whether p_j^{h+1} is equal to U^h . If equal, end the operation and complete the data granulation; otherwise, set $h = h + 1$ and return to step 3, and repeat the above steps until the above conditions are fulfilled and the data granulation is completed.

Step 7 : output the granulation results.

2.3.2. Data Particle Feature Calculation. Based on the above divided data particles, the data particle characteristics, including the offset degree, particle density, and interparticle space, are calculated [17]. Calculate these three features.

Offset. Offset degree refers to the case of the data particle offset particle center, with the following formula:

$$G_{i,O} = \frac{\sum_{i=1}^n f[g(i,O), q(O)]}{n}, \quad (6)$$

where $g(i,O)$ represents the accessible distance from the particle i to the particle center, n represents the particle number, $q(O)$ represents the particle center capacity, f represents the binary mapping function, and $G_{i,O}$ represents the particle i offset degree.

Grain Density. Particle density refers to the density of the particle distribution. The calculation formula is as follows:

$$\rho = \frac{Bn}{\sum_{i=1}^n B_i}, \quad (7)$$

where ρ represents the particle density, B_i represents the inverse operation representing the average accessible distance between the particle and the particle center, and B represents the average accessible distance between the particle and the particle center.

Inside the Grain Interval. Inside the grain interval, describe the degree of intimacy between the particles:

$$d = \frac{\sum_{i=1}^n (B \cdot w/r_i)}{n}, \quad (8)$$

where r_i represents the radius of the particle i and w represents the degree of membership.

Based on the above process, the embedded multifunctional data feature extraction work based on particle calculation is completed.

2.4. Embedded Multifunctional Data Mining Implementation. With the above extracted features used as input, the embedded multifunctional data mining is implemented to achieve data classification, abnormality detection, fault identification, and other goals [18]. Here, the neural network method is used to realize the embedded multifunctional data mining. The basic structure of the neural network is shown in Figure 3.

Embedded multifunctional data mining based on the neural network is divided into two steps, namely, training and testing.

- (1) The training uses the extracted three embedded multifunctional data features, namely, offset, particle density, and the input interval and the output of the processing and operation, and the neural network. If the matching results and the expected set results meet the end conditions, the training will end; otherwise, error backpropagation is performed until the training is successful.
- (2) The test is based on the former training of the good model, to complete the mining of the test samples.

3. Technical Testing and Analysis

For the embedded multifunctional data, the mining technology based on particle computing is taken as an example,

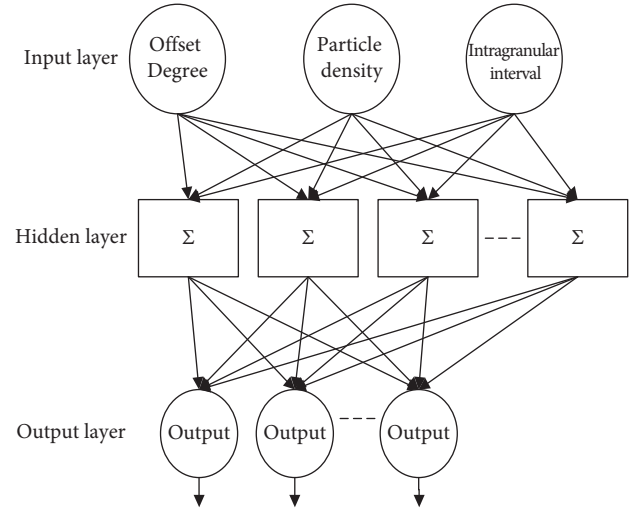


FIGURE 3: Basic structure of the neural network.

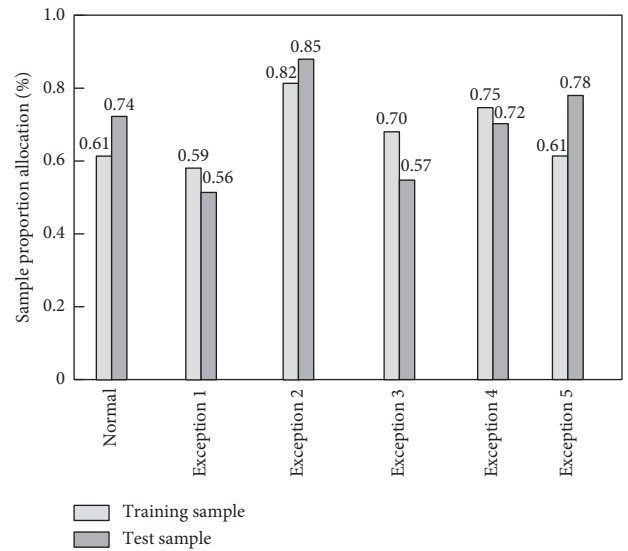


FIGURE 4: Test sample allocation plot.

which is applied to the network anomaly detection to test the effectiveness of the mining technology. The simulation test platform is Matlab 2016.

3.1. Simulation Sample. Six types of data were selected from the DARPA KDD CUP 99 dataset to form the embedded multifunctional data simulation samples, with a total number of 10,000 samples. The sample proportion allocation is shown in Figure 4.

Since the samples were obtained from the standard DARPA KDD CUP 99 dataset, the preprocessing process was not analyzed in detail.

3.2. Embedded Multifunctional Data Particle Feature. In Section 1, the study was used to granulate the embedded multifunctional data samples and then we calculated the data particle features. The results are shown in Figure 5.

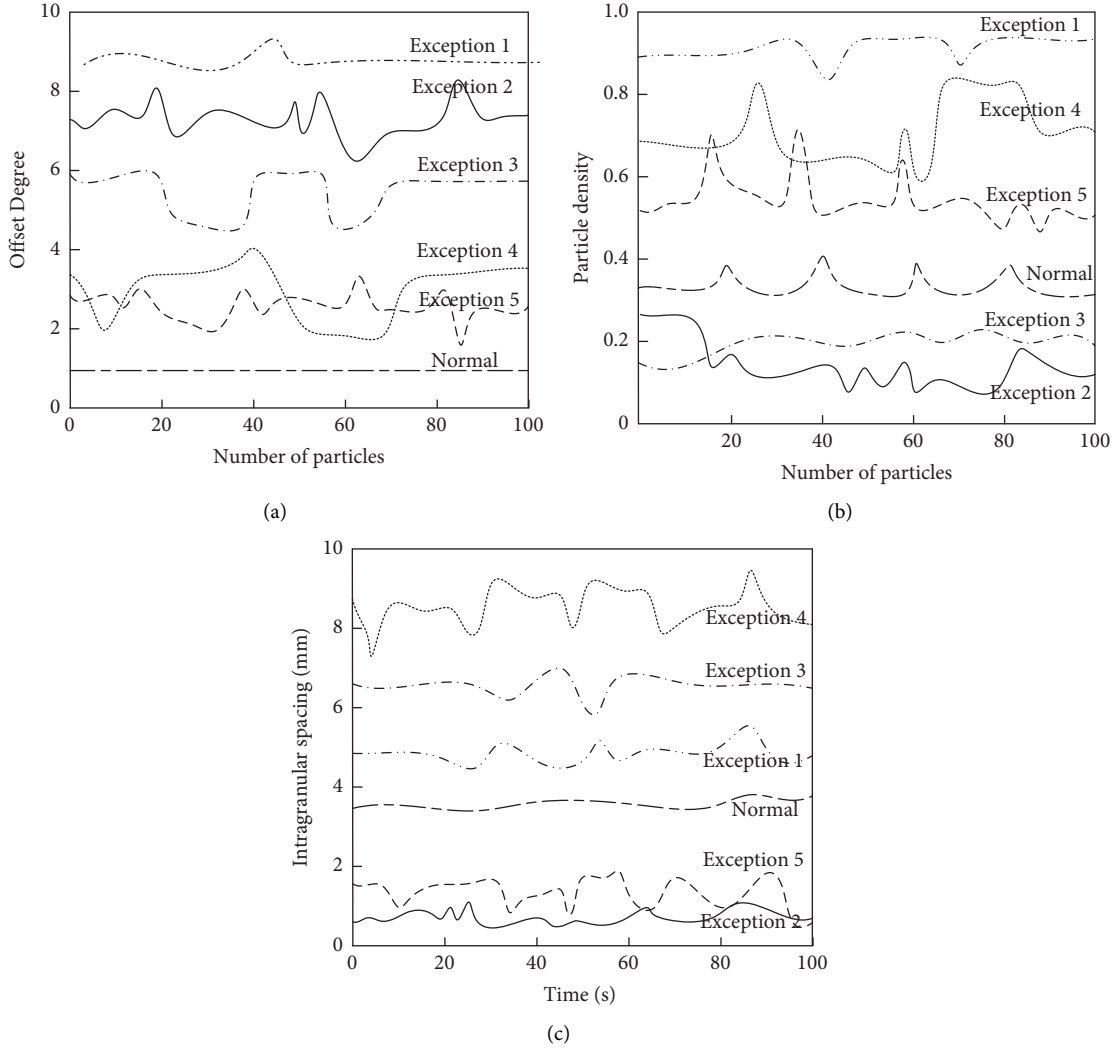


FIGURE 5: Embedded multifunctional data particle feature. (a) Drift rate. (b) Granule density. (c) Internal particle spacing.

3.3. *Test Indicators.* G -mean proposed by Kubat is the evaluation index of embedded multifunctional data mining technology. The calculation formula is as follows:

$$G\text{-mean} = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \quad (9)$$

The various index parameters in the equation are derived from the confusion matrix, as shown in Table 1.

G -mean takes the value (0,1]; when greater than 0.9, the mining technical accuracy is high.

4. Results and Analysis

The neural network was trained using the training samples, and the post-training weights were set to 0.25 and 0.36; the thresholds were set to 1.20 and 1.50. Taking the test sample as input, the trained neural network model was used for embedded multifunctional data mining to obtain anomalous mining results. Finally, the values were calculated from the anomalous mining results, as shown in Table 2 G -mean.

TABLE 1: Confounding matrix.

Class	Positive class	Negative class
Positive class	TP	FN
Negative class	FP	TN

TABLE 2: G -mean values for the statistical results.

Type	1 test	Two tests	Three tests	4 tests	5 tests
Normal	0.925	0.921	0.936	0.945	0.951
Abnormal 1	0.932	0.934	0.935	0.941	0.932
Abnormal 2	0.965	0.952	0.932	0.952	0.934
Abnormal 3	0.951	0.941	0.914	0.953	0.956
Abnormal 4	0.923	0.920	0.922	0.940	0.955
Abnormal 5	0.920	0.923	0.924	0.933	0.942
Average value	0.936	0.932	0.928	0.944	0.936

As can be seen from Table 2, the abnormal mining result of each type of data mining is greater than 0.9, thus indicating that the accuracy of the studied mining technology is high.

In conclusion, the studied mining technology was used to carry out abnormal mining of embedded multifunctional data many times, and in different cases, the results of each type of data mining were all greater than 0.9, showing high accuracy and good results.

5. Conclusion

Data mining is the most important issue in big data processing, where categories, rules, and even abnormalities can be found from the data. Current data mining is limited by the embedded multifunctional data features, and the mining accuracy is not high. For the above problems, an embedded multifunctional data mining technique based on particle calculation was studied. This technology has been tested and its effectiveness has been proved. It can cope well with the embedded multifunctional data mining technology based on particle computing, with high accuracy and good modification. However, this study only tested the technology in one field, and therefore, the test results have limitations. Further testing is needed, and in the future, particle calculation in embedded multifunctional data mining can be improved. From the perspective of collaborative innovation, data mining technology innovation ability can also be improved. The key is to face the characteristics of the Internet of Things industry and to explore the technology collaborative innovation process and behavior collaborative interaction mode through in-depth data mining and analysis to develop more intelligent applications.

Data Availability

The data for all figures 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.

Authors' Contributions

Juan Li carried out the constructions and drafted the manuscript. Xianghong Tian conceived of the study and participated in its conceptualization, supervision, and coordination and helped to improve the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the Jiangsu Higher Education Reform Research Project (2021jsjg641), the Jiangsu Educational Science "14th five-year plan" Project (B/2021/01/13), and the Industry University Cooperation Collaborative Education Project (202101225008).

References

- [1] L. Zhang, X. Li, and L. Jing, "Research and application of layered data mining method based on game teaching," *China Audio-visual Education*, vol. 385, no. 2, pp. 87–94, 2019.
- [2] M.Y. Lee, "Privacy protection exploration for educational data mining based on feated learning," *Audio-visual Education Research*, vol. 41, no. 11, pp. 94–100, 2020.
- [3] Z. Wang, Y. Feng, and C. Zhu, "Research on information analysis method based on data mining technology—takes the price prediction of container shipping as an example," *Intelligence Science*, vol. 37, no. 7, pp. 65–71, 2019.
- [4] L. Zhang and Y. Ma, "Research on aviation customer loss and segmentation and R language program implementation based on data mining technology," *Practice and understanding of mathematics*, vol. 49, no. 6, pp. 134–142, 2019.
- [5] L. Ma and J. Liu, "Research on anomaly data detection of optical fiber communication network based on data mining," *Applied Optics*, vol. 41, no. 6, pp. 1305–1310, 2020.
- [6] Y. Wang and J. Ku, "Research on collaborative innovation platform of Internet of things industry based on data mining technology," *Journal of Physics: Conference Series*, vol. 1881, no. 4, pp. 042072–42150, 2021.
- [7] L. Yi and J. Hu, "A neighborhood outlier detection for mixed attribute data," *Small Microcomputer system*, vol. 41, no. 04, pp. 855–860, 2020.
- [8] D. Liu, T. Li, X. Yang, and D. Liang, "Three-branch decision-making-based on rough set and grain computing research perspective," *Journal of Intelligent Systems*, vol. 14, no. 06, pp. 1111–1120, 2019.
- [9] L. Zhang, F. Qian, S. Zhao, J. Chen, Y. Zhang, and F. Liu, "Network representation learning based on a multi-granularity structure," *Journal of Intelligent Systems*, vol. 14, no. 06, pp. 1233–1242, 2019.
- [10] J. Shen, Q. Yan, Q. Sun, and Z. Wan, "Research on technology fusion identification and technology opportunity prediction based on patent data mining—takes the electric vehicle industry as an example," *Library Magazine*, vol. 38, no. 10, pp. 95–106, 2019.
- [11] J. Wang and T. Li, "Research and application of energy consumption based on big data mining," *Highways*, vol. 64, no. 04, pp. 228–232, 2019.
- [12] H. Zhou, B. Lin, Z. Zhang, J. Qi, L. Zheng, and C. Chang, "Discussion on the data processing methods of residential water and gas use in northern cold areas based on data mining technology," *HVAC*, vol. 49, no. 02, pp. 58–66, 2019.
- [13] Z. Liang, X. Guo, J. Guo, Y. Han, Q. Zhu, and X. Xiong, "Study on hyperspectral soil texture classification based on data mining technology," *Chinese Agricultural Science*, vol. 53, no. 21, pp. 4449–4459, 2020.
- [14] P. Tang and Y. Dong, "Online classification of traffic based on the granular relationship matrix," *Journal of Electronics*, vol. 49, no. 01, pp. 1–7, 2021.
- [15] J. Li, F. Wang, W. Wu, X. Yang, and Y. Zhe, "Summary of multi-granular data analysis methods based on particle calculation," *Data acquisition and processing*, vol. 36, no. 03, pp. 418–435, 2021.
- [16] Y. Xu and Y. Yao, "Division sequence product space: a particle calculation model based on division," *Computer Research and Development*, vol. 56, no. 04, pp. 836–843, 2019.
- [17] C. Ai, D. Jiang, and J. Wu, "Patent text data mining study based on subject model and association rules," *Journal of North University of China (Natural Science Edition)*, vol. 40, no. 06, pp. 524–530, 2019.
- [18] J. Yu, J. Fu, Baitana, S. Li, and W. Hong, "Data mining research on emotion and context interaction based on unique property characteristics," *Journal of Yanshan University*, vol. 43, no. 05, pp. 462–470, 2019.

Research Article

Knowledge Graph-Enabled Text-Based Automatic Personality Prediction

Majid Ramezani ¹, Mohammad-Reza Feizi-Derakhshi ¹,
and Mohammad-Ali Balafar ²

¹Computerized Intelligence Systems Laboratory, Department of Computer Engineering,
Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

²Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

Correspondence should be addressed to Majid Ramezani; m_ramezani@tabrizu.ac.ir and Mohammad-Reza Feizi-Derakhshi; mfeizi@tabrizu.ac.ir

Received 3 March 2022; Revised 24 April 2022; Accepted 26 May 2022; Published 20 June 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Majid Ramezani 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.

How people think, feel, and behave primarily is a representation of their personality characteristics. By being conscious of the personality characteristics of individuals whom we are dealing with or deciding to deal with, one can competently ameliorate the relationship, regardless of its type. With the rise of Internet-based communication infrastructures (social networks, forums, etc.), a considerable amount of human communications takes place there. The most prominent tool in such communications is the language in written and spoken form that adroitly encodes all those essential personality characteristics of individuals. Text-based Automatic Personality Prediction (APP) is the automated forecasting of the personality of individuals based on the generated/exchanged text contents. This paper presents a novel knowledge graph-enabled approach to text-based APP that relies on the Big Five personality traits. To this end, given a text, a knowledge graph, which is a set of interlinked descriptions of concepts, was built by matching the input text's concepts with DBpedia knowledge base entries. Then, due to achieving a more powerful representation, the graph was enriched with the DBpedia ontology, NRC Emotion Intensity Lexicon, and MRC psycholinguistic database information. Afterwards, the knowledge graph, which is now a knowledgeable alternative for the input text, was embedded to yield an embedding matrix. Finally, to perform personality predictions, the resulting embedding matrix was fed to four suggested deep learning models independently, which are based on convolutional neural network (CNN), simple recurrent neural network (RNN), long short-term memory (LSTM), and bidirectional long short-term memory (BiLSTM). The results indicated considerable improvements in prediction accuracies in all of the suggested classifiers.

1. Introduction

Personality is the enduring set of traits and styles that an individual exhibits, that is, those characteristics that represent his/her dispositions, namely, natural tendencies or personal inclinations [1]. Being aware of the personality characteristics of people will help them improve their relationship management skills and also ameliorate their interpersonal communications, regardless of the type of relationship, as it happens between two friends, the boss and employee, investor and investee, seller and buyer, between members of a family, and so on.

With the advent of social networks and their remarkable fortune among people, nowadays, a great deal of communication happens through social networks. Language, as the main communication tool among humans that competently represents their thoughts, emotions, opinions, and totally personality, is also used in written and spoken form among social networks' users to communicate with each other. Admittedly, having some information about the personality with whom you are communicating would be so advantageous. It can be carried out by analyzing the exchanged texts (which is also known as written language), among the users of such information infrastructures. Accordingly, the automatic prediction of human personality through

computational approaches is called Automatic Personality Prediction (APP).

What we know about text-based APP is largely based upon empirical studies that have investigated how to exploit different methodologies for the purpose of personality prediction of individuals in Internet-based infrastructures (like social networks). Actually, various hypotheses regarding this issue can be found that they are commonly concerned about achieving a more knowledgeable substitutions for text elements to deal with, rather than pure strings of characters.

In the history of text-based APP, initial investigations have mostly focused on linguistic features of text elements to achieve more knowing about them [2–5]. Over the years, it has received much attention; while some studies have applied a combination of linguistic features and machine learning methods [6–9], some others have focused solely on the machine (deep) learning methods [10–13]. In the last few years, we have witnessed a considerable rise in text-based APP, which have used embedding methods to transfer the text elements to a more meaningful space (rather than character space), in favor of better exploitation of computational methods [14–17].

Generally, it can be inferred that all of the investigations are intended to acquire more knowing about the text elements, each of which is done through applying miscellaneous methods. Indeed, they are absolutely right; namely, the knowing will be the basis of predictions.

Although various researches have been carried out on text-based APP, no study has been found that is essentially intended to focus on Knowledge Representation (KR). This paper for the first time (to the best of our knowledge) calls into question the application of knowledge representation and thereby knowledge graph in text-based APP. Specifically, this study makes a major contribution to research on automatic personality prediction by proposing a novel knowledge graph-enabled system. Indeed, it meticulously investigates knowledge representation as a novel solution for text-based personality assessment. We believe that there is knowledge, and then there is knowing. Therefore, at first, we should discover the world behind the words. It will provide an important opportunity to advance the understanding of text elements. In consequence, practically, the significance of our method is that it empowers an APP system to achieve a comprehensive representation of the appearing concepts in the input text, which entails the knowledge behind them and models the semantic relations among them, in a more comprehensible manner for the machine, as the basis of its predictions. In fact, the proposed method equips machine with the required knowledge to acquire a better understanding of the entailed concepts in the input text and accordingly achieve better results.

Knowledge representation is a field of artificial intelligence dedicated to representing information about the world in a form that a computer system can utilize to solve complex tasks [18]. In fact, knowledge representation is necessary to understand the nature of intelligence and cognition of concepts so well that computers can be made to exhibit human-like abilities [19]. Therefore, in the case of expecting

human-like abilities from artificial intelligence, it seems that we ought to represent the knowledge of the world for it. Meanwhile, the knowledge graph is actually the outcome of knowledge representation. It organizes the knowledge of concepts in a graph structure and integrates all existing information about them.

Therefore, aimed to design a knowledge graph-enabled text-based APP system, this paper proposes a three-phase approach that includes:

- (i) Phase 1: preprocessing that contains four steps of needed preprocessing, that is, tokenization, noise removal, normalization, and named entity recognition to make the input text ready for main processes in next phase.
- (ii) Phase 2: knowledge representation, the main contribution of this study that comprises three steps, that is, graph building, graph enriching, and graph embedding. In practice, this phase first attempts to build the corresponding knowledge graph for a given text, which is a knowledgeable representation of the input text and then enriches it to cover some neglected pieces of knowledge about concepts. At last, the acquired enriched graph is embedded to a more computationally applicable space, to facilitate the computations in next phase.
- (iii) Phase 3: automatic personality prediction that aims to predict the personality traits for each input text through a multilabel classification model. To do so, four base deep learning models were proposed that includes Convolutional Neural Network- (CNN-) based, simple Recurrent Neural Network- (RNN-) based, unidirectional Long Short-Term Memory- (LSTM-) based, and Bidirectional Long Short-Term Memory- (BiLSTM-) based classifiers.

This study aimed to address the following research questions:

RQ.1: How does knowledge graph enabling influence the performance of a text-based automatic personality prediction system?

RQ.2: What are the performances of popular deep learning models, including CNN, simple RNN, LSTM, and BiLSTM in multilabel classification of knowledge graphs' embedding matrices? Which of them outperforms the others?

RQ.3: Does knowledge graph enabling of an APP system affect equally the predictions in all five personality traits in Big Five model?

The remaining part of the paper proceeds as follows: Section 2 is concerned with the Big Five personality model. Section 3 provides an overview of text-based APP systems. The proposed knowledge representation-based APP system is meticulously demonstrated in Section 4. Section 5 presents the findings of this study, and then Section 6 includes a discussion of the implication of the findings as well as responses to the research questions. Finally, Section 7, namely,

the conclusion, gives a brief summary and critique of the findings.

2. The Big Five Personality Model

So far, various personality trait models have been introduced [20]. In this study, the Big Five model (Five Factor Model) [21], as the most widely accepted trait model that capably correlates with human traits that are presented in written language [22], is used. It basically demonstrates the individuals' personality in five categories: openness, conscientiousness, extroversion, agreeableness, and neuroticism. OCEAN is the acronym of the five categories, which we shall refer to, also. Each of the five personality trait represents a range between two extremes [23]; that is, extroversion represents a continuum between extreme extroversion and extreme introversion. To make it more clear, indicating some facets of each trait, which occurs in those people with high scores for each trait, may be useful (for more details, please refer to [24, 25]):

- (i) Openness (O): an inclination to embrace new ideas, arts, feelings, and behaviors; unconventional; focused on tackling new challenge; wide range of interests and so imaginative.
- (ii) Conscientiousness (C): an inclination to be so self-disciplined, well organized and dutiful; careful and hard-working; reliable, resourceful and on time.
- (iii) Extroversion (E): an inclination to be outgoing, energetic, assertive and talkative; affectionate, sociable and articulate; enjoys being the center of attention.
- (iv) Agreeableness (A): an inclination to agree and accompany the others; altruist and unselfish; friendly, loyal and patient; modest, considerate and cheerful.
- (v) Neuroticism (N): an inclination to experience negative emotions like anxiety, anger, depression, sadness, and envy; impulsive and moody; lack of confidence.

Furthermore, it is worth noting that the Big Five traits are mostly independent [23]. It means that being cognizant of someone's one personality trait does not provide so much information on the remaining traits of the Big Five model.

3. Literature Review

In recent years, there has been an increasing amount of literature on APP that mainly pays particular attention to predict the personality from text, speech, image, video, and social media activities (likes, visits, mentions, digital footprints, profile interpretation, etc.).

Text as an appearance of human language would competently reflect the writer's personality [22]. Due to this fact, it is always a matter of concern for personality psychologists. Spreading the Internet-based communication infrastructures increased the text-based communications among people. It opens the door for computational psychologists to investigate the personality of writers from

exchanged texts. Here, we will review the researches conducted on text-based APP.

Taking a glimpse into the reported investigations in APP, it can be claimed that, generally, all of them are intended to acquire more meaningful and knowing-full alternatives to input text's elements (namely, words, terms, or, generally, all the appearing concepts) to deal with. In simple words, dealing with more meaningful alternatives that convey more knowing and information rather than pure character strings is highly preferred. Actually, this knowing about written language elements may better represent the knowledge behind them and may lead better predictions about writer's personality. Tracing the evolution of text-based APP systems sheds more light on this claim. With respect to this claim, generally, we can classify the previous studies into five categories: lexicon-based methods, hybrid methods (combination of lexicon-based and deep learning-based methods), embedding methods, ensemble modeling methods, and network-based methods. A detailed analysis of these categories is given below.

3.1. Lexicon-Based Methods [2–5, 26]. Rudimentary techniques have mainly tended to utilize lexicons, which provide linguistic and statistical knowing about text elements. Lexicon-based methods primarily try to predict the personality of writer through assigning his/her words to pre-determined categories. Linguistic Inquiry and Word Count (LIWC) [27] is one of the most common tools that counts words in psychologically meaningful categories and calculates the degree to which people use different categories of words. It is simply a dictionary of words and word stems, each of them belonging to a one or more category. Given a text, LIWC calculates the percentage of included words in each category. The main idea behind LIWC is that the word usage in everyday language reveals the thoughts, personality, and feeling of individuals. There have been different versions available since 2001. Further information is available at <https://liwc.wpenline.com/> and there are more than 80 categories in LIWC2015. Mairesse features [2] and Structured Programming for Linguistic Cue Extraction (SPLICE) are other options that provide linguistic features for words.

In their analysis of APP from the words that people use, Yuan et al. [5] have investigated the personality of the characters in vernacular novels. They have created a vector for each dialog using LIWC features, which reflects the psychology of the characters. Finally, the vectors have been mapped to the Big Five personality traits, to predict the final personality labels. Mairesse et al. [2] have also investigated a miscellaneous variety of lexicon-based features in order to predict the Big Five personality traits from written text and spoken conversation.

Among the first reports on APP from the social media text, Golbeck et al. [3] have considered LIWC features over the 167 samples of Facebook text contents as well as the users' profile information. The results confirmed limited improvements in APP. In the same manner, the authors in [4] have studied the Big Five personality traits from Twitter posts besides the users' profile attributes. They have actually

intended to find antisocial traits of narcissism, Machiavellians, and psychopathy (commonly referred to as the dark triad) through using LIWC features. Perusing the reported results implies that prediction of personality traits from social media text using lexicon-based methods could not considerably improve the APP accuracy.

Later, in a study that has been set out to predict personality traits from social networks' microblogs, Han et al. [26] have found that the context-based knowledge of words may be advantageous to personality prediction. They have believed that since the traditional psychological lexicons (like LIWC) are appropriate for formal texts, they could not efficiently be applied in social networks' informal texts. Therefore, they have proposed an approach to automatically extract a personality lexicon from social networks, through using keyword extraction techniques and then semantically clustering the extracted keywords. At last, they have simply combined the extracted lexicon (as a prior-knowledge source) with the word embedding vectors and have fed them into a classification model, to predict the Big Five's personality traits' labels. They have partially enhanced the prediction accuracy, even though they have just taken the advantage of words' lexical knowledge.

3.2. Hybrid Methods [6, 9, 11, 28]. Generally, in the literature, there seems to be no a tendency among researchers to use lexicon-based methods, solely. Telling the truth, it is hardly fair to shift all the APP responsibility solely to them, because of their superficial knowledge of text elements. Consequently, a large and growing body of literature has investigated the combination of lexicon-based methods with more knowing-full methods that fairly has improved proportionally the predictions' accuracy.

Designing a convolution neural network (CNN), which uses the document-level Mairesse features (extracted from the input text) in an inner layer, has formed the central focus of a study by Majumder et al. [6]. They have trained a separate identical binary classifier for each of the five personality traits in the Big Five model that receives sentences of the input text one by one and then aggregates them into a document level vector. Besides, they have finally ignored all the emotionally neutral sentences, to improve the performance. Yuan et al. [9] have carried out a study to predict the personality of users from their Facebook status contents. Actually, they have combined the LIWC features with deeper features that have been extracted through a deep learning model. They firstly have extracted the language features via the LIWC tool, and then using a CNN, they have automatically extracted the features from textual contents. Subsequently, the two extracted features have been combined to predict the personality labels.

In another investigation into APP from texts in online social networks, the authors in [11] have proposed a bidirectional LSTM model, called 2CLSTM. In order to detect user's personality using the structures of texts, the model has been strengthened by a CNN as well as a latent sentence

grouping module, which has been applied to capture closely connected sentences. Xue et al. [28] studied the effects of semantic representation of words in APP systems. They acquired a word-level semantic representation of text elements and then fed them into a neural network to obtain higher-level semantics of text elements.

3.3. Embedding Methods [14–17, 29–31]. Alongside these researches, many attempts have been made with the purpose of utilizing complex methods that use even more knowing-full alternatives for text elements. Indeed, they have succeeded in achieving better results during predictions. They mostly pay particular attention to embedding methods, which transform the text elements from a textual space to a real-valued vector space. Overall, these methods, despite their variety, have better performance in APP, rather than previously mentioned methods. This ability is a consequence of embedding methods' adroitness in meaning acquisition and representation. In a study that has been set out to detect personality based on text content analysis, Ren et al. [14] have investigated a novel multilabel personality prediction learning model, which combines emotional and semantic features. In particular, they have leveraged a Bidirectional Encoder Representation from Transformers (BERT), to generate sentence-level embeddings for extracting semantic features from text, as well as a sentiment dictionary for the sake of text sentiment analysis purposes. Encoders primarily are designed for achieving a knowing-full representation of input text. They have used the Myers-Briggs Type Indicator (MBTI) and Big Five personality trait models in their study. Xue et al. [15] have also designed a deep learning-based method for personality prediction from text, which are posted in online social networks. They have recommended AttRCNN, a hierarchical model that uses a sentence-level encoder that is followed by a document level encoder in order to achieve the deep semantic features of text posts. Moreover, they have concatenated the deep semantic features with the statistical linguistic features obtained directly from the text posts and have fed them into a regression model to predict the Big Five personality traits' labels. Exploiting the embedding methods abilities, in their study, Christian et al. [17] have suggested a multimodel deep learning architecture for personality prediction, which was combined with various pretrained language model including BERT, RoBERTa, and XLNet as a feature extraction method on social media text. The main idea behind their investigations was that since the common deep learning models such as recurrent neural networks (RNNs) and LSTMs suffer from some drawbacks that are defeated using embedding methods, the embedding methods practically outperform them. Specifically, they mostly suffer long training times and inability to capture the context-based information of words and thereby the true meaning of words. At last, the final predictions have been taken based on averaging the output of different pretrained models. Other researches ([16, 29–31]) have also investigated designing embedding-based APP models that make predictions from text.

3.4. Ensemble Modeling Methods [13, 24, 32–34]. Meanwhile, taking the advantages of several classifiers and benefit their prediction abilities simultaneously was a matter of concerns for some studies. Utilizing different APP models predictions, the authors in [24] have proposed an ensemble modeling method. Specifically, they have suggested five separate APP models, including term frequency vector-based, ontology-based, enriched ontology-based, latent semantic analysis-based, and deep learning-based (BiLSTM) methods. Then, all of the individual five models have been gathered through a Hierarchical Attention Network (HAN) as the meta-model. In consequence, they have benefited the ability of five distinct APP model, to make the final decisions about Big Five personality traits. In their study, El-Demerdash et al. [13] have suggested a transfer learning-based APP method that have got the benefits of leading pretrained language models such as Elmo, ULMFiT, and BERT. To raise the overall personality prediction performance, they have applied a model consists of fusion strategies on data level and classifier level. Adopting the tree pretrained models, they have used the fusion of Essays and my personality datasets for further fine-tuning of the proposed models. Using independent classifiers, each model performs APP separately. Then, the results have been fed into an ensemble learning model that combines multiple classifiers' outputs, to acquire more reliable prediction. Having the same objectives, other researchers [32–34] have questioned the usefulness of such an approach.

3.5. Network-Based Methods [35, 36]. There are also a number of investigations that have aimed for achieving a different representation. They mainly have focused on modeling the network among the online social media users. The first report on group-level personality prediction was conducted by Sun et al. [35]. They have proposed an unsupervised feature learning method called AdaWalk that takes the advantage of independence from labeled dataset. Actually, it was designed based on Network Representation Learning (NRL) method, which was suggested by the authors. Practically, it constructs a complete graph in which its vertices are the users. The graph also possesses the generated texts for each user, the similarity between each users' texts, and the personality labels in Big Five model. Subsequently, applying random walks (AdaWalks) on the graph, they have transformed the network to a set of sequences and finally have predicted their personality labels after embedding all of them. In the same vein, Guan et al. [36] have suggested personality2vec, which predicts the personality labels based on NRL using online social networks' texts. The authors have intended to fully utilize the semantic, personality based, and structural information of user generated texts.

Regarding the evolution, the aforementioned claim that all of the contributions have been attempted to achieve more meaningful alternatives for text elements to deal with would thus seem to be defensible. Actually, the contributions provide strong experimental evidences that more knowing-full alternatives for text elements may lead more reliable results. What is not yet clear is the impact of an approach

that is fundamentally based on knowledge representation of text elements on APP. An approach that provides really knowledgeable alternatives for text elements conveys all of the related information and knowing about the concepts as well as their relations.

4. Material and Methods

The purpose of knowledge representation approach is to demonstrate the cognitive perceptions behind the key concepts in the world, as well as the relations among them. The dexterity of intelligent functionality is remarkably correlated with existed represented knowledge, both for human and seemingly for machine. We thus primarily decided to represent the knowledge behind the input text elements in favor of APP objectives. To do so, it was decided to manipulate RDF modeling. The aforementioned abilities of RDF model justify its competency in knowledge representation.

4.1. Dataset and Some Statistics about It. In this study, the provided essays in Essays Dataset [37] were used for training and testing the proposed APP model. It consists of 2,467 essays, which are written by psychology students. Afterwards, they were asked to fill out the Big Five Inventory Questionnaire. At the end for each essay, a binary label was assigned to each five personality traits. Throughout this paper, each individual essay will be referred to as text. Moreover, it should also be noted that the Big Five personality model was used all over the investigations.

Let us scrutinize much more information about the Essays Dataset. Figure 1 depicts the distribution of True and False labels throughout the dataset individually in each of the five personality traits. Slight difference between the number of True and False labeled essays reveals that the dataset is balanced and appropriate for learning the APP model.

Figure 2 compares the correlations among the five personality traits in Essays Dataset. As it can be seen, a correlation matrix is a symmetric matrix, in which all the values on the main diagonal are equal to 1. The correlation coefficient can range between -1 and $+1$. The larger the absolute value of coefficient, the stronger the relationship between two traits. Specifically, a positive coefficient between two traits means that being aware of one trait's label allows a correct prediction of the other; as close as possible to $+1$, it will conclude more correct predictions.

The UpSet [38] plot of five sets of personality traits is presented in Figure 3. An UpSet plot actually is considered as a substitution for Venn diagram, when dealing with more than 3 sets. Having five sets of personality traits (namely, O, C, E, A, and N), the UpSet plot makes it possible to provide an efficient way to visualize the intersections of five sets. Each row at the bottom of Figure 3 denotes to a set, and each column corresponds to one segment in Venn diagram, depicted with five light or black circles. A black circle indicates that the corresponding set is participating in the intersection, and a light circle vice versa. Indeed, a light circle

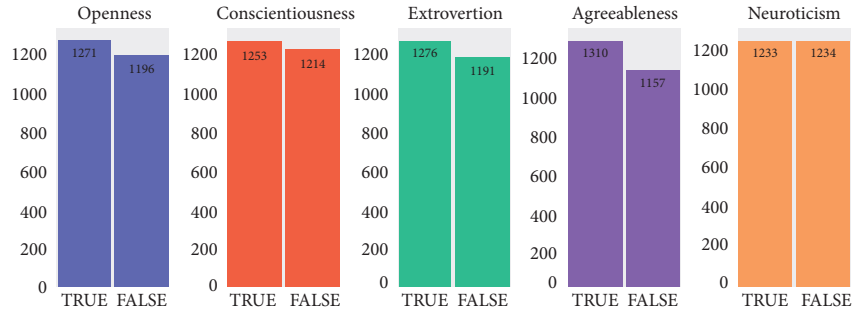


FIGURE 1: The distribution of labels in each five personality traits in Essays Dataset.

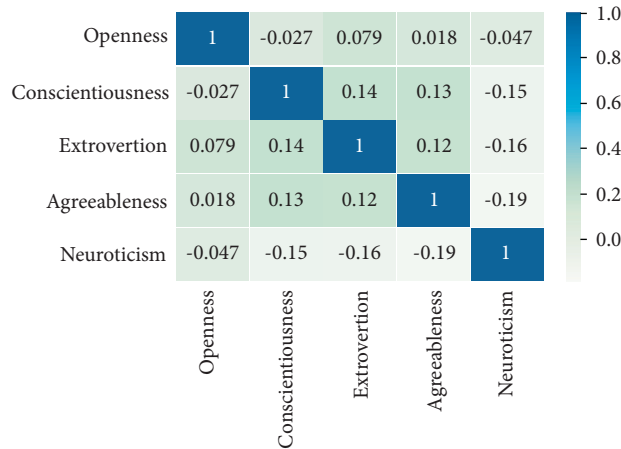


FIGURE 2: The correlation matrix for five personality traits in Essays Dataset.

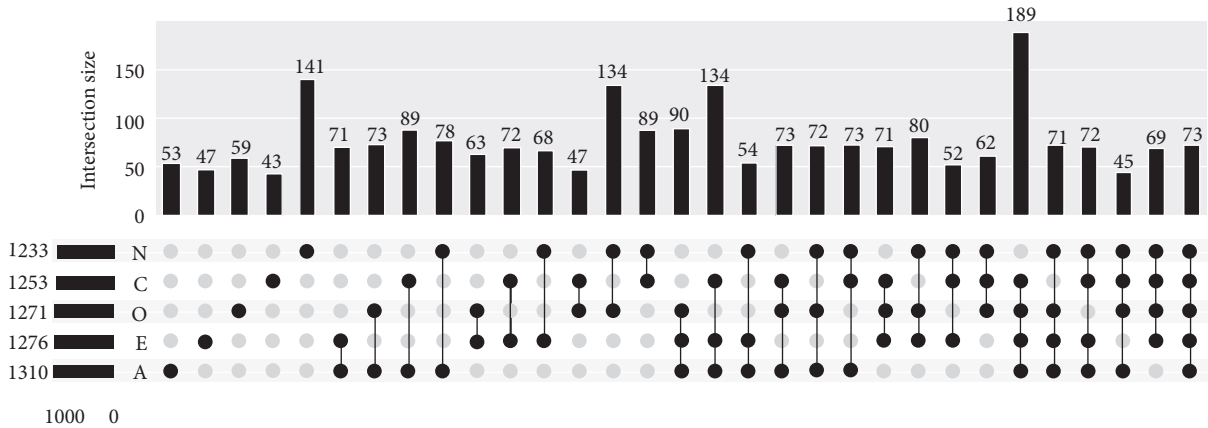


FIGURE 3: The UpSet plot of intersections among true labeled sets of personality traits in Essays Dataset. Notes: (i) sets {O, C, E, A, N} are sorted by their cardinality in ascending order; (ii) light (empty) circles indicate that the set is not part of that intersection.

indicates that the complement of the set (O' , C' , E' , A' , or N') is participating in the intersection. In particular, the rightmost column that has five black circle for all of the five sets is equal to $(O \cap C \cap E \cap A \cap N)$. The bar chart on top of Figure 3 represents the cardinality of each corresponding intersection. It worth mentioning that the plot depicts the intersections among true labeled essays' sets. That is to say, just the true labeled essays in OCEAN traits are taken into consideration.

4.2. System Architecture. Aimed to answer the research questions stated at the beginning of this study, we suggested a three-phase approach, which is outlined in Figure 4. The experiment proceeds with the following phases below.

4.2.1. Phase 1: Preprocessing. In this phase, the aim was to clean and transform the input texts into a more digestible form for machine to be processed in next phase. This

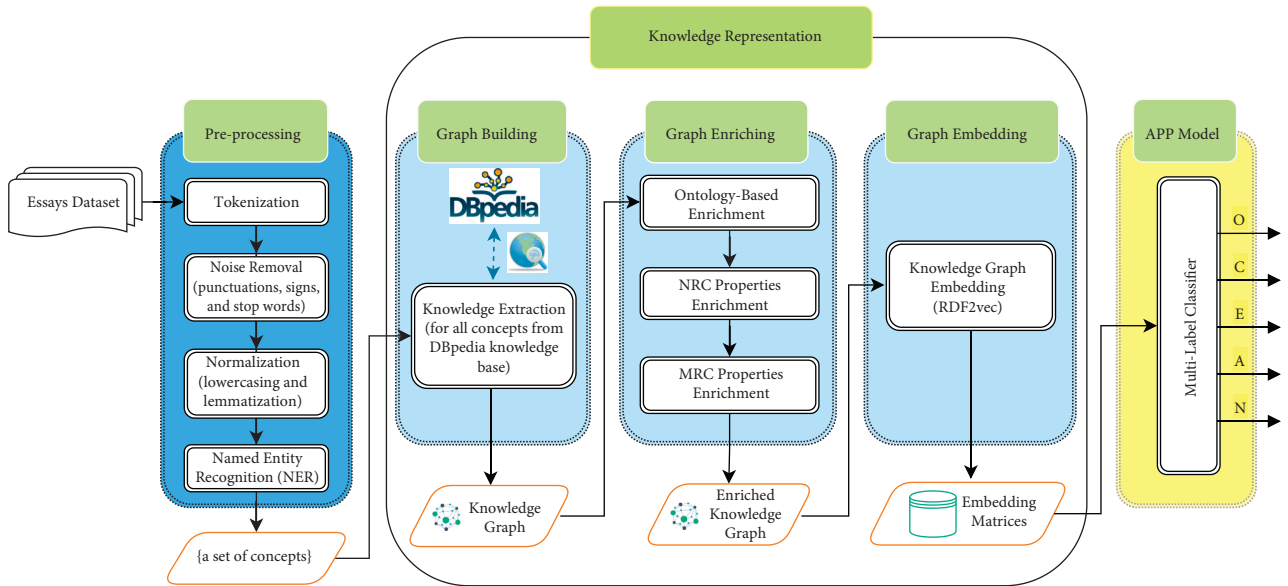


FIGURE 4: System architecture.

traditional prominent and common practice in natural language processing basically consists of miscellaneous activities depending on the existing task. What follows is a description of preprocessing activities that were carried out in first phase, as depicted in Figure 4.

(i) *Tokenization*. Having a text, “tokenization” is the task of chopping it up into pieces called tokens, which roughly correspond to words [39]. Tokens are also deemed as the smallest useful semantic unit for processing. For this purpose, the tokenizer, which is provided by Natural Language Toolkit (NLTK) [40], was used.

(ii) *Noise Removal*. In the interest of achieving more plain text, it was necessary to remove undesirable and interfering pieces of input text. Regarding the current task, we removed punctuations, signs, and stop words using NLTK.

(iii) *Normalization*. “Normalization” is the process of canonicalizing tokens to a more uniform sequence, so that matches occur despite superficial differences in the character sequences of the tokens [39]. It practically decreases the amount of information that the machine has to deal with, those that are conceptually similar, but morphologically different. In an attempt to normalize the input text, lowercasing and lemmatization were carried out.

Lemmatization is the morphological analysis of words that groups together their inflected forms and returns their bases or dictionary forms, which are called lemma. Since lemmatization converts words to their meaningful dictionary form and yields the correct form of concepts, which really exist in the world compared to stemming, which is an alternative method to reduce inflected words to their stems and usually is fulfilled through chopping off the ending characters of the word, it usually returns incorrect and misspelled forms of words, and it is appropriate in the current task. The resulting meaningful concepts will be

queried during knowledge graph building in the following phase. In this study, the lemmatization was also carried out using NLTK.

(iv) *Named Entity Recognition (NER)*. With an eye to achieving the knowledge behind the words, it will be necessary to recognize the named entities from the input text. The sequences of words are actually the name of things (that is to say, the name of organization, person, company, event, etc.). As a matter of fact, they convey more information than do the other words. In current study, the spaCy NER [41] was used to recognize named entities.

After the completion of preprocessing phase, what is extant constitutes a set of concepts, which convey the fundamental notions that have appeared in the input text. It should be mentioned that, after the NER, duplicate elements were removed from the set of concepts. Then, in order to prepare the elements to be matched with DBpedia knowledge base entries, first letter capitalization was performed for all of the elements, and white space replacement with underscore was done for multiword elements as well. Now, everything is ready to find out the world behind the words.

A brief summary of the phase 1, as it can be seen in Figure 4, may be described as follows:

- (i) Input: essays’ texts from Essays Dataset;
- (ii) Output: a set of extracted concepts for each text;
- (iii) Objective: to prepare a more digestible form of input text for main processes in next phases.

4.2.2. Phase 2: Knowledge Representation. As it was stated in Introduction, we chose the graph structure to represent the existing knowledge of concepts in the input text, as well as the relations among them, and to eventuate a knowledge graph for each text. The two first steps in bellow fully describe how a knowledge graph was built for each text. The

purpose of the current phase is to attain a comprehensive representation of existing knowledge of the input set of concepts, so that it could be applied for subsequent computations. Hence, the resulting knowledge graph was transferred to a numerical space using a graph embedding method in third step. The suggested three-step procedure is shown in Figure 4 and proceeds as follows.

Step 1. Knowledge Graph Building

As a matter of fact, the set of extracted concepts from the input text in phase 1 substantially organizes the existent notions in it. There is always knowledge behind every concept. The current step is intended to extract the knowledge of appeared concepts in input text from DBpedia knowledge base [42] and then tries to establish a knowledge graph, which effectively organizes and represents the knowledge of containing text elements [43].

A knowledge graph is a large-scale knowledge base composed of a large number of entities (objects, events, or concepts) and relationships between them [44]. Actually, it is a directed heterogeneous (having vertices/edges of different types) labeled multigraph (a graph, which is allowed to have multiple directed edges between the same pair of vertices), in which the labels have well-defined meanings [45]. The graph structure in knowledge graph adroitly possesses what is needed in knowledge representation [46]. Like all graphs, it consists of vertices and edges, in which the vertices represent the entities of real world, and the edges connect pairs of vertices according to their relationship. What is more, the labels convey the exact information (sometimes called semantics) about the existing relationship (edge) between the vertices. The encompassed knowledge in the knowledge graphs is stored in the form of triples same as (h, r, t) that stands for (head entity, relationship, tail entity). That is to say, having a set of vertices V , along with a set of labels L , the knowledge graph would be a subset of the cross product $V \times L \times V$; each member of this set is referred to as a triple [45]. Each triple may also be interpreted as (subject, predicate, object); for instance (Louvre, is located, Paris). [47] provides detailed information about knowledge graphs.

Meanwhile, there is a well-suited framework that matches as close as possible the knowledge graph's triple requirement, namely, Resource Description Framework or RDF. In essence, it is a standard for representing information in the Web. Equally, this framework is made up of (subject, predicate, object) triples. A set of RDF triples that constructs an RDF dataset can be also viewed as a directed heterogeneous labeled multigraph (like a knowledge graph), which is also referred to as RDF graph [48]. In an RDF graph, vertices (subjects and objects) are either Internationalized Resource Identifiers (IRIs), which stands for a Unicode string representing resources, or literals that contain values such as strings, numbers, and dates. Also, the edges (predicates or labels) are also IRIs representing predicates or relationships. More detailed information about RDFs is available at <https://www.w3.org/TR/rdf11-concepts/>.

Intended to build the knowledge graph of input text during graph building step in phase 2, as previously mentioned, the DBpedia knowledge base is used. DBpedia actually is a community effort to extract structured information from Wikipedia and to make them available on the Web. The 2016-04 release of the DBpedia contains 9.5 billion RDF triples that describe about 6 million entities.

One can easily query on DBpedia dataset online via SPARQL endpoint, which is a standard query language and protocol for linked open data and RDF databases. Thus, we queried all the elements of input concepts' set on DBpedia and extracted all the relevant knowledge of each concept. It was fulfilled through "DESCRIBE" in SPARQL query language (with no binding in SELECT clause and no pattern in WHERE). Specifically, it asks for a description about queried concept (sometimes called resource) and receives any concepts or resources, which are directly related to the queried concept (for further details about SPARQL query language, please refer to [49]). As previously mentioned in Section 1, the results returned from the queries are in the form of RDF triples that provide a set, which is also called RDF graph. RDF graph organizes the knowledge of concepts in a directed heterogeneous labeled multigraph, which is widely known as knowledge graph. It almost encompasses all the (existent) knowledge of concepts. One can find the results for a given query X on DBpedia at <https://dbpedia.org/page/X>. The abundance of resulting RDFs for one query prevented us from exhibiting the concluding results for a sample query. Please note that first letter capitalization and white space replacement with underscore for multiword concepts are necessary.

Step 2. Knowledge Graph Enriching

After building the knowledge graph for the input text, different pieces of information (likewise in form of RDF triples) enrich the current knowledge graph during this step. Enriching the representation inevitably gives more focus to some neglected aspects of facts about entities. In other words, having limited aspects of knowledge will bound the intelligent agent's perception of the world [50]. Consequently, the following graph enrichments were carried out on the resulting knowledge graph.

(i) *Ontology-Based Enrichment*. Ontology actually is a branch of metaphysics dealing with the nature and relations of beings [51]. It demonstrates how the things are related to each other's in a systematic hierarchical classification.

A great deal of attention must be paid that knowledge bases essentially are made up of instances, rather than concepts; what are the foundations of ontologies? Therefore, it would indisputably enhance the representation by means of providing a different aspect of knowledge about things.

To do so, we used the DBpedia ontology. It covers 768 classes (a complete list of covered classes is available in [52]), which are described by 3,000 properties for about 4,233,000 instances. One can easily find the ontology-based representation of a given concept X in DBpedia ontology at <https://dbpedia.org/ontology/X>. At the beginning of foundation, it had been created based on most commonly used

infoboxes within Wikipedia (in 2008) before it evolved into a crowd-sourcing effort. All the RDFs resulting from matching the concepts with DBpedia ontology were added to the previously achieved RDF graph.

(ii) *NRC Properties Enrichment*. Words can be associated with different intensities of an emotion. The NRC Emotion Intensity Lexicon [53], which is provided by National Research Council Canada (NRC), contains real-valued intensity scores for eight basic emotions (namely, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) for about 10,000 entries in English. The lexicon mainly includes more common English words and terms along with those that are more prevalent in social media. The aim of present section is to enrich the representation of input text through enhancing the eight provided emotions' degrees for included words. The emotions' scores for each concept were added to the existed RDF graph in the form of literal RDFs for each word, in case of inclusion in NRC.

(iii) *MRC Properties Enrichment*. The final knowledge graph enriching process was the enhancement of psycholinguistic properties to the RDF graph. It was carried out through MRC psycholinguistic database [54]. MRC is a publicly available machine useable dictionary, which contains (up to 26) linguistic and psycholinguistic attributes (like syntactic, phonological, orthographic, and semantic features) for 150,837 English words. These properties also were added to the existing RDF graph in the form of literal RDFs for each word.

Step 3. Knowledge Graph Embedding

So far, we have achieved the knowledge graph for a given text that basically is made up of RDFs, for both vertices and edges. This step transforms the resulting knowledge graph into a vector space and produces its equivalent embedding matrix. It strives to maximally persevere graph's structure, even though it practically performs dimensionality reduction on it. In this study, the knowledge graphs were embedded according to the method proposed by Ristoski et al. [55]. In their seminal contribution, they proposed RDF2vec, a tool for creating vector representations of RDF graphs. RDF2vec actually is inspired by the word2vec [56], which is a well-known word embedding method (representing words in numeric vector space). RDF2vec almost works similar to word2vec; the major difference is the input sequence. While word2vec receives a set of sentences for training the learning model as the input sequence, RDF2vec uses random walks on the RDF graph to create sequences of RDF vertices to feed them into the same learning model. As a consequence, similar vertices placed close to each other in the final vector space and dissimilar ones do not, like what happens for words after embedding in word2vec. To put it briefly, in this step, the corresponding embedding matrix for a given knowledge graph was achieved.

We set the maximum depth for each walk and the maximum number of walks per entity, both equal to 5 in all the random walks, which were carried out on knowledge graphs. It should be considered that, in practice, the two first

phases, namely, preprocessing and knowledge representation, were iteratively executed for all of the essays in Essays Dataset (please refer to Figure 4) and lasted for more than four months. Experiments were run on a computer with an Intel i7-7700K processor, using 64 GB of ram and running Windows 10. As a result of such iteration, a set of embedding matrices resulting from the knowledge graphs embeddings were achieved, in which the rows of each matrix are dedicated to existing concepts in corresponding essay, and the columns are dedicated to the embedding dimensions. The number of rows in each matrix is different depending on the existing number of concepts in corresponding essay. Therefore, to fix the number of rows and achieve embedding matrices with same number of rows, we selected the 10,000 most frequent concepts in all final resulted knowledge graphs for Essays Dataset's essays. The larger number of rows leads to sparsity of embedding matrices, and the smaller number leads to ignore the included concepts. The number of columns (embedding size), which is specified by RDF2vec, by default is equal to 500.

In consequence, a brief summary of the phase 2, as it can be seen in Figure 4, may be described as follows:

- (i) Input: a set of concepts for each text;
- (ii) Output: equivalent embedding matrix for each text;
- (iii) Objective: knowledge representation for each text; specifically building, enriching, and embedding the corresponding knowledge graph for each text.

4.2.3. *Phase 3: Automatic Personality Prediction*. Finally, four separate classification models were developed to carry out personality prediction, including convolutional neural network- (CNN-) based, simple recurrent neural network- (RNN-) based, long short-term memory- (LSTM-) based, and bidirectional LSTM- (BiLSTM-) based classifiers. To appraise the competency of suggested knowledge graph-enabled APP merely, some of the base and most well-known deep learning classification models, with maximally similar architectures and same configurations, were used. Classification in all Big Five traits was fulfilled concurrently. In fact, each model performs a multilabel binary classification, which assigns five labels to each of the OCEAN traits for a given text. Some common settings, which were applied in all suggested APP models, are presented in Table 1. Furthermore, as shown in Figure 5, the architecture of each model is composed of two stacked classifiers (like CNN), which leads to better results rather than single classifier. The classifiers are then followed by a batch normalization, to expedite the training and regularize the model. Next, applying a pooling layer as well as a dropout layer will help to avoid overfitting through providing an abstracted form of the representation. Finally, the models are followed by two consecutive dense layers to classify the extracted features from previous layers, change the dimensions of the vectors, and make possible the final prediction in the output layer.

(i) *Convolutional Neural Network- (CNN-) Based Classifier*. Convolutional neural networks, as a model with impressive

TABLE 1: Common parameters' settings among all of the proposed APP models (including CNN, RNN, LSTM, and BiLSTM).

Parameter	Setting	Parameter	Setting
Train-test split ratio (%)	80–20	Optimizer	Stochastic gradient descent (SGD)
Number of epochs	30	Learning rate	0.01
Early stopping	Applied on validation loss	Loss function	Binary_cross-entropy
Patience value	4	Batch size	32
Activation function	Sigmoid	Cross validation	10-Fold

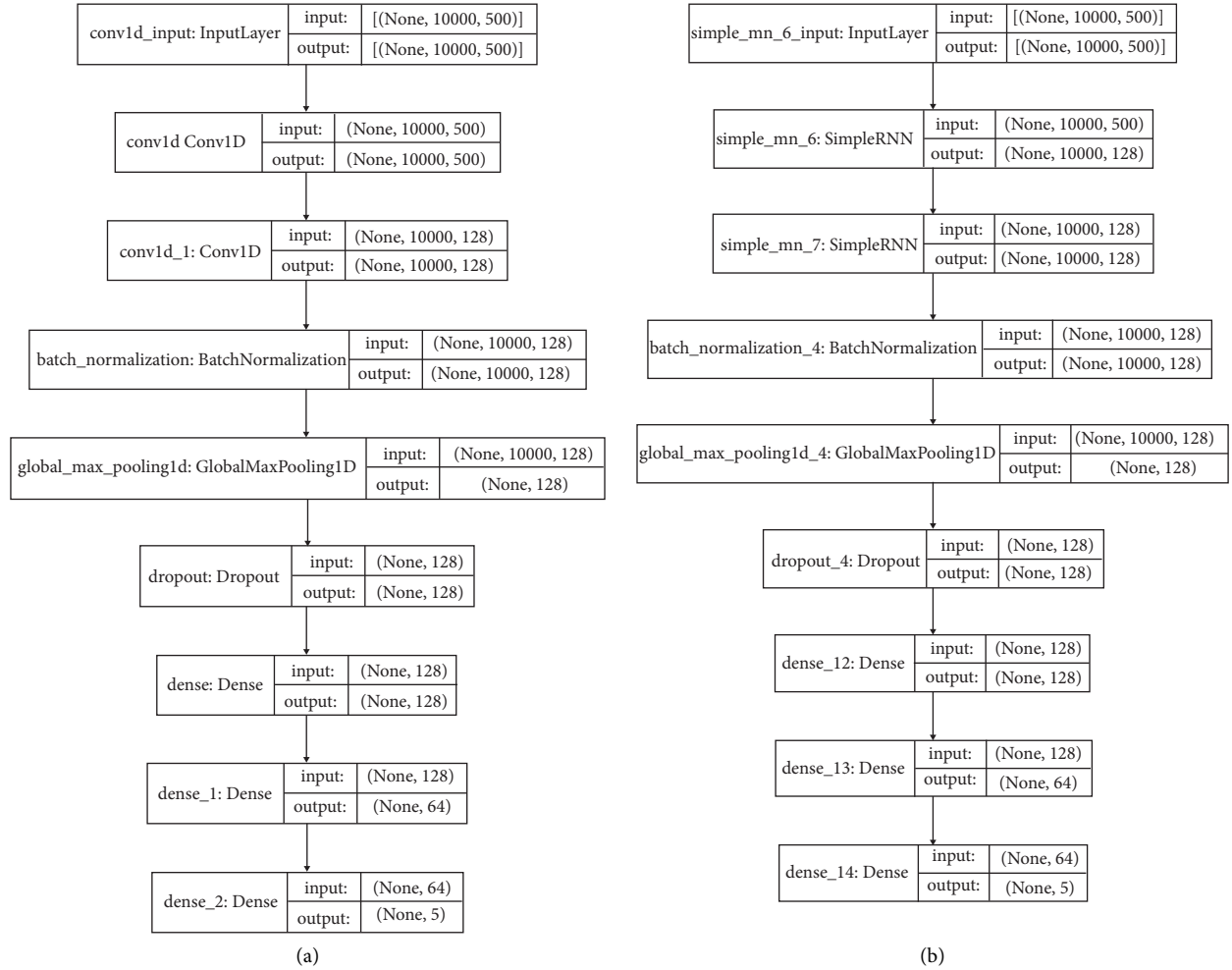


FIGURE 5: Continued.

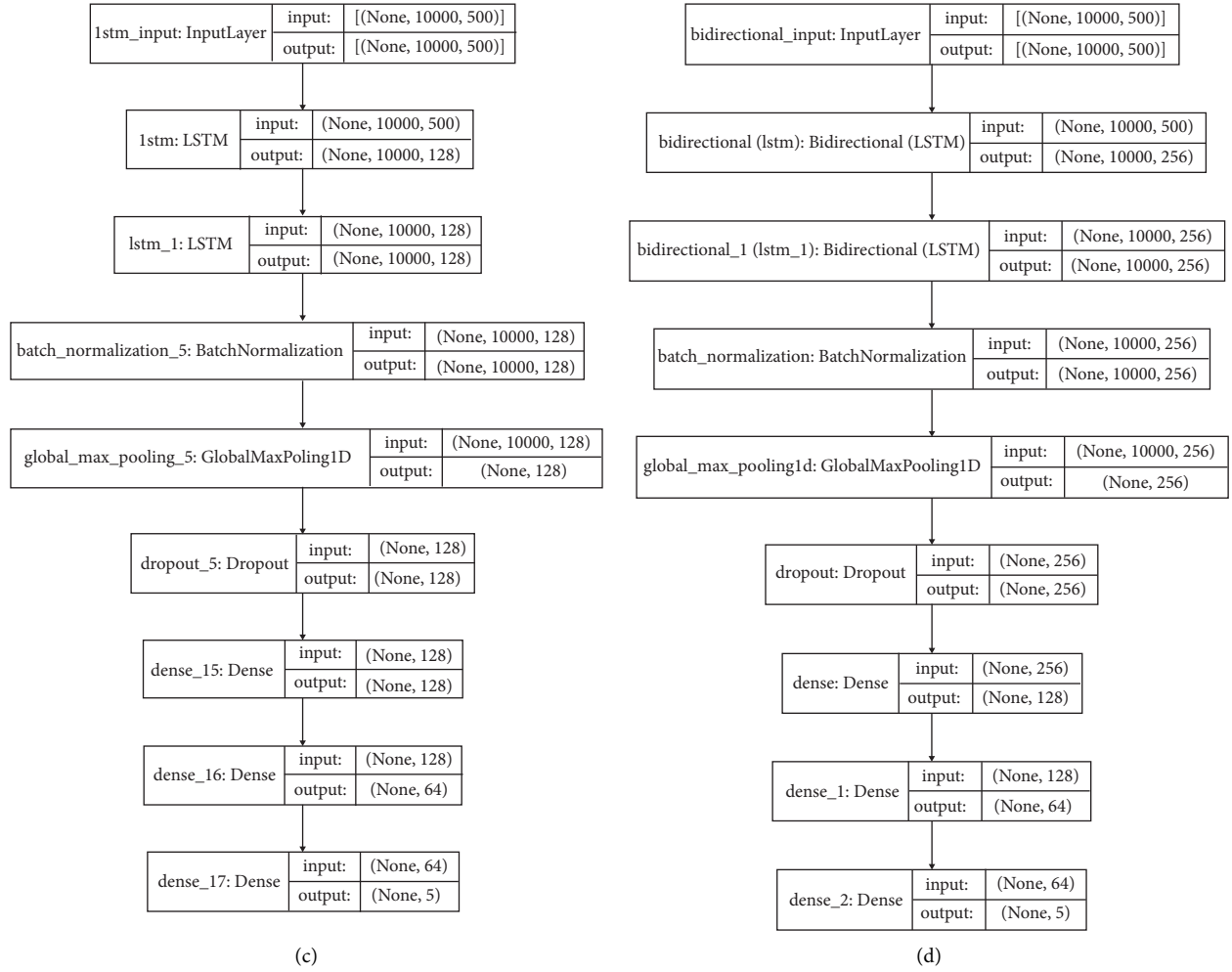


FIGURE 5: The Summaries of proposed APP classifiers. (a) CNN model. (b) RNN model. (c) LSTM model. (d) BiLSTM model.

performance, have been extensively investigated in various problems including visual recognition, speech recognition, and natural language processing [57]. To classify the resulted embedding matrices, a model with two one-dimensional convolutional layers followed by a batch normalization layer, a pooling layer, a dropout layer for regularization, and finally two fully connected layers was developed. In each convolutional layer, 128 parallel feature maps and a kernel size of 7, along with same padding, were applied. Figure 5(a) presents a summary of the model.

(ii) *Recurrent Neural Network- (RNN-) Based Classifier.* Recurrent neural networks have shown excellent dexterity in text classification tasks. The foundation of RNN [58] makes it possible to utilize previous step's outputs as inputs in current step. To rephrase it, while traditional neural networks deal with the inputs independently of one another, RNNs manipulate a set of previous inputs. Furthermore, the internal state of an RNN, which acts as memory, empowers it to learn from previous information and grants a privilege of processing the sequential inputs like text. The suggested simple RNN-based classifier encompasses two simple RNN layers followed by a batch normalization layer, a pooling

layer, a dropout layer for regularization, and finally two fully connected layers. A summary of the model is shown in Figure 5(b).

(iii) *Long Short-Term Memory- (LSTM-) based classifier.* Long short-term memory networks as a kind of RNNs are suggested to deal with learning long-term dependencies problem [58]. In simple terms, simple RNNs suffer one major drawback; they can not remember information for a long period of time, what is resolved capably by LSTMs. Two stacked LSTMs followed by a batch normalization layer, a pooling layer, a dropout layer for regularization, and finally two fully connected layers construct the design of proposed LSTM-based classification model. A plot of model is depicted in Figure 5(c).

(iv) *Bidirectional Long Short-Term Memory- (BiLSTM-) based classifier.* Indeed, as can be inferred, BiLSTMs are a bidirectional form of LSTMs. In simple words, LSTM is a unidirectional network, which utilizes previous information that has already passed through it in forward direction within sequence processing, while a BiLSTM network exploits both previous and future information in forward and

backward directions, respectively. Telling the truth, it consists of two LSTMs: one analyzes the input sequence from beginning to the end in forward direction, and the other one, from end to beginning in backward direction [59]. The final output is the concatenation of the two LSTMs. Two stacked BiLSTMs, followed by a batch normalization layer, a pooling layer, a dropout layer, and finally two fully connected layers, comprise the architecture of the proposed BiLSTM-based classifier. Figure 5(d) depicts a summary of the model.

At last, a brief outline of phase 3 as can be seen in Figure 4 is as follows:

- (i) Input: a set of embedding matrices;
- (ii) Output: predicted labels for OCEAN traits for each embedding matrices;
- (iii) Objective: personality prediction using multilabel classification model.

Algorithm 1 details a step-by-step flow of the proposed method that would assist towards a better comprehension of the method.

5. Results

5.1. Evaluation Metrics. Traditionally, classification models are evaluated through some well-known evaluation metrics including precision, recall, f-measure, and accuracy [39]. There are two determining sets, which play crucial role in their values, specifically the set of essays' "actual labels," which is sometimes referred to as gold standard and the set of "system predicted labels." Practically, for each prediction in a given class (namely, O, C, E, A, and N), there are four possible combinations of actual labels and system predicted labels, including:

- (i) True Positive (TP): that occurs when the actual label is true, and the system predicted label is also true;
- (ii) True Negative (TN): that occurs when the actual label is false, and the system predicted label is also false;
- (iii) False Positive (FP): that occurs when the actual label is false, while the system predicted label is true;
- (iv) False Negative (FN): that occurs when the actual label is true, while the system predicted label is false.

Essentially, in an APP system evaluation, the TP and TN play a leading part, due to the fact that, in such classification systems, it is prominent to truly predict that a given text really belongs or does not belong to the class. In both of TP and TN, the system predicted labels are equal to the actual labels; hence, it can be stated that the total number of TPs and TNs denotes the APP system's correct predictions. Consequently, the ratio of systems correct predictions to the total number of predictions reveals the quality of prediction; it is actually known as accuracy. That is to say, $accuracy = (TP + TN)/(TP + TN + FP + FN)$.

Moreover, the precision and recall as well as their weighted harmonic mean, which are called f-measure, convey some facts about the performance of classification

system. Precision (P) mainly concerns system's true labeled predictions. It reveals that the proportion of system's true labeled predictions has actual true labels. In other words, $P = TP/(TP + FP)$, while recall (R) mainly concerns true labels in gold standard. It tries to reveal that proportion of true labeled samples in gold standard has achieved true labels, after the system prediction. It means that $R = TP/(TP + FN)$.

Both of the precision and recall are unreliable metrics in classification systems' evaluation when they are considered separately. To put it another way, there may be some cases with high values of precision and low values of recall simultaneously, and vice versa. It is principally because of their partial coverage and incomplete reports. Hence, f-measure is suggested to address this problem. In fact, it makes a tradeoff between precision and recall and combines their included facts; precisely, $f - measure = (2 \times P \times R)/(P + R)$. However, it still suffers from a significant drawback. Actually, TN as a prominent factor in evaluation is completely neglected. As an illustration, it ignores all of the correctly false labeled samples by system. Thereby, accuracy is preferred to f-measure in APP system evaluation.

5.2. Evaluation Results. This study was undertaken to design a knowledge graph-enabled automatic personality prediction system and evaluate the efficacy of knowledge graph-enabling of a personality prediction system. Accordingly, a three-phase approach was proposed, which by receiving a text proceeds to carry out some preprocessing in the first phase and then build, enrich, and embed the corresponding knowledge graph consecutively in second phase, as it is completely scrutinized in Section 4.2.2. Figure 6 provides the results obtained from the second phase for a sample essay in Essays Dataset. Eventually, the resulting embedding matrix was classified through four independent classification models in third phase, and the predicted labels in each OCEAN traits were assigned. This section summarizes the findings and contributions made.

Specifically, there were four APP classification models suggested, namely, CNN-based, RNN-based, LSTM-based, and BiLSTM-based classifiers. However, accuracy outperforms precision, recall, and f-measure in APP systems' evaluation, and we will report the evaluation results for all of them. Albeit that, we will mainly rely on accuracy. Of course, in spite of the facts behind the precision, recall, and f-measure, availability of their values would be helpful when comparing those studies, which have just reported the evaluation results for them, rather than accuracy.

Table 2 provides the results obtained from the evaluation of four APP classifiers. Comparing accuracy values among four suggested classifiers, the most striking results were achieved through BiLSTM. Specifically, it had the most accurate predictions in all OCEAN traits compared to other classifiers. Hence, the first highest average accuracy in five traits was achieved by BiLSTM. In addition, the second highest average accuracy was attained by LSTM. However, comparing the accuracies in each trait individually reveals

- (1) **Foreach** essay $e_i \in$ Essays Dataset
- (2) Phase-1: Perform preprocessing activities, including:
 - (i) Tokenization
 - (ii) Noise removal (punctuations, signs, and stop words)
 - (iii) Normalization (lowercasing and lemmatization)
 - (iv) Named Entity Recognition (NER)
- (3) Phase-2: Perform knowledge representation for e_i , more specifically:
 - (i) Build the corresponding knowledge graph (KG) for e_i , through DBpedia
 - (ii) Enrich the acquired KG using DBpedia ontology, NRC, and MRC
 - (iii) Embed the acquired enriched KG using RDF2vec, and save it in **Embeddings set**
- (4) **End foreach**
- (5) Using the **Embeddings set**, train and test:
 - (i) A CNN-based multilabel binary classification model to perform APP
 - (ii) An RNN-based multilabel binary classification model to perform APP
 - (iii) An LSTM-based multilabel binary classification model to perform APP
 - (iv) A BiLSTM-based multilabel binary classification model to perform APP

ALGORITHM 1: Algorithm of the proposed method

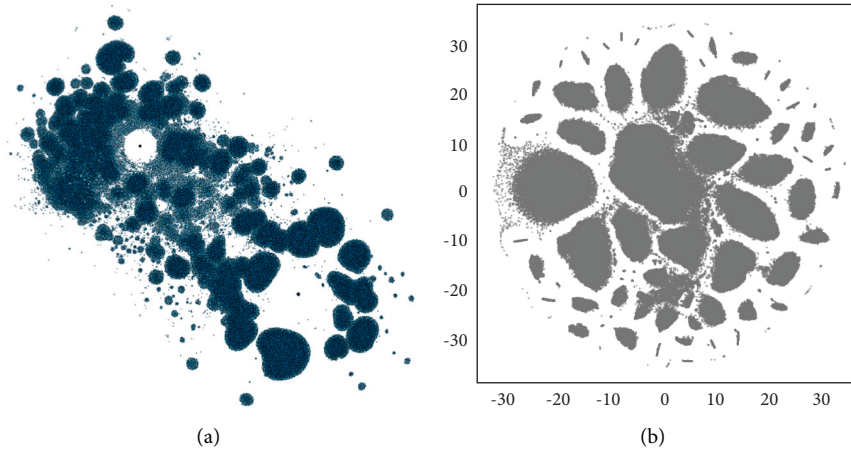


FIGURE 6: The resulting final knowledge graph from phase 2 for a sample essay (2004_139) in Essays Dataset (including 226,763 vertices and 532,146 edges). The edges and labels in knowledge graph are emitted for better visualization. (a) Knowledge graph's vertices (provided by Gephi, the ForceAtlas2 algorithm). (b) Embedded knowledge graph in 2D space.

TABLE 2: Evaluation results for suggested APP classifiers, including CNN-based, RNN-based, LSTM-based, and BiLSTM-based classifiers.

Metric	Classification model	O	C	E	A	N	Avg.
Precision	CNN	61.22	60.70	60.13	59.80	62.42	60.85
	RNN	62.41	62.80	64.49	61.63	55.52	61.37
	LSTM	66.56	67.65	63.79	67.58	59.62	65.04
	BiLSTM	69.12	71.43	73.05	67.75	62.69	68.81
Recall	CNN	82.68	79.72	78.36	74.37	78.48	78.72
	RNN	77.73	71.69	77.06	74.30	76.67	75.49
	LSTM	81.20	79.31	82.76	78.88	71.82	78.79
	BiLSTM	78.80	80.46	83.03	76.33	75.12	78.75
<i>F</i> -measure	CNN	70.35	68.92	68.05	66.29	69.53	68.63
	RNN	69.23	66.95	70.22	67.37	64.40	67.64
	LSTM	73.15	73.02	72.05	72.79	65.15	71.23
	BiLSTM	73.64	75.68	77.72	71.78	68.34	73.43
Accuracy	CNN	67.34	68.36	65.52	63.49	66.94	66.33
	RNN	69.17	68.56	69.37	68.76	63.90	67.95
	LSTM	69.78	68.97	69.57	69.98	65.72	68.44
	BiLSTM	71.40	72.62	73.83	70.18	69.37	71.48

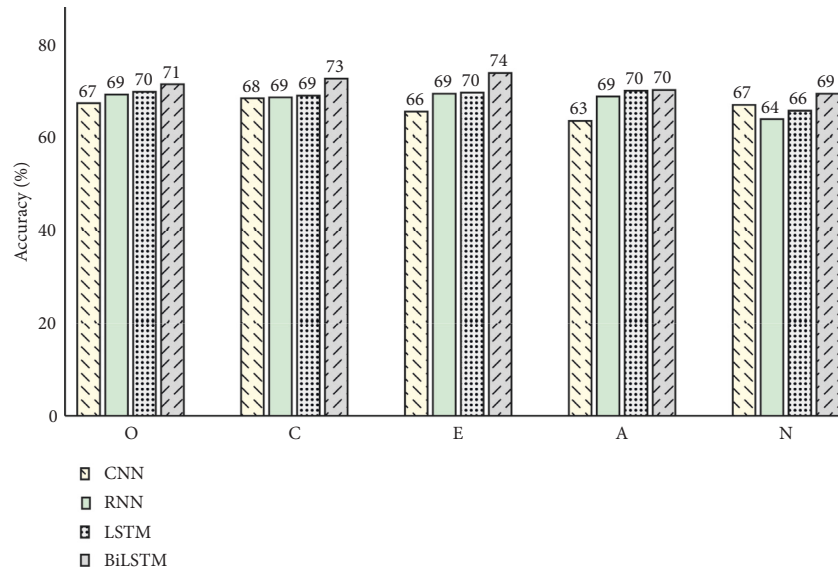


FIGURE 7: Accuracy values for four suggested APP classifiers, in each of the five personality traits in Big Five model (results are rounded).

that it had more accurate predictions in *O*, *C*, *E*, and *A* rather than RNN and CNN, while CNN in *N* practically had better predictions. However, LSTM concluded more accurate results rather than simple RNN in *N*. Afterwards, RNN outperformed CNN in all traits except *N*. Figure 7 compares the accuracy values in five personality traits resulting from four classification models.

As Table 2 shows, the same ranking as it happens when perusing the accuracy was achieved by classification models when taking the average f-measure into consideration. That is to say, BiLSTM, LSTM, RNN, and at last CNN were ranked first to fourth, respectively, although the ranks do not last when comparing the f-measure values individually in each trait. Regarding recall average values, LSTM with a slight difference to BiLSTM and CNN had better performance. At last, among the four suggested classifiers, BiLSTM, LSTM, RNN, and CNN had the most precise predictions, respectively, as it can be seen from Table 2.

6. Discussion

The major objective of current study was to investigate the efficacy of knowledge graph-enabled automatic personality prediction system. Thus, we used four simple base deep learning classifiers, which were designed maximally similar to each other. Furthermore, we intentionally avoided designing complex networks to merely appraise the efficacy of knowledge graph-enabling of an APP system. Accordingly, we suggested one CNN-based classifier as well as three recurrent classifiers, in particular, a simple RNN-based classifier along with one LSTM-based and one bidirectional LSTM- (BiLSTM-) based classifiers.

Regarding the resulting accuracy values in Table 2 for each of the classifiers, it is clear that generally recurrent classifiers lead to better results, rather than CNN. It seems possible that these results are due to their ability in

processing temporal information that is presented in input sequences. To put it simply, recurrent networks are basically designed for sequence prediction problems like text. More specifically, they can capture sequential information, which pinpoints the existing dependencies among the words throughout the input sequence of words.

Among the three suggested recurrent networks, superior results are seen for BiLSTM. In fact, it outperforms LSTM and simple RNN, in which this does seem to be because it is actually an enhanced version of LSTM, in which it itself is an enhanced version of simple RNN. To rephrase it, LSTMs were proposed to tackle RNNs' problem in preserving information over several timesteps; and BiLSTMs were also proposed to tackle LSTMs' problem in ignoring future information for a given word in input sequence. As a consequence, it is fair that BiLSTMs show better results than LSTMs and simple RNNs, and LSTMs show better results than simple RNNs. This is what happened in all personality traits, though with slight differences in some traits. So, the obtained results confirm the expectations.

Besides, CNN-based classifier leads to comparable results. Despite the fact that it was ranked fourth among four classifiers, its results are so close to some recurrent classifiers in *C* and even it outperforms simple RNN and LSTM-based classifiers in *N*. We speculate that this might be due to its filters' good ability in feature extraction from input embedding matrices.

The results obtained from the four proposed classifiers can be compared with the state-of-the-art APP systems, which were performed on Essays Dataset in Table 3. These results go beyond previous contributions, showing that all of the suggested methods give clearly better results than all of them. On the other hand, whereas our first ranked proposed method considerably yields better results, the fourth ranked proposed method also outperforms that of

TABLE 3: Comparing the results obtained from our proposed methods and state-of-the-art reports in APP from text, which were performed on Essays Dataset.

APP method	F-measure						Accuracy					
	O	C	E	A	N	Avg.	O	C	E	A	N	Avg.
CNN-based classifier (proposed method)	70.35	68.92	68.05	66.29	69.53	68.63	67.34	68.36	65.52	63.49	66.94	66.33
RNN-based classifier (proposed method)	69.23	66.95	70.22	67.37	64.40	67.64	69.17	68.56	69.37	68.76	63.90	67.95
LSTM-based classifier (proposed method)	73.15	73.02	72.05	72.79	65.15	71.23	69.78	68.97	69.57	69.98	65.72	68.44
BiLSTM-based classifier (proposed method)	73.64	75.68	77.72	71.78	68.34	73.43	71.40	72.62	73.83	70.18	69.37	71.48
Majumder et al. [6]							62.68	57.30	58.09	56.71	59.38	58.83
Yuan et al. [9]							62.00	57.00	58.00	56.00	59.00	58.40
Ramezani et al. [24]	57.37	59.74	65.80	61.62	60.69	61.04	56.30	59.18	64.25	60.31	61.14	60.24
Xue et al. [28]	67.84	63.46	71.50	71.92	62.36	67.42	63.16	57.49	58.91	57.49	59.51	59.31
El-Demerdash et al. [30]							63.30	57.97	58.85	59.25	59.88	59.85
Jiang et al. [31]							65.86	58.55	60.62	59.72	61.04	61.16
El-Demerdash et al. [13]							65.60	59.52	61.15	60.80	62.20	61.85
Kazameini et al. [34]							62.09	57.84	59.30	56.52	59.39	59.03
Wang et al. [60]	67.00	68.00	67.00	69.00	69.00	68.00	64.80	59.10	60.00	57.70	63.00	60.92
Tighe et al. [61]	61.90	56.00	55.60	55.70	58.30	57.50	61.95	56.04	55.75	57.54	58.31	57.92

previous reports. This is an important finding in the understanding of the knowledge graph-enabling of an automatic personality prediction system. Moreover, it is anticipated that utilization of more complex classification models (like hybrid models) would lead to more accurate predictions.

Ultimately, we are going to answer the research questions (as stated in Introduction) according to our observations as follows:

RQ.1: The results of the experiment found clear support for knowledge graph-enabling of an APP system. Actually, it empowers an APP system to yield considerably more accurate results. It is also worth noting that, in this study, we have just utilized the embeddings of resulting knowledge graphs to perform personality predictions, while the knowledge graphs inherently comprise miscellaneous knowledges of concepts, which may be effectively utilizable in automatic personality prediction.

RQ.2: The most interesting finding was that, in classification of knowledge graphs' embedding matrices, all of the proposed deep learning classifiers, namely, CNN-based, RNN-based, LSTM-based, and BiLSTM-based classifiers, substantially outperform the state-of-the-art contributions in APP, in spite of the models' simple design. This is obviously confirmed when comparing our results to those of older studies. Besides, experimental observations demonstrated that the classifiers, which are based on BiLSTM, LSTM, simple RNN, and CNN, yield better results, respectively, when they were utilized in classification of knowledge graphs' embeddings.

RQ.3: Regarding the obtained results from several classifiers, it is clear that knowledge graph-enabling of an APP system totally enhances the number of accurate predictions in all personality traits of Big Five model, albeit the enhancement pattern is not similar in all of the classifiers.

7. Conclusion

The current study aimed to determine the effect of knowledge graph-enabling on an automatic personality prediction system. To do so, a three-phase approach was proposed, in which a given text performs some pre-processing (including tokenization, noise removal, normalization, and named entity recognition) in its first phase. The second phase is aimed toward achieving a knowledgeable representation of input text, trying to build the corresponding knowledge graph, then enriching it (utilizing DBpedia ontology, NRC Emotion Intensity Lexicon, and MRC psycholinguistic database), and finally embedding the enriched knowledge graph. At last, in the third phase, the embedding knowledge graph is fed into some base deep learning models (namely, CNN-based, simple RNN-based, LSTM-based, and BiLSTM-based classifiers) to perform personality prediction. The results demonstrate a strong effect of knowledge graph-enabling on an automatic personality prediction system. More specifically, the findings definitely confirmed the proposed method's ability to predict all five personality traits of the Big Five model.

As the greatest practical significance of this study, it provides the basis of human-like behavior for machines in a specific task, namely, automatic personality prediction. Since human intelligent behavior is a consequence of his/her cognitive abilities, in which it is an outcome of representation of the knowledge of the world's concepts; therefore, our method will help machines mimic human behavior as it is, which is a big step forward. That is to say, providing a comprehensive representation of appearing concepts in the input text models the human cognition for the machine, which enables it to show human-like performance. The obtained results, as well as the comparison of the findings with those of other studies, confirmed this claim.

In future work, we intend to investigate more complex deep learning models, to achieve more accurate predictions. The current study has only examined the efficacy of a

knowledge graph-enabled automatic personality prediction system, and hence to minimize the effect of extrinsic factors as far as possible, it just relied on simple base deep learning models. As well, since the resulted knowledge graph usually is very large, more research is also needed to find a way to cope with it. Besides, further experimental investigations are needed to peruse other graph embedding methods and determine their effectiveness. Moreover, the application of the suggested method over different datasets in different personality models could shed more light on the efficacy of the proposed method. More broadly, the proposed knowledge representation method potentially is capable of performing other tasks, which deal with text, since it provides a more knowledgeable representation of text elements for machines. Hence, the issue of knowledge representation is an intriguing one, which could be usefully explored in several researches.

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

No funds, grants, or other support was received.

References

- [1] R. M. Bergner, "What is personality? two myths and a definition," *New Ideas in Psychology*, vol. 57, Article ID 100759, 2020.
- [2] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of Artificial Intelligence Research*, vol. 30, pp. 457–500, 2007.
- [3] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from twitter," in *Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pp. 149–156, Boston, MA, USA, October 2011.
- [4] C. Sumner, A. Byers, R. Boochever, and J. Gregory, "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets," in *Proceedings of the 2012 11th International Conference on Machine Learning and Applications*, pp. 386–393, Boca Raton, Florida, USA, December 2012.
- [5] Y. Yuan, B. Li, D. Jiao, and T. Zhu, "The personality analysis of characters in vernacular novels by sc-liwc," in *Human Centered Computing*, Q. Zu and Bo Hu, Eds., pp. 400–409, Springer International Publishing, New York, NY, USA, 2018.
- [6] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.
- [7] D. Sewwandi, K. Perera, S. Sandaruwan, O. Lakchani, A. Nugaliyadde, and S. Thelijagoda, "Linguistic features based personality recognition using social media data," in *Proceedings of the 2017 6th National Conference on Technology and Management (NCTM)*, pp. 63–68, Malabe, Sri Lanka, January 2017.
- [8] B. B. C. Da Silva and I. Paraboni, "Personality recognition from facebook text," in *Computational Processing of the Portuguese Language*, A. Villavicencio, V. Moreira, A. Abad et al., Eds., pp. 107–114, Springer International Publishing, New York, NY, USA, pp. 107–114, 2018.
- [9] C. Yuan, J. Wu, L. Hong, and L. Wang, "Personality recognition based on user generated content," in *Proceedings of the 2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6, Hangzhou, China, July 2018.
- [10] J. Yu and K. Markov, "Deep learning based personality recognition from facebook status updates," in *Proceedings of the 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, pp. 383–387, Taichung, China, November 2017.
- [11] X. Sun, B. Liu, J. Cao, J. Luo, and X. Shen, "Who am i? personality detection based on deep learning for texts," in *Proceedings of the 2018 IEEE International Conference on Communications (ICC)*, pp. 1–6, Kansas City, MO, USA, May 2018.
- [12] T. Yilmaz, A. Ergil, and B. İlgen, "Deep learning-based document modeling for personality detection from Turkish texts," in *Proceedings of the Future Technologies Conference (FTC) 2019*, K. Arai, R. Bhatia, and S. Kapoor, Eds., pp. 729–736, Springer International Publishing, New York, NY, USA, 2020.
- [13] K. El-Demerdash, R. A. El-Khoribi, A. Mahmoud, S. Ismail, and S. Abdou, "Deep learning based fusion strategies for personality prediction," *Egyptian Informatics Journal*, vol. 23, 2021.
- [14] Z. Ren, Q. Shen, X. Diao, and H. Xu, "A sentiment-aware deep learning approach for personality detection from text," *Information Processing & Management*, vol. 58, no. 3, Article ID 102532, 2021.
- [15] D. Xue, L. Wu, Z. Hong et al., "Deep learning-based personality recognition from text posts of online social networks," *Applied Intelligence*, vol. 48, no. 11, pp. 4232–4246, 2018.
- [16] N. H. Jeremy and D. Suhartono, "Automatic personality prediction from Indonesian user on twitter using word embedding and neural networks," *Procedia Computer Science*, vol. 179, pp. 416–422, 2020.
- [17] H. Christian, D. Suhartono, A. Chowanda, and K. Z. Zamli, "Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging," *Journal of Big Data*, vol. 8, no. 1, p. 68, 2021.
- [18] M. K. Bergman, M. K. Bergman, and L. Fife, *Knowledge Representation Practionary*, Springer, New York, NY, USA, 2018.
- [19] F. Van Harmelen, V. Lifschitz, and B. Porter, *Handbook of Knowledge Representation*, Elsevier, Amsterdam, Netherlands, 2008.
- [20] S. Matz, W. F. Chan, and M. Kosinski, *Models of Personality*, pp. 35–54, Springer International Publishing, New York, NY, USA, 2016.
- [21] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, vol. 60, no. 2, pp. 175–215, 1992.
- [22] J. D. Moreno, J. Á Martínez-Huertas, R. Olmos, G. Jorge-Botana, and J. Botella, "Can personality traits be measured analyzing written language? a meta-analytic study on

- computational methods,” *Personality and Individual Differences*, vol. 177, Article ID 110818, 2021.
- [23] J. A. Cummings and L. Sanders, *Introduction to Psychology*, University of Saskatchewan Open Press, Saskatchewan, 2019.
- [24] M. Ramezani, M. R. Feizi-Derakhshi, M. A. Balafar et al., “Automatic Personality Prediction; an Enhanced Method Using Ensemble Modeling,” *Neural Computing and Applications*, 2022.
- [25] A.-R. Feizi-Derakhshi, M.-R. Feizi-Derakhshi, M. Ramezani et al., “The State-Of-The-Art in Text-Based Automatic Personality Prediction,” 2021, <https://arxiv.org/abs/2110.01186>.
- [26] S. Han, H. Huang, and Y. Tang, “Knowledge of words: an interpretable approach for personality recognition from social media,” *Knowledge-Based Systems*, vol. 194, Article ID 105550, 2020.
- [27] J. W. Pennebaker, M. E. Francis, and R. J. Booth, “Linguistic inquiry and word count: Liwc 2001,” *Mahway: Lawrence Erlbaum Associates*, vol. 71, 2001.
- [28] X. Xue, J. Feng, and X. Sun, “Semantic-enhanced sequential modeling for personality trait recognition from texts,” *Applied Intelligence*, vol. 51, pp. 1–13, 2021.
- [29] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, and S. Eetemadi, “Bottom-up and top-down: predicting personality with psycholinguistic and language model features,” in *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*, pp. 1184–1189, Sorrento, Italy, November 2020.
- [30] Kamal El-Demerdash, R. A. El-Khoribi, A. Mahmoud, S. Ismail, and S. Abdou, “Psychological human traits detection based on universal language modeling,” *Egyptian Informatics Journal*, 2020.
- [31] H. Jiang, X. Zhang, and J. D. Choi, “Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings (student abstract),” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 10, Article ID 13821, 2020.
- [32] A. Kunte and S. Panicker, “Personality prediction of social network users using ensemble and xgboost,” in *Progress in Computing, Analytics and Networking*, H. Das, P. Kumar Pattanaik, S. S. Rautaray, and K.-C. Li, Eds., pp. 133–140, Springer, Singapore, 2020.
- [33] A. Sood and R. Bhatia, “Baron-cohen model based personality classification using ensemble learning,” in *Data Science and Big Data Analytics*, D. K. Mishra, X.-S. Yang, and A. Unal, Eds., pp. 57–65, Springer, Singapore, 2019.
- [34] A. Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria, “Personality Trait Detection Using Bagged Svm over Bert Word Embedding Ensembles,” 2020, <https://arxiv.org/abs/2010.01309>.
- [35] X. Sun, B. Liu, Q. Meng, J. Cao, J. Luo, and H. Yin, “Group-level personality detection based on text generated networks,” *World Wide Web*, vol. 23, no. 3, pp. 1887–1906, 2019.
- [36] Z. Guan, B. Wu, B. Wang, and H. Liu, “Personality2vec: network representation learning for personality,” in *Proceedings of the 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*, pp. 30–37, July 2020.
- [37] J. W. Pennebaker and L. A. King, “Linguistic styles: language use as an individual difference,” *Journal of Personality and Social Psychology*, vol. 77, no. 6, pp. 1296–1312, 1999.
- [38] A. Lex, N. Gehlenborg, H. Strobel, R. Vuillemot, and H. Pfister, “Upset: visualization of intersecting sets,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1983–1992, 2014.
- [39] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*, Cambridge University Press, Cambridge, 2008.
- [40] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O’Reilly Media, Inc, Sebastopol, California, 2009.
- [41] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, “SpaCy: Industrial-Strength Natural Language Processing in Python,” 2020.
- [42] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, “Dbpedia: a nucleus for a web of open data,” in *The Semantic Web*, M. Peter, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, Eds., pp. 722–735, Springer, Berlin, Heidelberg, 2007.
- [43] X. Chen, S. Jia, and Y. Xiang, “A review: knowledge reasoning over knowledge graph,” *Expert Systems with Applications*, vol. 141, Article ID 112948, 2020.
- [44] Z. Chen, Y. Wang, B. Zhao, J. Cheng, X. Zhao, and Z. Duan, “Knowledge Graph Completion: A Review,” *IEEE Access*, vol. 8, 2020.
- [45] D. Vrandečić, J. Aasman, and M. Galkin, *What Is a Knowledge Graph?*, Stanford University, Stanford, California, 2020.
- [46] M. Chein and M.-L. Mugnier, *Graph-based Knowledge Representation: Computational Foundations of Conceptual Graphs*, Springer Science & Business Media, Berlin, Heidelberg, 2008.
- [47] M. Kejrival, C. A. Knoblock, and S. Pedro, *Knowledge Graphs: Fundamentals, Techniques, and Applications*, MIT Press, Cambridge, MA, USA, 2021.
- [48] H. Arnaout and S. Elbassuoni, “Effective searching of rdf knowledge graphs,” *Journal of Web Semantics*, vol. 48, pp. 66–84, 2018.
- [49] A. Hogan, “SPARQL query language,” *The Web of Data*, Springer International Publishing, pp. 323–448, New York, NY, USA, pp. 323–448, 2020.
- [50] G. Jakus, V. Milutinović, S. Omerović, and S. Tomažič, *Concepts, Ontologies, and Knowledge Representation*, Springer, New York, NY, USA, 2013.
- [51] Merriam-Webster, “Ontology,” in *Merriam-Webster.com Dictionary*, Springer, New York, NY, USA, 2022.
- [52] L. L. C. MultiMedia, *Dbpedia Ontology Classes*, Springer London, London, 2021.
- [53] S. M. Mohammad, “Word affect intensities,” in *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan, October, 2018.
- [54] M. Wilson, “Mrc psycholinguistic database: machine-useable dictionary, version 2.00,” *Behavior Research Methods Instruments & Computers*, vol. 20, no. 1, pp. 6–10, 1988.
- [55] P. Ristoski, J. Rosati, T. Di Noia, R. De Leone, and H. Paulheim, “Rdf2vec: rdf graph embeddings and their applications,” *Semantic Web*, vol. 10, no. 4, pp. 721–752, 2019.
- [56] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” 2013, <https://arxiv.org/abs/1301.3781>.
- [57] J. Gu, Z. Wang, J. Kuen et al., “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [58] A. Sherstinsky, “Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network,” *Physica D: Nonlinear Phenomena*, vol. 404, Article ID 132306, 2020.

- [59] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [60] Z. Wang, C.-H. Wu, Q.-B. Li, B. Yan, and K.-F. Zheng, "Encoding text information with graph convolutional networks for personality recognition," *Applied Sciences*, vol. 10, no. 12, Article ID 4081, 2020.
- [61] E. P. Tighe, J. C. Ureta, B. A. Pollo, C. K. Cheng, and R. D. Bulos, "Personality trait classification of essays with the application of feature reduction," in *Proceedings of the SAAIP @ IJCAI*, pp. 22–28, New York, USA, July 2016.

Research Article

Optimal Control and Stability Analysis of an SEIR Model with Infectious Force in Latent Period

Li Jiayi , Li Sixian, Shi Weixuan, Hu Manfeng, and Zhang Jingxiang 

School of Science, Jiangnan University, Wuxi, Jiangsu 214122, China

Correspondence should be addressed to Zhang Jingxiang; zhangjingxiang@jiangnan.edu.cn

Received 21 April 2022; Revised 19 May 2022; Accepted 24 May 2022; Published 15 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Li Jiayi 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, an SEWIR epidemic model with the government control rate and infectious force in latent period is proposed. The conditions to the existence and uniqueness of disease-free and endemic equilibrium points in the SEWIR model are obtained. By using the Hurwitz criterion, the locally asymptotic stability of disease-free and endemic equilibrium points is proved. We show the global asymptotic stability of the disease-free equilibrium point by the construction of Lyapunov function and LaSalle invariance principle. The globally asymptotic stability of the endemic equilibrium is verified by numerical simulation. Several optimal control strategies are proposed on controlling infectious diseases.

1. Introduction

Since the outbreak of COVID-19, the global economy and social stability have been severely affected. Subsequently, the control strategies of COVID-19 infection have become the focus of research. SEIR is a commonly used model in epidemiology. It splits the entire group into four compartments: susceptible ones, exposed ones, infected ones, and recovered ones. Early in the research, many researchers considered fewer factors, and most of them directly used SEIR model to study and analyze epidemic diseases. With the deepening of research, some mathematicians found that there are more factors which affect the prediction and control of epidemic diseases, and even there are coupled dynamic changes among the factors. In the process of analyzing the transmission paths and control strategies of epidemic diseases, they have constructed lots of models for different situations to formulate the optimal strategies. Almeida et al. [1] studied an SEIR epidemic model which splits the recovery rate of infected population into two categories: without and with medical treatment. The numerical change curve of the four groups was simulated under the condition of controlling the economic cost. Carcione et al. [2] investigated an SEIR epidemic model similar to the one described in [1]. They

replicated the infection and death curves by altering the four populations' baseline values, the transition rate from exposure to infection, and the recovery rate of infectious population. The importance and effectiveness of isolation and medical level in stopping the spread of the virus were verified. Khan et al. [3] considered nonlinear morbidity with a saturation constant and introduced the susceptibility for recovered individuals and medical control function into the SEIR model. It can be concluded that the optimal method to control the disease is the proper use of treatment. Several studies have demonstrated that boosting the quality of medical care during an outbreak of epidemic significantly increases the rate of recovery; however, it has no effect on limiting the transmission of the infection. Considering the factors affecting the outbreak, one of the most effective and cost-efficient control strategies, such as centralized or home isolation, are frequently implemented. Auger and Mousaoui [4] studied an SEIR epidemic model with three scenarios: individual residences, workplaces, and high-density public places based on the law of population density changing with time. Simultaneous, improving immunity through vaccination can be held accountable as an effective strategy for controlling epidemic diseases. In infectious disease dynamics, there exists two categories: continuous

vaccination [5–10] and pulse vaccination [11–13]. Sen et al. [5] constructed an SIR model with constant vaccination control, and a vaccination control method was proposed. In [6], the authors studied an SEIR model including asymptomatic and dead-infectious subgroups. And, feedback vaccination as well as antiviral treatment control measures was included in the model. Nistal et al. [7] presented a discrete SEIADR model by considering diseases where infected corpses are still infectious. They introduced two types of vaccination. It was indicated that susceptible populations decrease, while recoveries increase under constant vaccination. Jiao and Shen [9] studied a more general SEIR model by introducing disease transmission rate with seasonal forcing and the rate of losing immunity of recovered population. They proposed and evaluated some control strategies by adjusting the transfer rate and simulating the population change curve under different isolation measures. However, it is practically difficult to achieve universal continuous vaccination. Hence, pulse vaccination was introduced to prevent diseases by regular vaccination. The effectiveness of this preventive measure has been proved successively [11–13].

Nowadays, most countries have implemented control measures such as vaccination, medical testing, isolation of foreign populations, close contacts and subclose contacts of diagnosed cases, and medical treatment. As a general rule, the strength of the government to implement epidemic control measures are closely related to the spread and control of epidemic. To further investigate the impact of government intervention and vaccination on the transmission kinetics, we construct a government intervention model. Additionally, the model is improved based on the traditional SEIR model under the measures taken to centrally quarantine populations tested positive after regional outbreaks in China. Given that the exposed population is still infectious, if a portion of the exposed population is medically quarantined in time, it will greatly limit the infectivity of the virus. Thus, the probability of further spread of the epidemic will be reduced, thus allowing for rapid identification and disposal and cutting off the transmission chain. Based on the above, we divide the exposed parts into two compartments: unquarantined and quarantined. Also, we introduce the government control rate and construct an SEWIR epidemic disease model with vaccination and government control.

In this paper, the whole population $N(t)$ can be split into five divisions, designated by $S(t)$, $E(t)$, $W(t)$, $I(t)$, and $R(t)$. $S(t)$ is the quantity of susceptible individuals who have no immunity to the COVID-19 virus. $E(t)$ denotes the quantity of unquarantined exposed individuals. $W(t)$ stands for the quantity of quarantined exposed individuals. $I(t)$ denotes the quantity of infected individuals who exhibit symptoms and are capable of propagating the disease. $R(t)$ represents the quantity of recovered individuals with normal medical test and immunity to COVID-19 virus. Furthermore, we obtain $N(t) = S(t) + E(t) + W(t) + I(t) + R(t)$. In Figure 1, the SEWIR model's transmission mechanism is depicted.

Based on the transmission mechanism, the SEWIR epidemic model with vaccination and government control is constructed with the form:

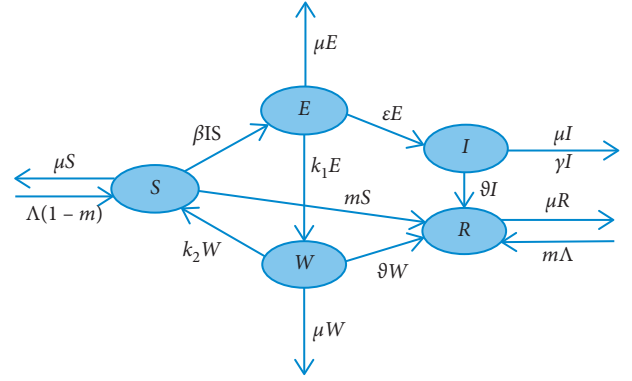


FIGURE 1: The transmission mechanism figure of the VGC-SEWIR model.

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -\beta I(t)S(t) + \Lambda(1-m) + k_2 W(t) - mS(t) - \mu S(t), \\ \frac{dE}{dt} = \beta I(t)S(t) - k_1 E(t) - \varepsilon E(t) - \mu E(t), \\ \frac{dW}{dt} = k_1 E(t) - \vartheta W(t) - k_2 W(t) - \mu W(t), \\ \frac{dI}{dt} = \varepsilon E(t) - \vartheta I(t) - \gamma I(t) - \mu I(t), \\ \frac{dR}{dt} = \vartheta W(t) + \vartheta I(t) + m(\Lambda + S(t)) - \mu R(t), \end{array} \right. \quad (1)$$

where β denotes the ratio of infected individuals infecting susceptible ones, m denotes the vaccination success rate of the susceptible individuals which represents the probability of getting immunity after vaccination, ε denotes the conversion rates from unquarantined exposed population to infected populations, k_1 denotes the government control rate, ϑ denotes the recovery rate, k_2 denotes the autoimmune virus rate, Λ denotes the population replenishment rate including foreign population and newborns, μ is the natural mortality, and γ represents the disease mortality.

The overall structure of this paper is listed as follows: Section 2 recalls some related analysis tools. Section 3 addresses the existence and uniqueness of equilibrium points for (2). Section 4 analyses the stability of (2) about equilibrium points. Section 5 simulates the relevant results and analyze the impact of vaccination and government control on infectious diseases. Section 6 proposes some optimal control strategies for infectious illnesses.

2. Preliminary

In this section, we briefly recall some related analysis tools. For more details, the interested reader may refer to [14–16].

Lemma 1 (Routh–Hurwitz theorem [14]). *Consider the characteristic equation*

$$\lambda^n + b_1\lambda^{n-1} + b_2\lambda^{n-2} + \dots + b_{n-1}\lambda + b_n = 0. \quad (2)$$

There exist negative real parts for all λ in case

$$\Phi_k = \begin{vmatrix} b_1 & b_3 & b_5 & \dots & b_{2k-1} \\ 1 & b_2 & b_4 & \dots & \vdots \\ 0 & b_1 & b_3 & \dots & \vdots \\ 0 & 1 & b_2 & \dots & b_{2k-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & b_k \end{vmatrix} > 0. \quad (3)$$

If $k > n$, then $b_k = 0$, where $k = 1, 2, 3, \dots, n$.

Lemma 2 (Lyapunov's stability theorem [15]). *Let V be a continuous differentiable positive definite function. If it enables \dot{V} to be negative semidefinite, then the origin is stable; if it enables \dot{V} to be negative definite, then the origin is asymptotically stable. When the condition of asymptotic stability holds globally as well, while V is radially unbounded, then the global asymptotic stability of the origin can be obtained.*

Lemma 3 (LaSalle invariance principle [16]). *Given a system $\dot{x} = f(x)$, in which f is continuous. Let $\Omega \subset D$ be a positive invariant tight set of the system. Let $V: D \rightarrow \mathbb{R}$ be a continuous differentiable function and satisfy positive definiteness. M is defined as the largest invariant set within $E: = \{x | \dot{V}(x) = 0\}$. If the system $\dot{x} = f(x)$ starts in Ω , then with $t \rightarrow \infty$, the system is bound to converge to M .*

3. Existence and Uniqueness of Equilibrium Points

Since the equation for $R(t)$ is independent from other equations, then the dynamics of model (1) is qualitatively equivalent to the dynamics of model given by

$$\begin{cases} \frac{dS}{dt} = -\beta I(t)S(t) + \Lambda(1-m) + k_2W(t) - mS(t) - \mu S(t), \\ \frac{dE}{dt} = \beta I(t)S(t) - k_1E(t) - \varepsilon E(t) - \mu E(t), \\ \frac{dW}{dt} = k_1E(t) - \vartheta W(t) - k_2W(t) - \mu W(t), \\ \frac{dI}{dt} = \varepsilon E(t) - \vartheta I(t) - \gamma I(t) - \mu I(t). \end{cases} \quad (4)$$

The unknown $R(t)$ can be determined correspondingly from

$$\frac{dR}{dt} = \vartheta W + \vartheta I + m(\Lambda + S) - \mu R. \quad (5)$$

It follows from (4) that

$$\begin{aligned} \frac{dS}{dt} + \frac{dE}{dt} + \frac{dW}{dt} + \frac{dI}{dt} &= \Lambda(1-m) - \vartheta(W+I) \\ &\quad - \gamma I - mS - \mu(S+E+W+I) \\ &\leq \Lambda(1-m) - \mu(S+E+W+I). \end{aligned} \quad (6)$$

Furthermore, we have

$$\limsup_{t \rightarrow \infty} (S(t) + E(t) + W(t) + I(t)) \leq \frac{\Lambda(1-m)}{\mu}. \quad (7)$$

Considering the biological significance of the model, the dynamic properties of model (2) are only discussed in the closed set Ω , which is defined by

$$\Omega = \left\{ (S, E, W, I) \mid 0 \leq S + E + W + I \leq \frac{\Lambda(1-m)}{\mu}, \right. \\ \left. S \geq 0, E \geq 0, W \geq 0, I \geq 0 \right\}. \quad (8)$$

It can be shown that Ω is positive and invariant. Obviously, there exists a disease-free equilibrium point $P_0(0, 0, S_0, 0) \in \partial\Omega$ with $S_0 = \Lambda(1-m)/m + \mu$ for model (2). Furthermore, the basic reproduction number R_0 will be applied to the proof of the existence and uniqueness of equilibriums. Relevant methods are presented in [17, 18].

Let us set $\mathbf{x} = (E, I, S, W)^T$. Then, model (2) becomes

$$\frac{d\mathbf{x}}{dt} = \Gamma(\mathbf{x}) - \Psi(\mathbf{x}), \quad (9)$$

with

$$\begin{aligned} \Gamma(\mathbf{x}) &= (\beta IS \ 000)^T, \\ \Psi(\mathbf{x}) &= (\mu + k_1 + \varepsilon)E, \\ &\quad (\mu + \vartheta + \gamma)I - \varepsilon E, \\ &\quad \beta IS + (\mu + m)S - k_2W - \Lambda(1-m), \\ &\quad (\mu + \vartheta + k_2)W - k_1E. \end{aligned} \quad (10)$$

The Jacobian matrix of $\Gamma(\mathbf{x})$ and $\Psi(\mathbf{x})$, when evaluated at $P_0(0, 0, S_0, 0)$, is as follows:

$$\begin{aligned} D\Gamma(P_0) &= \begin{pmatrix} F_{2 \times 2} & O_{2 \times 2} \\ O_{2 \times 2} & O_{2 \times 2} \end{pmatrix}_{4 \times 4}, \\ D\Psi(P_0) &= \begin{pmatrix} \mu + k_1 + \varepsilon & 0 & 0 & 0 \\ -\varepsilon & \mu + \vartheta + \gamma & 0 & 0 \\ 0 & \beta S_0 & \mu + m & -k_2 \\ -k_1 & 0 & 0 & \mu + \vartheta + k_2 \end{pmatrix}, \end{aligned} \quad (11)$$

where $O_{2 \times 2}$ denotes second-order zero matrix and

$$F_{2 \times 2} = \begin{pmatrix} 0 & \beta S_0 \\ 0 & 0 \end{pmatrix}.$$

Let

$$V_{2 \times 2} = \begin{pmatrix} \mu + k_1 + \varepsilon & 0 \\ -\varepsilon & \mu + \vartheta + \gamma \end{pmatrix}. \quad (12)$$

Consequently, the basic reproduction number is

$$R_0 = \rho(\mathbf{FV}^{-1}) = \frac{\beta \varepsilon S_0}{(\mu + k_1 + \varepsilon)(\mu + \vartheta + \gamma)}. \quad (13)$$

Theorem 1. *If $R_0 > 1$, model (2) has $P_0(0, 0, S_0, 0)$ and endemic equilibrium point $P^*(E^*, I^*, S^*, W^*)$. Otherwise, there exists only $P_0(0, 0, S_0, 0)$, where $S_0 = (\Lambda(1 - m)/m + \mu)$, $S^* = (T_1 T_2 / \beta \varepsilon)$, $E^* = T_3 T_1 T_2 \theta (1 - R_0) / \beta \varepsilon (k_1 k_2 - T_1 T_3)$, $I^* = (\varepsilon / T_2) E^*$, $W^* = (k_1 / T_3) E^*$, $T_1 \triangleq \mu + k_1 + \varepsilon$, $T_2 \triangleq \mu + \vartheta + \gamma$, and $T_3 \triangleq \mu + \vartheta + k_2$, $\theta \triangleq m + \mu$.*

Proof. The equilibrium point satisfies the following system of equations:

$$\begin{cases} -\beta IS + \Lambda(1 - m) + k_2 W - mS - \mu S = 0, \\ \beta IS - k_1 E - \varepsilon E - \mu E = 0, \\ k_1 E - \vartheta W - k_2 W - \mu W = 0, \\ \varepsilon E - \gamma I - \vartheta I - \mu I = 0. \end{cases} \quad (14)$$

Noticing that if $I = 0$, we can obtain $P_0(0, 0, S_0, 0)$.

If $I \neq 0$, as a result of the third and fourth equations of (14), we have

$$I^* = \frac{\varepsilon}{T_2} E^*, \quad (15)$$

$$W^* = \frac{k_1}{T_3} E^*.$$

Substituting I^* and W^* in the first and second equations of (14), we conclude that

$$\begin{aligned} S^* &= \frac{T_1 T_2}{\beta \varepsilon}, \\ E^* &= \frac{T_3 [T_1 T_2 \theta - \beta \varepsilon \Lambda (1 - m)]}{\beta \varepsilon (k_1 k_2 - T_1 T_3)}, \\ &= \frac{T_3 T_1 T_2 \theta (1 - R_0)}{\beta \varepsilon (k_1 k_2 - T_1 T_3)}, \end{aligned} \quad (16)$$

where $T_1 \triangleq \mu + k_1 + \varepsilon$, $T_2 \triangleq \mu + \vartheta + \gamma$, and $T_3 \triangleq \mu + \vartheta + k_2$, $\theta \triangleq m + \mu$. Thanks to

$$\begin{aligned} k_1 k_2 - T_1 T_3 &= k_1 k_2 - (\mu + k_1 + \varepsilon)(\mu + \vartheta + k_2) \\ &= -(\mu + \varepsilon)(\mu + \vartheta) - k_1(\mu + \vartheta) - k_2(\mu + \varepsilon) < 0. \end{aligned} \quad (17)$$

Suppose that $R_0 > 1$; obviously, there is $E^* > 0$. Then, in this case, it follows that $P^*(E^*, I^*, S^*, W^*)$. Otherwise, there would be no P^* . \square

4. Stability of Equilibrium Points

It is worth noting that the existence and unique conditions of equilibrium points have been proved in Theorem 1. Next, we will focus on the stability of model (2) with respect to the equilibrium points in this section.

Theorem 2. *If $R_0 < 1$, then $P_0(0, 0, S_0, 0)$ of model (2) is locally asymptotically stable in Ω .*

Proof. We linearize system (2) around $P_0(0, 0, S_0, 0)$. The matrix of the linearization at $P_0(0, 0, S_0, 0)$ is given by

$$J(P_0) = \begin{pmatrix} -T_1 & \beta S_0 & 0 & 0 \\ \varepsilon & -T_2 & 0 & 0 \\ 0 & -\beta S_0 & -\theta & k_2 \\ k_1 & 0 & 0 & -T_3 \end{pmatrix}, \quad (18)$$

where $T_1 = \mu + k_1 + \varepsilon$, $T_2 = \mu + \vartheta + \gamma$, $T_3 = \mu + \vartheta + k_2$, and $\theta = m + \mu$.

The formula of characteristic equation for $J(P_0)$ being

$$\begin{aligned} f(\lambda) &= (\lambda + T_1)(\lambda + T_2)(\lambda + T_3)(\lambda + \theta) \\ &\quad - \beta S_0 \varepsilon (\lambda + \theta)(\lambda + T_3) = 0, \end{aligned} \quad (19)$$

which leads to

$$(\lambda + \theta)(\lambda + T_3)[(\lambda + T_1)(\lambda + T_2) - \beta S_0 \varepsilon] = 0. \quad (20)$$

Clearly, (20) always has negative roots $\lambda_1 = -\theta$ and $\lambda_2 = -T_3$. All other roots are determined by

$$(\lambda + T_1)(\lambda + T_2) - \beta \varepsilon S_0 = 0. \quad (21)$$

Substituting (13) into (21) yields

$$(\lambda + T_1)(\lambda + T_2) - T_1 T_2 R_0 = 0, \quad (22)$$

which is of the form $\lambda^2 + b_1 \lambda + b_2 = 0$, where $b_1 = T_1 + T_2$ and $b_2 = T_1 T_2 (1 - R_0)$.

When $R_0 < 1$, we conclude that $\Phi_1 = b_1 = T_1 + T_2 > 0$ and $\Phi_2 = b_1 b_2 = T_1 T_2 (T_1 + T_2) (1 - R_0) > 0$. According to Lemma 1 [14], there exist negative real parts for all roots of (21). Therefore, $P_0(0, 0, S_0, 0)$ of (2) is locally asymptotically stable. \square

Theorem 3. *If $R_0 > 1$, $P^*(E^*, I^*, S^*, W^*)$ is locally asymptotically stable in Ω^0 .*

Proof. From Theorem 1, there exists P^* in model (2) only if $R_0 > 1$. Then, the Jacobian matrix of model (2) at $P^*(E^*, I^*, S^*, W^*)$ is presented as follows

$$J(P^*) = \begin{pmatrix} -T_1 & \beta S^* & \beta I^* & 0 \\ \varepsilon & -T_2 & 0 & 0 \\ 0 & -\beta S^* & -\beta I^* - \theta & k_2 \\ k_1 & 0 & 0 & -T_3 \end{pmatrix}, \quad (23)$$

where $T_1 = \mu + k_1 + \varepsilon$, $T_2 = \mu + \gamma + \vartheta$, $T_3 = \mu + \vartheta + k_2$, and $\theta = m + \mu$.

The characteristic equation of $J(P^*)$ is

$$\begin{aligned} g(\lambda) &= (\lambda + T_1)(\lambda + T_2)(\lambda + T_3)(\lambda + \beta I^* + \theta) \\ &\quad - \beta \varepsilon S^* (\lambda + T_3)(\lambda + \theta) - k_1 k_2 \beta I^* (\lambda + T_2) = 0, \end{aligned} \quad (24)$$

that is,

$$\lambda^4 + b_1\lambda^3 + b_2\lambda^2 + b_3\lambda + b_4 = 0, \quad (25)$$

where

$$\begin{aligned} b_1 &= T_1 + T_2 + T_3 + \beta I^* + \theta > 0, \\ b_2 &= T_1T_3 + T_2T_3 + (\beta I^* + \theta)(T_1 + T_2 + T_3) > 0, \\ b_3 &= \beta I^*(T_1T_2 + T_2T_3) + \theta(T_1T_3 + T_2T_3) + \beta I^*(T_1T_3 - k_1k_2), \\ b_4 &= \beta I^*T_2(T_1T_3 - k_1k_2). \end{aligned} \quad (26)$$

Considering that $T_1T_3 - k_1k_2 = (\mu + k_1 + \varepsilon)(\mu + \vartheta + k_2) - k_1k_2 > 0$, we arrive at $b_3 > 0, b_4 > 0$.
Owing to

$$\begin{aligned} \Phi_1 &= b_1 = T_1 + T_2 + T_3 + \beta I^* + \theta > 0, \\ \Phi_2 &= b_1b_2 - b_3 \\ &= (T_1 + T_2 + T_3 + \beta I^* + \theta)[T_1T_3 + T_2T_3 + (\beta I^* + \theta)(T_1 + T_2 + T_3)] \\ &\quad - [\beta I^*(T_1T_2 + T_2T_3) + \theta(T_1T_3 + T_2T_3) + \beta I^*(T_1T_3 - k_1k_2)] \\ &= (T_1 + T_2 + T_3)[(T_1 + T_2 + T_3 + \beta I^* + \theta)(\beta I^* + \theta) + (T_1T_3 + T_2T_3)] \\ &\quad + \beta I^*k_1k_2 - \beta I^*T_1T_2 > 0, \\ \Phi_3 &= b_3(b_1b_2 - b_3) - b_1^2b_4 = b_3\Phi_2 - b_1^2b_4 \\ &= [\beta I^*(T_1T_2 + T_2T_3) + \theta(T_1T_3 + T_2T_3) + \beta I^*(T_1T_3 - k_1k_2)] \\ &\quad \{(T_1 + T_2 + T_3)[(T_1 + T_2 + T_3 + \beta I^* + \theta)(\beta I^* + \theta) + (T_1T_3 + T_2T_3)] + \beta I^*(k_1k_2 - T_1T_2)\} \\ &\quad - (T_1 + T_2 + T_3 + \beta I^* + \theta)^2[\beta I^*T_2(T_1T_3 - k_1k_2)] \\ &= \beta I^*(\beta I^* + \theta)^2(T_1T_3 - k_1k_2)(T_1 + T_3) \\ &\quad + \beta I^*(\beta I^* + \theta)(T_1 + T_2 + T_3)[(T_1T_2 + T_2T_3)(T_1 + T_2 + T_3) + T_2k_1k_2 - T_1T_2T_3] \\ &\quad + \beta I^*(T_1 + T_2 + T_3)(T_2T_3)^2 + \Phi_2[\beta I^*(k_1k_2 - T_1T_3) + \theta(T_1T_3 + T_2T_3)] > 0, \\ \Phi_4 &= b_4\Phi_3 = \beta I^*T_2(T_1T_3 - k_1k_2)\Phi_3 > 0. \end{aligned} \quad (27)$$

It follows from Lemma 1 [14] that there exist negative real parts for all roots of (24). Therefore, $P^*(E^*, I^*, S^*, W^*)$ is locally asymptotically stable. The corresponding global asymptotic stability will be verified in the numerical simulation. \square

Theorem 4. *If $R_0 < 1$, $P_0(0, 0, S_0, 0)$ of model (2) is globally asymptotically stable in Ω .*

Proof. Define the following Lyapunov function as

$$\mathcal{L}(t) = \varepsilon E(t) + (\mu + k_1 + \varepsilon)I(t). \quad (28)$$

Obviously, we have $\mathcal{L}(t) \geq 0$. Furthermore, $\mathcal{L}(t) = 0$ if and only if both $E(t) = 0$ and $I(t) = 0$.

The derivative of $\mathcal{L}(t)$ along model (2) is

$$\begin{aligned} \frac{d\mathcal{L}(t)}{dt} \Big|_{(4)} &= \varepsilon(\beta I(t)S(t) - \mu E(t) - k_1E(t) - \varepsilon E(t)) \\ &\quad + (\mu + k_1 + \varepsilon)(\varepsilon E(t) - \mu I(t) - \gamma I(t) - \vartheta I(t)), \\ &= \beta \varepsilon I(t)S(t) - (\mu + k_1 + \varepsilon)(\mu + \gamma + \vartheta)I(t), \\ &= I(t)[\beta \varepsilon S(t) - (\mu + k_1 + \varepsilon)(\mu + \gamma + \vartheta)] \\ &\leq I(t)[\beta \varepsilon S_0 - (\mu + k_1 + \varepsilon)(\mu + \gamma + \vartheta)] \\ &= I(t)(\mu + k_1 + \varepsilon)(\mu + \gamma + \vartheta)(R_0 - 1). \end{aligned} \quad (29)$$

When $R_0 < 1$, we have $d\mathcal{L}(t)/dt \leq 0$. Furthermore, we obtain $d\mathcal{L}(t)/dt = 0$ if and only if $I(t) = 0$. Consequently, the largest set of compact invariants in $\{(S, E, W, I) | d\mathcal{L}(t)/dt = 0\}$ is the single point set $\{P_0\}$ when $R_0 < 1$. With the aid of Lemma 2 [15] and Lemma 3 [16], $P_0(0, 0, S_0, 0)$ of (2) is globally asymptotically stable in Ω when $R_0 < 1$. \square

5. Model Simulations

In the section, first the relevant results in Section 4 will be verified. Subsequently, the impact of vaccination and government control on infectious diseases will be analyzed. Parameters are listed in Table 1.

The initial values for $(S(0), E(0), I(0), R(0))$ is set as $(120, 0, 15, 1, 0)$, $(70, 0, 15, 1, 50)$, and $(20, 0, 15, 1, 100)$. m and k_1 are variables, and the other parameters are set to constant values in Table 1.

In Case 1, we set $m = 0.8$ and $k_1 = 0.6$. Then, $R_0 = 0.9802 < 1$ and $P_0(0, 0, S_0, 0) = (0, 0, 3.4091, 0)$. Then, we simulate the changes in the number of these four populations under the condition of $R_0 < 1$. In Case 2, we set $m = 0.35$ and $k_1 = 0.3$. Then, $R_0 = 8.1769 > 1$ and $P^*(E^*, I^*, S^*, W^*) = (7.2730, 6.1898, 2.7730, 2.4516)$. Similarly, the population size for the condition $R_0 > 1$ is simulated as well. The relative results are illustrated in Figure 2. The first row

TABLE 1: Parameters definition and estimated values.

Parameters	Definition	Value
Λ	Population replenishment rate	15
β	Infected population level infection rate	0.5
μ	Natural mortality rate	0.08
ε	Conversion from unquarantined virus exposed population to infected populations	0.8
γ	Disease mortality	0.06
m	Vaccination success rate	Variable
k_1	Government control rate	Variable
ϑ	Recovery rate	0.8
k_2	Autoviral immunity rate	0.01

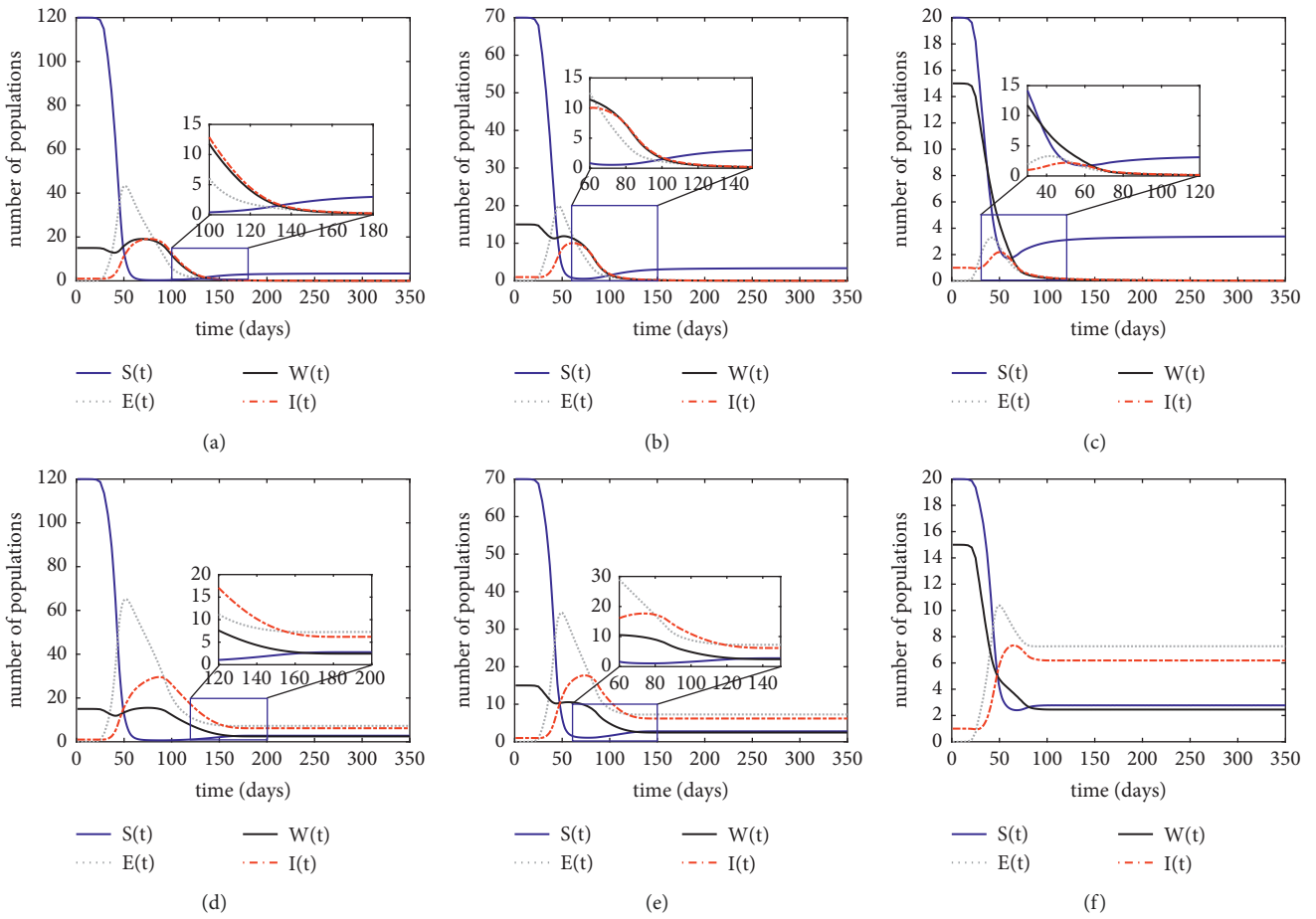


FIGURE 2: Equilibria global asymptotic stability of model (2). The first row is the case 1, and the second row is the case 2. (a, d) Initial values set as (120, 0, 15, 1, 0); (b, e) initial values set as (70, 0, 15, 1, 50); (c, f) initial values set as (20, 0, 15, 1, 100). (a) Case 1: initial value (120, 0, 15, 1, 0). (b) Case 1: initial value (70, 0, 15, 1, 50). (c) Case 1: initial value (20, 0, 15, 1, 100). (d) Case 2: initial values (120, 0, 15, 1, 0). (e) Case 2: initial values set as (70, 0, 15, 1, 50). (f) Case 2: initial values set as (20, 0, 15, 1, 100).

shows the changes in population size for different initial settings in case 1, and the second row shows the changes in population size for case 2.

From Figure 2, it can be concluded that the decrease in the initial value of susceptible individuals at the beginning of the outbreak leads to a shorter duration of the epidemic. Meanwhile, the peak of confirmed cases reaches a lower level in almost unanimous time.

Focusing on Figures 2(a) and 2(d), it is evident that the equilibria of (2) are globally asymptotically stable. Correspondingly, three conclusions can be drawn: (1) compared to Case 1, the amount of the four populations in Case 2 starts to stabilize after about day 160, however, about 20 days later than in Case 1. (2) In Case 2, neither the exposed nor the infected populations can achieve disappearance, i.e., the epidemic will persist. However, as illustrated in Case 1, once

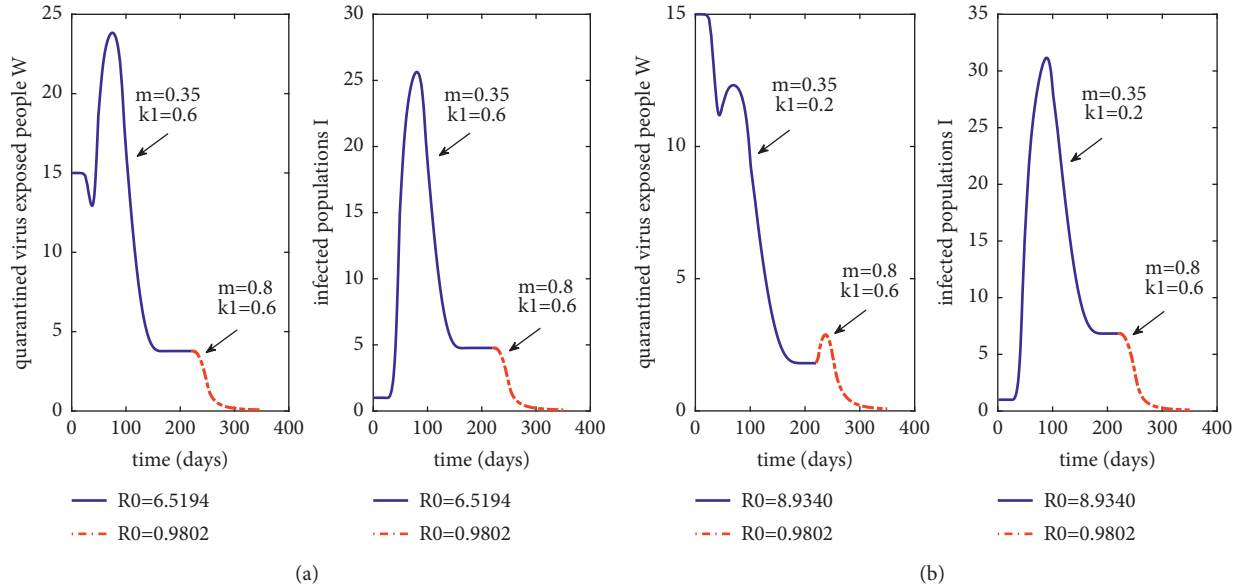


FIGURE 3: Effect of parameters m and k_1 on quarantined virus exposed population and infected population. (a) Case 3: adjustment of parameter m only. (b) Case 4: simultaneous adjustment of parameters m and k_1 .

the four populations have stabilized, the total population will consist entirely of the susceptible and recovered population that has been immunized. This is a high-priority objective for epidemic control. (3) The exposed and infected populations in Case 2 consume a similar amount of time as in Case 1, but reach a higher peak. Hence, we can infer from these findings that government involvement and vaccination both have an impact on the epidemic dynamics.

From the analysis above, we can obtain that m as well as k_1 has a beneficial effect on the prevention and control of the epidemic. Consequently, we will specifically investigate the effectiveness of m and k_1 for the control of COVID-19. The initial value is set as (120, 0, 15, 1, 0). Changes in the amount of quarantined virus exposed population $W(t)$ and infected population $I(t)$ will be observed.

In Case 3, we set $m = 0.35$ and $k_1 = 0.6$ initially. And, the basic reproduction number $R_0 = 6.5149 > 1$. From Figure 3(a), we find the blue parts of the $W(t)$ and $I(t)$ curves converge quickly after about day 150 and reach dynamic equilibrium, in which case extinction of the epidemic disease cannot be achieved. On day 220, we raise m to $m = 0.8$, at which point $R_0 = 0.9802 < 1$. Obviously, $I(t)$ and $W(t)$ both started to decline rapidly from day 220 and dropped below 1 around day 260. Hence, it can be concluded that the increase of vaccination success rate m has a strong and efficient effect on the implementation of the COVID-19 control efforts.

In Case 4, we set $m = 0.35$ and $k_1 = 0.2$ initially, i.e., reduce the government control rate k_1 on the basis of case 3 at the stage when no epidemic control work is carried out. Subsequently, both m and k_1 were increased to be consistent with the parameter values on day 220 in Case 3. Thus, Figure 3(b) can be obtained. Comparison Figures 3(a) and 3(b), it can be found that, with a higher the initial k_1 , the peak number of $W(t)$ will rise by about 93.5%, but the peak number of $I(t)$ will drop by about 17.7%. When the epidemic control work

begins on day 220, the number of $W(t)$ exhibits an increasing trend followed by a declining trend, which would result in higher epidemic control costs compared to that of Case 3.

By calculating the time points of no new $W(t)$ and $I(t)$ addition after the control work, the results of day 258 and day 261 of Case 3 can be obtained, respectively. Meanwhile, we can obtain the results of day 264 and day 269 of Case 4. Consequently, if the implementation of the government's control work intensity can reach a high level in the early stage of the outbreak of COVID-19, it will help to shorten the time of control work with a reduction in new infection cases.

In Cases 3 and 4, we first simulate the process of epidemic from an outbreak to a convergence to equilibrium without extinction. Then, we insert control means immediately afterwards including vaccination and government intervention which make the epidemic go to extinction, thus achieving the purpose of control. Next, we will investigate the role of m and k_1 for epidemic control individually.

In Case 5, we fix $k_1 = 0.8$ and adjust m under the condition of ensuring $R_0 < 1$, so as to monitor variations in the number of $I(t)$. Then, Figure 4(a) can be obtained. Under the limited condition of $R_0 < 1$, we can determine that the three time points when the number of $I(t)$ peaks all fell within the range of day 50 to day 100.

In Case 6, similarly, we fix $m = 0.8$ and adjust k_1 to obtain Figure 4(b). From Figure 4(b), we can conclude that, with the improvement of k_1 , the time points when the three curves reach their peaks tend to be earlier. In other words, when m reaches a high level, the increase of k_1 will simultaneously lead to two positive effects: an earlier point of time at which the number of $I(t)$ peaks and peaks decline.

At the same time, we find, to m and k_1 , the corresponding sensitivity of R_0 are not identical. Then, according to (4) in Section 3, the relationship between these two factors and R_0 can be obtained, as shown in Figure 5. It can be

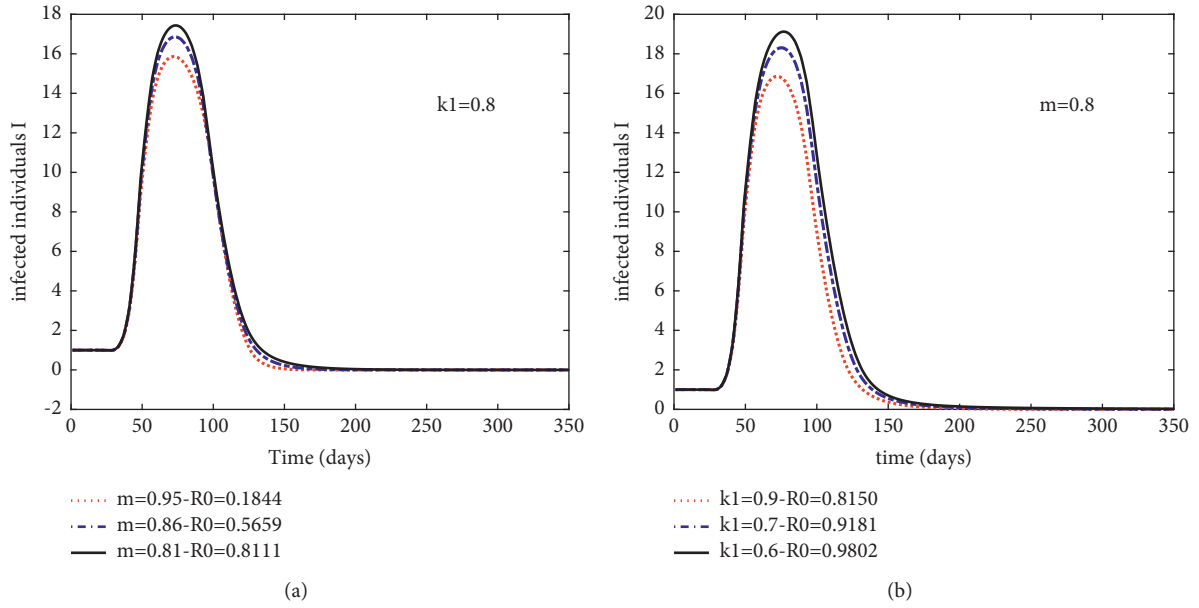


FIGURE 4: Sensitivity analysis of parameter m and k_1 to infected population when $R_0 < 1$. (a) Case 5: sensitivity of parameter m . (b) Case 6: sensitivity of parameter m and k_1 .

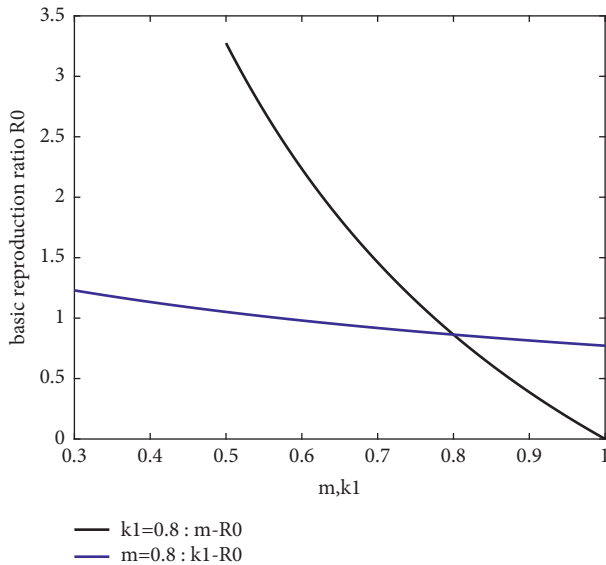


FIGURE 5: The relationship between parameter m and k_1 on the basic regeneration number R_0 .

obtained that R_0 is more sensitive to m . Then, when k_1 and m are, respectively, increased within the same numerical range, m has a greater benefit for the mitigation of the epidemic.

According to the simulation results obtained with, we find, the expansion of vaccination coverage, as well as the increase in government intervention contribute to the extinction of infectious diseases. Under the condition that $R_0 < 1$, once there exists an increase in m and k_1 , it will bring a positive effect of reducing the number of confirmed cases. And, if government intervention continues to expand in the face of vaccination for universal access, an added effect of an earlier demise of the epidemic will be gained.

6. Conclusion and Strategies

In the present work, a SEWIR epidemic model with vaccination and government control is presented. The basic regeneration number R_0 is calculated as the threshold which describes the extinction or persistence of the disease. By constructing the Lyapunov function and numerical simulations, it can be demonstrated that the disease-free equilibrium point of model (2) is globally asymptotically stable when $R_0 < 1$ and the infectious disease will tend to die out over time; when $R_0 > 1$, the endemic equilibrium point of model (2) is also globally asymptotically stable but will not achieve extinction but maintain dynamic equilibrium. In parallel, through extensive simulation experiments, we further study the impact of vaccination success and government control rates on the dynamics of the COVID-19 epidemic. In Case 3, we simulate the effect of vaccination on epidemic control. When only the vaccination success rate is adjusted from 0.35 to 0.8 and the government control rate is kept constant, the basic regeneration number R_0 is reduced from 6.5194 to 0.9802, and the epidemic tends to extinction in about 40 days. Hence, we can obtain the first control strategy: advancing the progress of vaccine development for COVID-19 and improving the immunization success rate of the vaccine. In Case 4, we adjust both the vaccination success rate from 0.35 to 0.8 and the government control rate from 0.2 to 0.6. Compared to Case 3, the number of quarantined exposed population experiences an increase and then a decrease. For both cases, we can conclude that if the government double the rate of control in the early stage of the outbreak, the number of quarantined exposed population will increase by 93.5%, but the number of infected population will decrease by 17.7% successively. Therefore, a second control strategy can be obtained: promoting the progress of governmental efforts to contain the epidemic

during its early stage: for example, to increase the frequency and scope of medical testing, to expand the investigation of people who have been in contact with positive patients, and to implement home quarantine for all people in certain region. Both of these strategies have good effect on the control of the epidemic.

The results above suggest that improving the immunization success rate of vaccine, enhancing vaccination efforts and government intervention can be used as strategies to control the epidemic during COVID-19. However, the huge impact of a long-term region-wide quarantine on the daily life of public and the enormous consumption of human and material resources during the quarantine cannot be ignored. Consequently, it can only be used as a short-term expedient measure. In order to fundamentally eliminate the epidemic, it is necessary to vigorously promote vaccination and improve the quality of vaccines to reduce the diagnosis rate and mortality rate.

Nevertheless, these results must be interpreted with caution, and there are several limitations that cannot be ignored. First, cultural differences and social factors between countries were not taken into account in the modeling process. Large-scale and long-term medical isolation measures are not applicable to all countries. Second, economic factors were also not considered. The proposed optimal prevention and control strategy is relatively idealistic and has some implementation limitations. In subsequent studies, it will be considered again from this aspect as well as the perspective of vaccination methods and vaccination coverage.

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 declare that they have no conflicts of interest.

References

- [1] R. Almeida, "Analysis of a fractional SEIR model with treatment," *Applied Mathematics Letters*, vol. 84, no. 1, pp. 56–62, 2018.
- [2] J. M. Carcione, J. E. Santos, C. Bagaini, and J. Ba, "A simulation of a COVID-19 epidemic based on a deterministic SEIR model," *Frontiers in Public Health*, vol. 8, no. 230, pp. 1–13, 2020.
- [3] M. A. Khan, Y. Khan, and S. Islam, "Complex dynamics of an SEIR epidemic model with saturated incidence rate and treatment," *Physica A: Statistical Mechanics and Its Applications*, vol. 493, pp. 210–227, 2018.
- [4] P. Auger and A. Moussaoui, "On the threshold of release of confinement in an epidemic SEIR model taking into account the protective effect of mask," *Bulletin of Mathematical Biology*, vol. 83, no. 4, pp. 25–18, 2021.
- [5] M. De la Sen, A. Ibeas, S. Alonso-Quesada, and R. Nistal, "On a SIR model in a patchy environment under constant and feedback decentralized controls with asymmetric parameterizations," *Symmetry*, vol. 11, no. 3, p. 430, 2019.
- [6] M. De la Sen, S. Alonso-Quesada, A. Ibeas, and R. Nistal, "On an SEIADR epidemic model with vaccination, treatment and dead-infectious corpses removal controls," *Mathematics and Computers in Simulation*, vol. 163, pp. 47–79, 2019.
- [7] R. Nistal, M. De la Sen, S. Alonso-Quesada, and A. Ibeas, "On a new discrete SEIADR model with mixed controls: study of its properties," *Mathematics*, vol. 7, no. 1, pp. 18–19, 2018.
- [8] D. Otoo, J. A. Kessie, E. K. Donkoh, E. Okyere, and W. Kumi, "Global dynamics of an SEIRS compartmental measles model with interrupted vaccination," *Applied Mathematics*, vol. 10, no. 7, pp. 588–604, 2019.
- [9] H. Jiao and Q. Shen, "Dynamics analysis and vaccination-based sliding mode control of a more generalized SEIR epidemic model," *IEEE Access*, vol. 8, pp. 174507–174515, 2020.
- [10] A. I. NasimUllah, M. I. M. A. Ibeas, M. Shafi, M. Ishfaq, and M. Ali, "Vaccination controllers for SEIR epidemic models based on fractional order dynamics," *Biomedical Signal Processing and Control*, vol. 38, pp. 136–142, 2017.
- [11] M. H. Darassi, M. A. Safi, and B. Al-Hdaibat, "A delayed SEIR epidemic model with pulse vaccination and treatment," *Nonlinear Studies*, vol. 25, no. 3, pp. 521–534, 2018.
- [12] J. J. Jiao, S. H. Cai, and L. M. Li, "Dynamics of a delayed SEIR epidemic model with pulse vaccination and restricting the infected dispersal," *Communications in Mathematical Biology and Neuroscience*, vol. 18, pp. 1–23, 2017.
- [13] L. Wang, "Existence of periodic solutions of seasonally forced SEIR models with pulse vaccination," *Discrete Dynamics in Nature and Society*, vol. 2020, no. 12, 11 pages, Article ID 9381375, 2020.
- [14] F. Bauer and J. A. Nohel, *The Qualitative Theory of Ordinary Differential Equations: An Introduction*, pp. 1–314, Dover Publications, New York, NY, USA, 1989.
- [15] A. M. Lyapunov, "The general problem of the stability of motion," *International Journal of Control*, vol. 55, no. 3, pp. 531–534, 1992.
- [16] O. Pardo, "Global stability for a phytoplankton-nutrient system," *Journal of Biological Systems*, vol. 8, no. 2, pp. 195–209, 2000.
- [17] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, no. 1–2, pp. 29–48, 2002.
- [18] Z. Shuai and P. van den Driessche, "Global stability of infectious disease models using Lyapunov functions," *SIAM Journal on Applied Mathematics*, vol. 73, no. 4, pp. 1513–1532, 2013.

Research Article

Residual Life Prediction of Lithium Batteries Based on Data Mining

Dandan Ma ¹ and Xiangge Qin ²

¹Information Science & Electronic Technology, Jiamusi University, Jiamusi 154007, Heilongjiang, China

²Materials Science & Engineering, Jiamusi University, Jiamusi 154007, Heilongjiang, China

Correspondence should be addressed to Dandan Ma; madandan@jmsu.edu.cn and Xiangge Qin; qinxiangge@jmsu.edu.cn

Received 7 March 2022; Revised 1 April 2022; Accepted 6 April 2022; Published 13 June 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Dandan Ma and Xiangge Qin. 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.

Lithium-ion batteries are an important part of smartphones, and their performance has a great impact on the life of the phone. The longevity of lithium-ion batteries is key to ensuring their reliability and extending their useful life. This paper built a lithium battery life prediction model and grey model MDGM(1,1) based on data mining. Then, experimental data were selected for testing, and the prediction error reached 10.5% at the minimum. It showed that the prediction model had higher precision and could provide help for the prediction and development of mobile phone battery life.

1. Introduction

With the advent of the era of “information explosion,” data mining arises at a historic moment to deal with the challenge of “knowledge shortage.” Data mining is a process of extracting valuable information and knowledge from a large amount of data. It has been widely used in society, economy, production, life, and other aspects. As we know, data mining is the massive historical data existing in databases, data warehouses, and external source files. However, some data have a small sample amount or are disabled, or the overall law is particularly complex, but the data in a certain time or space has strong regularity. At present, there is no effective data processing method, and the grey system theory can solve these problems very well. Taking advantage of grey system, it can be widely used in data mining. The two complementary advantages can make the discovered knowledge more effective and reliable.

Because of its advantages of lightweight, low discharge, and long life, lithium-ion battery is an energy storage device for portable electronics (such as notebook computers, digital cameras, tablet computers, mobile phones, and other handheld electronic products) [1]. As the core component

of electronic products, in long-term use, due to various factors, the performance and life of lithium batteries will be affected, which may cause some troubles for users. That is why battery life has become a growing concern in recent years [2]. The basic composition of the battery pack is a single lithium battery. Therefore, the research of battery pack management system technology usually includes the monitoring of temperature, current and voltage of a single battery, capacity prediction, and charge state prediction of a single battery [3].

The research significance of this paper is mainly reflected in two aspects. Firstly, from the theoretical aspect, this paper combined the relevant theoretical data of data mining and grey system; at the same time, this paper also combines these two methods to establish the remaining life prediction model of lithium battery. Secondly, from the practical point of view, the lithium battery life prediction model constructed in this paper based on grey data mining can predict the life of lithium battery in advance and predict the battery life, which can effectively predict the future working ability of electronic products, timely find problems, and avoid unnecessary troubles and losses. This is very meaningful for the research and maintenance of lithium batteries.

2. Literature Review

2.1. Battery Life Prediction Technology

2.1.1. Virtual Sample Technology. Traditional statistics are based on a sufficient sample number, but in reality, the sample number is limited or very small, which cannot meet the basic requirements of statistics. Therefore, a small sample is a classic problem [4]. In the era of big data, people are increasingly demanding for learning methods of two types of extreme sample data. These two types of sample data are massive samples that need to quickly obtain decision-making information due to big data and small samples that cannot obtain more data for learning in order to achieve fast response [5]. As the name implies, the problem of the small sample can be understood as the lack of information, but the key factors leading to the learning difficulty of the small sample are not all in the aspect of sample size. If there are a lot of data, but its distribution presents a discrete loose structure and there are gaps between sample points, the problem of incomplete and unbalanced samples will not be able to obtain effective information [6]. The problem of the small sample is mainly caused by unreasonable experimental design, high cost of obtaining sample data, or even the inability to obtain more sample data. In addition, it may also be due to the low probability of data occurrence, or although there are much data, it is data duplication. The problem of the small sample can be further divided into data scarcity, unbalanced data, and stability of feature selection of high-dimensional small sample [7]. High-quality data is the basis of effective decision-making. Therefore, to improve model-based accuracy and operability, it is an important work to expand the amount of effective data based on small sample data information to meet the requirements of modeling data.

The virtual sample is difficult to approximate the real sample completely and accurately. If too few virtual samples are generated, the additional information of unknown space carried by virtual samples is insufficient, and the generalization ability of the final model will be limited. If too many virtual samples are generated, on the one hand, the influence of real samples will be weakened, and on the other hand, errors introduced by virtual samples will deteriorate the generalization ability of the final model [8]. The more virtual samples are generated, the more untrusted information is brought in. Therefore, in order to maximize the generalization performance of the final model, there is an optimal number of virtual samples. The determination of the optimal virtual sample number has become an open problem in the research of the small sample problem [9].

2.1.2. Battery Life Prediction. The capacity of a battery is an important indicator of battery performance, and its lifespan is an important indicator of its health. When the battery capacity decreases to a certain critical point (typically 80% of the rated capacity, but different battery types have different thresholds), the battery is considered to be invalid [10]. Remaining service life refers to the remaining service life of a battery after it has been used for a period of time [11–16]. For

example, the power lithium battery has a cycle life of 500 times; that is, it can last 500 times under normal charging and discharging conditions. If it has been used 100 times, the remaining service life is 400 times. Generally, at present, the life prediction methods of lithium-ion batteries mainly include physical model and data-driven methods.

The physical model method mainly studies the internal structure of the lithium-ion battery and analyzes its physical and chemical changes. The formation of SEI film is considered to be the main cause of battery decay. The formation of SEI film consumes lithium ions inside the battery, and the internal resistance between electrode and electrolyte increases with the thickening of SEI film, which further leads to the decline of battery capacity [7, 17–19]. The method of the physical model involves the internal molecular level of the battery, which requires obtaining a large amount of information about the material properties and failure mechanism of lithium-ion batteries. However, internal chemical reactions are difficult to measure; it is difficult to predict the battery life through a physical model [12, 20]. The data-driven approach eliminates the need to study the internal chemical reactions of the battery and avoids the complex model of studying the interior of the battery, and on the basis of the existing experimental data, a functional model is obtained by using the method of data fitting [21–23]. Commonly used data-driven methods include grey model, BP neural network, SVM, correlation vector machine, ARMA model, and so on. The research object of this paper is the prediction of the remaining life of the lithium battery based on data-driven [24]. The capacity decay of lithium-ion batteries reflects the aging of batteries. Capacity refers to the amount of charge released in the complete process of discharging from full charge to empty charge, usually expressed in ampere-hour (Ah), as follows:

$$C = \int_k^{k+1} i dt. \quad (1)$$

When the current is constant, capacity $C=It$, where I represents the battery discharge current and T represents the discharge time. The capacity of the battery decreases with the life cycle of the battery. When the discharge capacity drops to 80%, it is considered to be the end of the service life. For example, when the battery power rating is 4 Ah, the aging threshold is 3.2 Ah. This paper uses experimental data to predict the capacity decay of the battery during charging and discharging cycles. [7].

2.2. Review of Lithium Battery Life Prediction. At present, although there are many kinds of lithium battery life prediction algorithms, according to the principle of modeling, it is divided into two types: physical model and data-driven. Each of these approaches is described below [14]. (1) Physical model-based methods refer to the prediction of remaining life (RUL) based on the physical properties of lithium batteries (e.g., material properties and loading conditions) and degradation mechanisms. (2) The data-driven RUL prediction method is to extract the characteristic parameters that can reflect the battery health status from the monitored variables such as voltage and current and build a

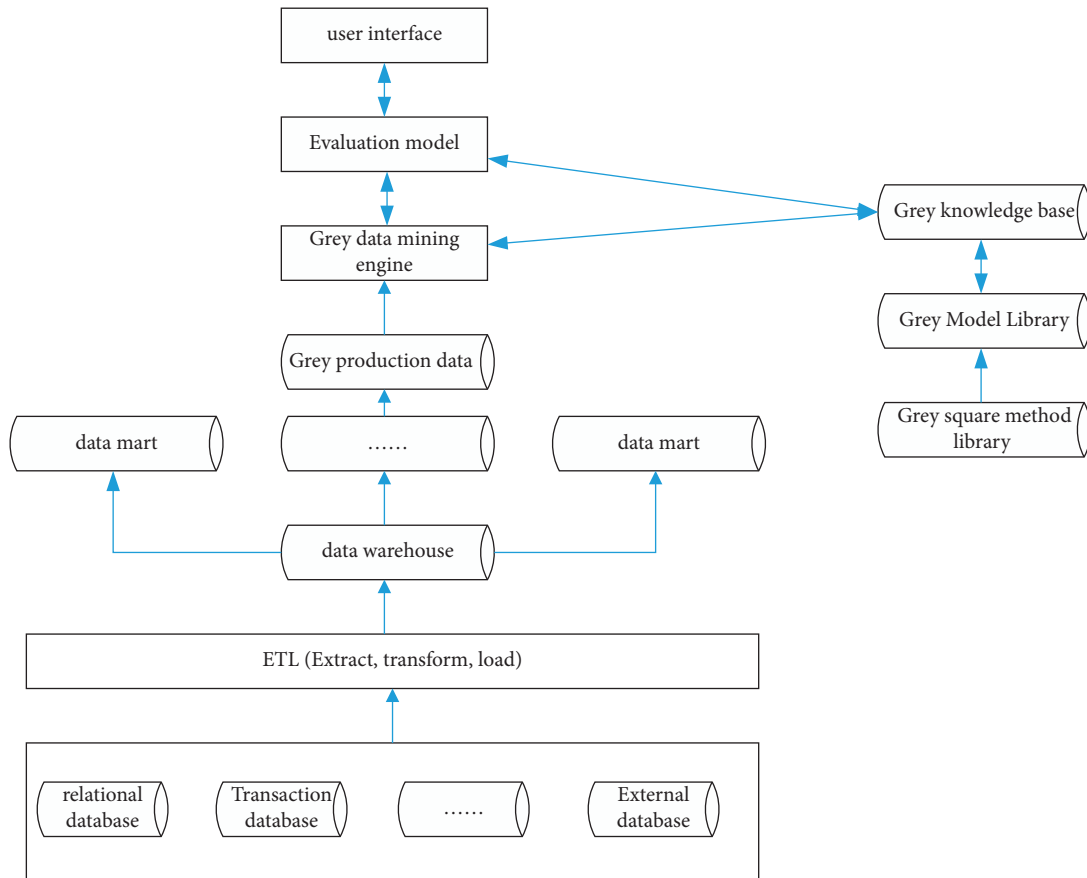


FIGURE 1: Structure of the grey data mining system.

statistical model of the system to extrapolate the prediction of RUL [25]. Because the data-driven method does not require the establishment of a complex system physical model, it is suitable for RUL prediction of complex and changeable internal lithium battery systems. Data-driven forecasting methods generally fall into two categories: artificial intelligence methods and statistics-driven methods [26].

Professor Deng Julong first proposed the grey theory in 1982. The theory is first applied in uncertain systems and then widely used in other systems. Its advantage is that it requires less data and has obvious advantages. The prediction methods based on grey theory include the metabolic Markov residual grey model and metabolic grey model. By selecting different models under different conditions, the applicability of the models is improved, and the estimation accuracy is also improved [27].

According to the principle of non-equilibrium thermodynamics, Virkar built a degradation mechanism model of the lithium battery by analyzing the electrochemical reaction inside the battery and added SEI membrane and chemical potential into the process of predicting RUL. Gong has proposed a method for predicting the residual life of lithium batteries based on a gas generation model.

According to the analysis and research on the actual gas generation mechanism of the battery, the types of chemical reactions in the battery were determined, and the gas generation equation was established. Based on these gas

equations, a relationship model between battery capacity and gas production characteristics was established, and a prediction model for the remaining life of lithium battery was established finally. TPF is a probability density function represented by random samples in the state space [28, 29].

To sum up, the fusion model-based method is still the focus of future research. It not only improves the prediction accuracy but also improves the robustness of model output. However, its disadvantages are high algorithm complexity and large computation. In addition, most of the research on battery life prediction are still in the laboratory stage, so it is still a long way to go.

3. Lithium Battery Life Prediction Model Construction Based on Data Mining

3.1. Grey Data Mining System. The grey data mining system draws on the structure of the traditional data mining system and makes full use of the grey system method to carry out data mining. It is based on a database and data warehouse and provides effective methods for data mining tools of data warehouse [16]. Figure 1 shows the system architecture.

3.2. Establishment of Grey Model MDGM(1,1). Since it was proposed by Deng Julong in 1982, the grey system has been widely used in various forecasting fields. Metabolism

discrete grey prediction MDGM(1,1) model is the core of grey system theory. MDGM(1,1) quantifies the abstract concept of the known information in the system first, then processes the quantified concept through modeling, and finally optimizes the model to predict the unknown data. Grey prediction MDGM(1,1) model has long been concerned and valued by people.

The MDGM(1,1) model is formed by the first-order differential equation with only one variable. First of all, the original sequence is accumulated and generated to present a certain rule; then the first-order linear differential equation model is established; and finally, the fitting curve is obtained to predict the system, which is the process of establishing MDGM(1,1) model. Details are as follows:

Let $X(0)$ be a nonnegative sequence:

$$X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)). \quad (2)$$

First-order accumulation is

$$X^{(1)} = (X^{(1)}(1), X^{(1)}(2), \dots, X^{(1)}(n)). \quad (3)$$

Adjacent to the mean-generating sequence

$$Z(1) = Z^{(1)}(2), Z^{(1)}(3), \dots, Z^{(1)}(n) \quad (4)$$

if

$$\hat{a} = [a, b]^T. \quad (5)$$

The GM(1,1) model is

$$X^{(0)}(k) + aZZ^{(1)}(k) = b. \quad (6)$$

The least-square estimation parameter column satisfies

$$\hat{a} = [B^T, B]^{-1} B^T Y \quad (7)$$

if the set is

$$X^{(0)}(k) = (\beta - aX^{(0)}(1))e^{-a(k-2)}. \quad (8)$$

The above equation is the basis for prediction in this paper.

3.3. Accuracy Test of Grey Model MDGM(1,1). After calculation, the small error probability p is 0.983; the ratio of post-test square difference C is 0.251; and the accuracy level is 1; and it shows that the model can be used for prediction. The grey GM(1,1) model is used to predict the life of lithium-ion batteries, as shown in Figure 2.

According to Figure 2, Battery18 reaches the failure point after 100 cycles of charging and discharging, so its real life is 100 cycles. When the starting point of prediction is 60 cycles, the life failure point predicted by the grey GM(1, 1) model is 120 cycles. The error is 20 cycles; the prediction accuracy is ordinary; for the convenience of comparison, the predicted results at $T = 40, 60,$ and 80 cycles were put on the same graph, as shown in Figure 3.

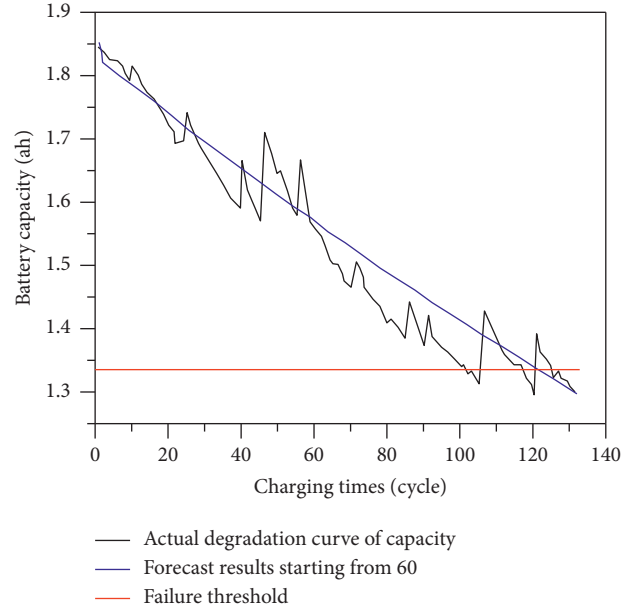


FIGURE 2: Prediction results of lithium-ion battery life based on grey model.

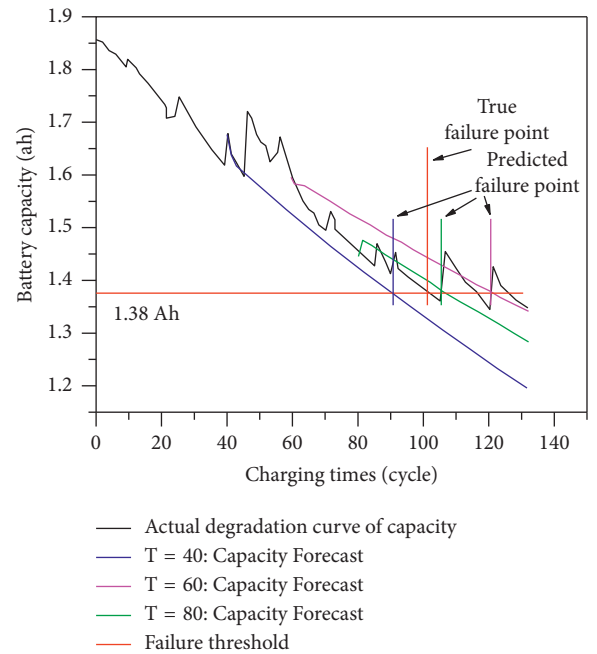


FIGURE 3: Life prediction of lithium-ion battery based on grey model (B18).

To sum up, GM(1,1) model in grey theory is used to predict the life of the lithium-ion battery in this section. The experimental results show that the prediction effect of this method is acceptable, but the accuracy needs to be improved. As the capacity decay trajectory of the lithium-ion battery is nonlinear and random, the predicted trajectory of the grey model is close to a straight line, which can be regarded as the prediction of the trend term in the capacity degradation trajectory, but the random term is not predicted.

Considering the degradation of battery capacity can also be regarded as a series of time series, the life prediction of the lithium-ion battery can be further studied by time series analysis.

As an important grey theory, the grey prediction method is widely used in the engineering field. Sun Tao et al. applied the grey system theory to the project cost and established the grey model to realize the estimate of the project cost with high precision. Xu hui et al. conducted a grey correlation analysis on the measured data of the dam and established the GM(1,1) model to predict the dam settlement on this basis. The results show that the method is effective. Li Peng and others. Based on the improvement of the traditional GM(1, 1) model, the grey relational model of GM(1, N) is put forward and applied to the capacity prediction of lead-acid batteries. Gu Weijun and so on. In view of the small sample of battery life data and the long test of battery life, the grey GM(1, 1) model is used to predict the cycle times when the battery reaches the specified life end value to estimate the battery lifetime.

4. Test Results and Analysis

4.1. Battery Capacity Degradation Data Analysis. Prediction is to estimate the future state of a product based on its past state of change. Specifically, RUL prediction mainly refers to the estimation of the remaining time from the current time to the final failure according to the monitored historical data of the product itself or similar products when the product runs to moment i (time here refers to the generalized time). Among them, historical data can be status monitoring data, failure time, maintenance time, or other event data. Given the status monitoring data of a product up to the current moment T , the method based on artificial intelligence is further subdivided into the statistical regression method and the similarity method.

Statistical data-driven RUL prediction methods can be divided into direct monitoring data-based RUL prediction method and indirect monitoring data-based RUL prediction method from the perspective of state monitoring data type. Direct state monitoring data is essentially the deterioration data of products. However, the literature does not consider the RUL prediction method combined with the historical data of similar products. This section summarizes the relevant RUL prediction methods from the perspective of degradation data, analyzes the advantages and disadvantages of each method, and discusses some problems worthy of study. Therefore, in this paper, the curve-grey model is used to analyze and predict the capacity degradation data, and the data are used to verify the model algorithm.

Using the grey mathematical model, a mathematical model of the capacity decay of lithium-ion batteries under constant current discharge, constant temperature, and constant depth of discharge is established. B5, B6, and B18 batteries were selected as samples to verify the model and algorithm. The discharge currents of B5, B6, and B18 batteries were all 2 A, and the three batteries carried out 168, 168, and 132 charge and discharge cycles, respectively, with discharge voltages of 2.7 V, 2.5 V, and 2.5 V.

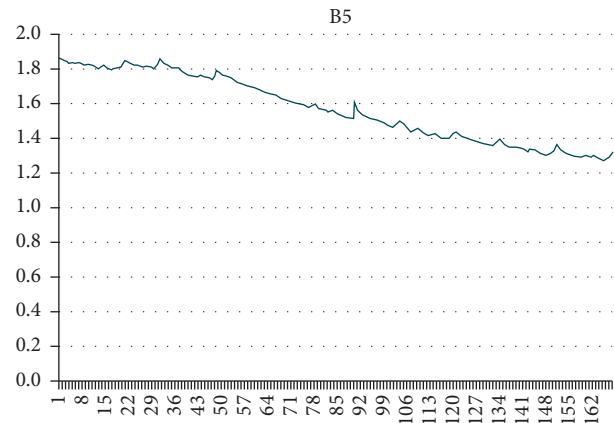


FIGURE 4: Capacity degradation of the B5 lithium battery.

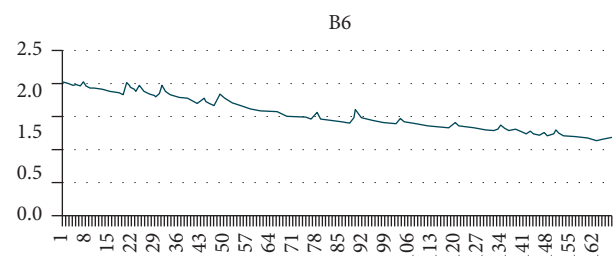


FIGURE 5: B6 lithium battery capacity degradation data.

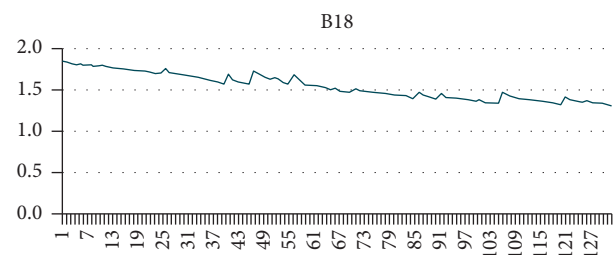


FIGURE 6: B18 lithium battery capacity degradation data.

As can be seen from Figures 4–6, the capacity of a single lithium-ion battery decreases gradually with the increase of charge and discharge cycles. Among them, B5 and B6 show an obvious linear decline trend, while B18 (after different processing) shows an exponential degradation trend during the first 10 charge and discharge cycles and then shows a linear degradation trend.

4.2. Analysis of Test Methods. In order to verify the performance of the mobile phone battery life prediction algorithm and prove its effectiveness, experimental data of different types of batteries under different experimental conditions would be selected in this paper for experimental analysis and evaluation of the grey data mining model construction method.

- (1) Firstly, using the measured data of four lithium-ion batteries of NASA, the relationship between constant discharge voltage and battery capacity before and

after Box-Cox conversion is analyzed, and the relationship between Box-Cox conversion parameters is also studied. And, based on GM(1, 1), the effectiveness and correctness of the method are verified by comparison.

- (2) Secondly, in order to verify the adaptability of the method, the battery test data of the GM(1,1) model is used to verify the adaptability of the method. In order to fully verify the performance of the proposed method throughout the experiment, the data of equal discharge voltage difference at three time intervals were used for experimental verification of each battery.

4.3. Analysis of Test Results

4.3.1. Grey Model Prediction Process. The following is the process for predicting the remaining life of lithium-ion batteries using the GM(1, 1) model.

Step 1. Original data of lithium-ion capacity degradation assuming cyclic charging and discharge is $C = \{C_i, I = 1, 2, \dots, N\}$, and the original data are, respectively, accumulated once:

$$C_i^{(1)}(k) = \sum_{i=1}^k C_i, \quad i = 1, 2, \dots, k. \quad (9)$$

Get a summation-generated sequence vector.

Step 2. Identify the unknown parameters in the model. Calculate the background value of the GM(1,1) model as follows:

$$Z(k) = 0.5 \times (C_K^{(1)} + C_{K-1}^{(1)}). \quad (10)$$

The function expression of GM(1,1) is

$$\frac{dc_i^{(1)}}{dt} + aC_i^{(1)} = b. \quad (11)$$

For the first-order grey model GM(1,1), the following equation can be obtained:

$$B_N a = Y_N. \quad (12)$$

Step 3. Solve the GM(1,1) model. The time response vector is

$$C_i^{(1)} = \left(C_0^{(1)} - \frac{b}{a} \right) \exp(-at) + \frac{b}{a}. \quad (13)$$

Step 4. Predict the remaining service life of the lithium-ion battery through the established grey model.

The capacity degradation model of the lithium-ion battery can be expressed as follows:

$$\hat{C}_{N+P} = \left(C_1 - \frac{\hat{b}_N}{\hat{a}_N} \exp(-\hat{a}_N(N+P)) \right) \exp(-\hat{a}_N - 1). \quad (14)$$

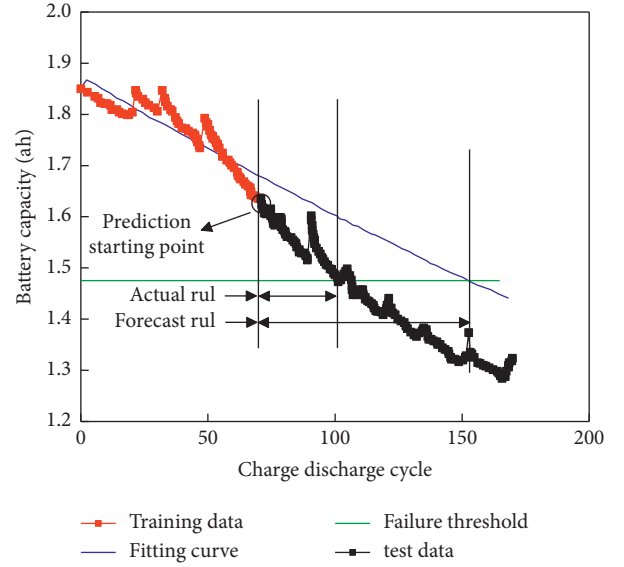


FIGURE 7: Battery prediction result (the first 70 volume data).

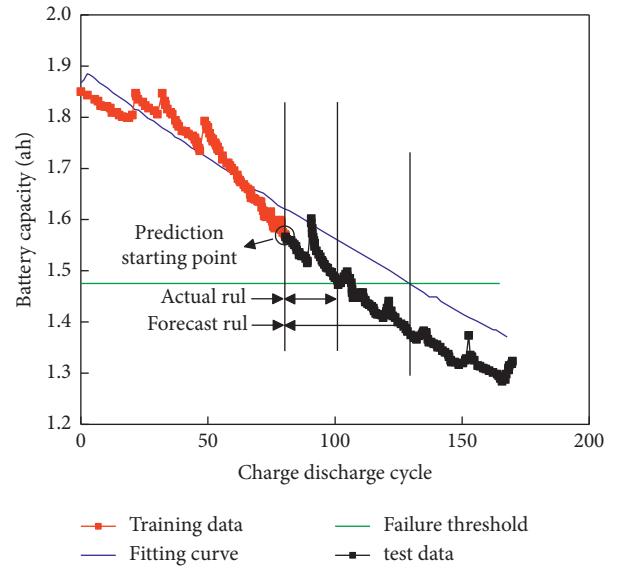


FIGURE 8: Battery prediction result (the first 80 volume data).

Also, the prediction accuracy of the proposed method was evaluated according to the relative error of the remaining service life prediction of lithium-ion batteries as follows:

$$\text{RUL error} = \frac{(\text{RUL true} - \text{RUL prediction})}{\text{RUL true}}. \quad (15)$$

4.3.2. Prediction Effect of the Discrete Grey Model. The following is an introduction to the feasibility of using NASA data to validate the grey model to predict lithium-ion battery RULs. Select the capacity degradation data for B6 and B18. The training data of the first 70 times, and the first 80 times

TABLE 1: Prediction errors.

Training data	The actual	RUL to predict	Error (%)
70	105	89	17.9
80	105	88	19.3

were selected, and the failure threshold was set to 80%. Figures 7 and 8 show the battery prediction results.

The feasibility of the algorithm using NASA data is shown below. B6 and B18 battery capacity degradation data are selected. The first 70 and 80 times of training data are selected, and the failure threshold is set to 80%. The prediction errors obtained are shown in Table 1.

5. Conclusion

Aiming at the life prediction of the lithium-ion battery in smartphone, a life prediction method based on grey data mining model is proposed, and the proposed method is verified and evaluated. In this paper, data mining and grey system theory are studied. Grey MDGM(1, 1) model is used as the life prediction method of the lithium-ion battery. Based on the above method, a grey data mining model is proposed to predict the degradation trajectory of lithium-ion batteries, which is nonlinear and random. The grey MDGM(1, 1) model has been used to describe the trend items in the degradation data to further improve the prediction accuracy. The battery test data for NASA PCoE is used to validate the model.

Due to the influence of time and equipment conditions, the research on the life prediction of lithium-ion batteries of smartphones in this paper is not sufficient. There are still the following problems, which need to be further improved in the future work:

- (1) In this paper, the life prediction methods of lithium-ion batteries are all about a single lithium-ion battery, while batteries in mobile phones are often in the form of battery packs, so battery packs should be considered in the future.
- (2) The data used in this paper are all obtained under the experimental conditions of constant current and constant voltage, while the actual working conditions of lithium-ion batteries on mobile phones are very complicated. Different charging and discharging mechanisms and different environments will affect its capacity degradation process. Therefore, capacity degradation data under different charging and discharging mechanisms should be used to verify the model in the future.

Data Availability

The data set can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] B. Y. Liaw and M. Dubarry, "Battery life prediction: ECS meeting abstracts," *IOP Publishing*, vol. 5, p. 328, 2006.
- [2] R. A. Dunstan, "Smart Battery Providing Battery Life and Recharge Time Prediction," 1996, <https://patents.google.com/patent/US5565759A/en>.
- [3] D. Rakhmatov, S. Vrudhula, and D. A. Wallach, "Battery lifetime prediction for energy-aware computing," in *Proceedings of the International Symposium on Low Power Electronics and Design*, IEEE, Monterey, CA, USA, August 2002.
- [4] H. Homayoun and G. N. Mills, "Wrist-worn ECG Monitor with Battery End of Life prediction," 1994, <https://patents.google.com/patent/US5289824A/en>.
- [5] Y. Wen, R. Wolski, and C. Krintz, "History-based, Online, Battery Lifetime Prediction for Embedded and Mobile Devices," in *Proceedings of the International Workshop on Power-Aware Computer Systems*, Springer, Cambridge, MA, USA, February 2003.
- [6] J. M. Kang, S. S. Seo, J. W. K. Hong, and W. James, "Personalized battery lifetime prediction for mobile devices based on usage patterns," *Journal of Computing Science and Engineering*, vol. 5, no. 4, pp. 338–345, 2011.
- [7] D. Z. Li, W. Wang, and F. Ismail, "A mutated particle filter technique for system state estimation and battery life prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 8, pp. 2034–2043, 2014.
- [8] W. Wilson, F. Ismail, and Z. Li, "A Mutated Particle Filter Technique for System State Estimation and Battery Life Prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, 2014.
- [9] F. Gao, L. I. Jianling, and S. Zhao, "Research Progress on Lithium-Ion Power Battery Life prediction," *Electronic Components and Materials*, vol. 24, 2009.
- [10] S. P. Perone and W. C. Spindler, "Battery lifetime prediction by pattern recognition. Application to lead-acid battery life-cycling test data," *Journal of Power Sources*, vol. 13, no. 1, pp. 23–38, 1984.
- [11] X. Zhang, Q. Miao, and Z. Liu, "Remaining useful life prediction of lithium-ion battery using an improved UPF method based on MCMC," *Microelectronics Reliability*, *PAGS*, vol. 75, no. Aug, pp. 288–295, 2017.
- [12] Y. Xing, E. Ma, K. L. Tsui, and M. Pecht, "A Case Study on Battery Life Prediction Using Particle Filtering," in *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing)*, pp. 1–6, Beijing China, May 2012.
- [13] D. D. Friel, "Battery management and life prediction," *Industrial Applications of Batteries*, Elsevier B.V, Amsterdam, The Netherlands, 2007.
- [14] A. Mohamed and W. Wang, "An Enhanced Mutated Particle Filter Technique for System State Estimation and Battery Life Prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, 2019.
- [15] H. Takatsuji, "Secondary Battery Lifetime Prediction Apparatus," *Battery System and Secondary Battery Lifetime Prediction method*, vol. 44, 2013.
- [16] P. M. Attia, K. A. Severson, and J. D. Witmer, "Statistical Learning for Accurate and Interpretable Battery Lifetime prediction," 2021, <https://arxiv.org/abs/2101.01885>.
- [17] P. G. Anselma, P. Kollmeyer, J. Lempert, and Z. Zhao, "Battery State-Of-Health Sensitive Energy Management of

- Hybrid Electric Vehicles: Lifetime Prediction and Ageing Experimental validation,” *Applied Energy*, vol. 285, 2021.
- [18] Q. Yue, “From Material Modeling to Li-Ion Battery Life Prediction - A Closer Look at the Degradation Mechanisms,” vol. 18, 2015.
- [19] P. Zhang, J. Liang, and F. Zhang, “An overview of different approaches for battery lifetime prediction,” *IOP Conference Series: Materials Science and Engineering*, vol. 199, Article ID 012134, 2017.
- [20] S. S. Mansouri, P. Karvelis, G. Georgoulas, and G. Nikolakopoulos, “Remaining useful battery life prediction for UAVs based on machine learning this work has received partial funding from the European union’s horizon 2020 research and innovation programme under the grant agreement No.644128, AEROWORKS,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 4727–4732, 2017.
- [21] B. Y. Liaw and D. D. Friel, “Battery Management and Life prediction,” *Industrial Applications of Batteries*, Elsevier B.V, Amsterdam, The Netherlands, 2007.
- [22] S. Voronov, E. Frisk, and M. Krysander, “Data-Driven Battery Lifetime Prediction and Confidence Estimation for Heavy-Duty Trucks,” *IEEE Transactions on Reliability*, vol. 67, pp. 1–17, 2018.
- [23] G. Wei, L. I. Zhipeng, and L. U. Tao, “Research on Battery Life Prediction Method Based on Markov model,” *Automation & Instrumentation*, vol. 36, 2019.
- [24] M. D’Arpino, M. Cancian, and A. Sergent, “A Simulation Tool for Turbo-Hybrid-Electric Aircraft Battery Life Prediction for the NASA ULI Program,” in *Proceedings of the 2019 AIAA/IEEE Electric Aircraft Technologies Symposium (EATS)*, IEEE, Indianapolis, IN, USA, August 2019.
- [25] D. Wang, F. Yang, Y. Zhao, and K. L. Tsui, “Battery remaining useful life prediction at different discharge rates,” *Microelectronics Reliability*, vol. 78, no. nov, pp. 212–219, 2017.
- [26] P. G. Anselma, P. Kollmeyer, J. Lempert, Z. Zhao, G. Belingardi, and A. Emadi, “Battery state-of-health sensitive energy management of hybrid electric vehicles: lifetime prediction and ageing experimental validation,” *Applied Energy*, vol. 285, Article ID 116440, 2021.
- [27] D. Ivanov, K. G. Larsen, S. Schupp, and J. Srba, “Analytical Solution for Long Battery Lifetime Prediction in Nonadaptive Systems,” *Quantitative Evaluation of Systems*, Springer, New York, NY, USA, 2018.
- [28] S. Voronov, E. Frisk, and M. Krysander, “Data-driven Battery Lifetime Prediction and Confidence Estimation for Heavy-Duty Trucks Ieee Transactions on Reliability 1 Data-Driven Battery Lifetime Prediction and Confidence Estimation for Heavy-Duty trucks,” *IEEE Transactions on Reliability*, vol. 67, 2019.
- [29] A. Li, H. Fang, and G. Yu, “Research on Data-Driven Life Prediction Methods of Satellite Lithium-Ion Battery,” *Computer Measurement & Control*, vol. 39, 2015.

Research Article

A Novel Stress State Assessment Method for College Students Based on EEG

Li Liu ^{1,2}, Yunfeng Ji,¹ Yun Gao,¹ Tao Li,¹ and Wei Xu¹

¹Jiangsu Vocational College of Information Technology, Wuxi, Jiangsu 214153, China

²Jiangsu Key Laboratory of Media Design and Software Technology (Jiangnan University), Wuxi 214122, China

Correspondence should be addressed to Li Liu; liul2@jsit.edu.cn

Received 4 April 2022; Revised 22 May 2022; Accepted 23 May 2022; Published 7 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Li 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.

Stress is an unavoidable problem for today's college students. Stress can arouse strong personal emotional and behavioral responses. Compared with other groups of the same age, college students have a special way of life and living environment. They have complex interpersonal relationships and relatively weak social support systems. At the same time, they also face fierce competition in both academic and employment. However, they lack the skills to deal with the crisis and are reluctant to ask others for help, which leads to a simultaneous increase in mental stress. The pressure on college students mainly comes from study, family, social, employment, society, and economy. When students face multiple pressures from family, school, society, etc., some students are prone to some psychological problems due to their own personality or external environment and other reasons. Therefore, regular assessment of students' stress status is an important means to prevent college students' psychological problems. Considering that in real life, the number of students whose pressure is within the tolerable range is the majority, while the number of students who are under too much pressure is a minority. Therefore, the actual dataset to be identified belongs to a kind of imbalanced data. In this study, an improved extreme learning machine (IELM) is used to improve the performance of the recognition model as much as possible. IELM takes the idea of label weighting as the starting point, introduces the AdaBoost algorithm, and combines its weight distribution with the label weighted extreme learning machine (ELM). During the weight update process, the advantage of the imbalanced nature of multi-label datasets is taken. IELM was used to classify EEG data to determine the stress level of college students. The experimental results demonstrate that the algorithm used in this study has excellent classification performance and can accurately assess students' stress levels. The accurate assessment of stress has provided a solid foundation for the development of students' mental health and has significant practical implications.

1. Introduction

College students play an important role in the development and construction of various projects across the country. Their mental health has an impact not only on individual growth and education but also on the country's and society's long-term stability. However, in recent years, college students have experienced psychological crises on occasion, and the subject has grown in prominence, attracting widespread attention from society and university student employees. To address the psychological issues that college students face, we must first comprehend their stress levels. Second, according to the stress condition, the source of psychological stress in college students is precisely evaluated, and the reasons for psychological stress in college students are

thoroughly examined. Finally, corresponding intervention strategies are proposed. This is the main connotation of college students' work. Psychological stress is a person's physiological changes and emotional fluctuations caused by changes in the external environment and the stress response of the body. The transition of college students from high school to university is an important turning point from school to society. There are all kinds of stress involved in adapting to a new environment, learning new courses, facing new challenges, and developing new relationships. The sources of psychological pressure on college students mainly include the following aspects: first, the psychological pressure caused by role change and adaptation disorders; second, the learning pressure caused by the change in learning style and weak learning motivation; third, the interpersonal

pressure caused by communication difficulties and weak communication skills; fourth, the economic pressure and mental pressure caused by the family's economic difficulties; fifth, the employment pressure caused by the severe employment situation and the fierce competition for talents; sixth, the psychological pressure caused by lack of sexual knowledge and immature concept of love; seventh, the psychological pressure caused by personality and emotional problems; and eighth, psychological pressure caused by personality and emotional problems. The psychological stress of college students directly affects their learning effect and quality of life during their school days. For schools, it is about whether they can produce graduates with excellent mental health and professional quality. In the research on the stress of college students, some scholars believe that college students are a high-stress group. Other studies believe that college students only feel less stress, but the sense of stress and the way of coping with stress are indeed the most important factors that cause college students' mental health problems. If psychological pressure cannot be relieved in time, it may lead to high blood pressure and cardiovascular disease and endanger physical health. Depression, pessimism, and hopelessness accompanying psychological pressure will affect students' academic performance and interpersonal communication. This will lead to college students' life satisfaction and happiness. It can even lead to vicious incidents such as wounding and suicide.

Most of the methods of psychological stress assessment use objective scales. Reference [1] compiled a social readjustment scale (SRRS) containing 43 items, which opened the first instance of using questionnaires to measure stress. Subsequently, some researchers have compiled some stress questionnaires with relatively high reliability and validity, such as the Student Stress Scale prepared in [2] and the Graduate Stress Scale prepared in [3]. Commonly used psychological scales are perceived stress scale (PSS) [4], relative stress scale (RSS) [5], psychological stress measure (PSM) [6], and so on. When the method of filling in the scale is used to evaluate the psychological stress state of the subjects, a large-scale test can be administered in a short period of time, and the test results can be obtained quickly. Compared with other methods, it has the advantages of high efficiency and good scientificity. However, many objective scales used in China are compiled based on the psychological characteristics of Westerners. However, the psychological characteristics of social approval, default tendency, and strong conformity in Chinese people's response to the scale have led to a considerable proportion of students' answers with false elements in the scale test. This affects the authenticity and validity of the questionnaire. Therefore, it is difficult to truly evaluate their psychology and behavior through questionnaires. Due to the shortcomings of the scale test method and the complexity of the respondents' motivation to answer the questions, many scales have already added a certain number of fraud identification questions during the compilation process.

To find a more objective stress assessment method, many scholars began to study stress assessment methods based on physiological signals. Studies have shown that there is indeed

a close relationship between psychological stress and brain activity [7]. In particular, when negative emotions appear, the activity of the right hemisphere of the brain is abnormally active [8]. Therefore, scholars have begun to boldly hypothesize that the relative activity of EEG on the right side of the brain predicts changes in psychological stress and a higher risk of mental illness [9]. Based on this conclusion, many EEG-based stress recognition studies have appeared one after another. Reference [10] studied a number of subjects who faced examination pressure and found that the subjects' right forehead EEG activity was more intense under high-intensity examination pressure. Reference [11] studied the stability of prefrontal EEG asymmetry for detecting psychological stress and depression levels, and the results showed that resting EEG α wave asymmetry could be used as a reliable indicator for detecting stress and depression levels. However, some literature studies also pointed out that gender and age differences have a greater impact on the study of psychological stress [12]. Reference [13] proposed that gender differences in prefrontal asymmetry and negative emotion processing may be related to human genes.

To sum up, we found that the existing researches on stress state recognition are based on different types of data such as scales, EEG signals, ECG signals, speech signals, video signals, and facial expressions. The evaluation models used are also different for different data types. Common models are mainly based on machine learning [14–16] and deep learning [17–19]. Because the physiology is more realistic, this study mainly chooses the data based on the physiological signal to identify the stress state. Among the physiological data, EEG and ECG are the most common. Considering that the ECG signal acquisition equipment requires multiple electrodes to collect signals, even a single-lead wearable product is not very convenient. In contrast, EEG signals are easier to acquire, and the signals are sensitive, which can quickly reflect changes in pressure. Therefore, this study chooses EEG as the data type used in the research. In the selection of the recognition model, considering that although the recognition rate of the deep learning algorithm is better, its model training time and high requirements for the hardware performance of the device are high, and the model has a lot of parameters. In this study, the theory is simple and easier to implement the machine learning algorithm. Considering that in real life, the number of students whose pressure is within the tolerable range is the majority, while the number of students who are under too much pressure is a minority. Therefore, the actual dataset to be identified belongs to a kind of imbalanced data. In this study, IELM is used to improve the performance of the recognition model as much as possible. IELM takes the idea of label weighting as the starting point, introduces the AdaBoost algorithm, and combines its weight distribution with the label weighted extreme learning machine. During the weight update process, the advantage of the imbalanced nature of multi-label datasets is taken. IELM was used to classify EEG data in order to determine the stress level of college students. The experimental results demonstrate that the algorithm used in this study has excellent classification performance and can accurately assess students' stress levels.

TABLE 1: Brain wave band details.

Name	Frequency (Hz)	Location	Generated time
δ	0.5 < 4	Forehead in adults, back of brain in children	Occurs mostly in the brains of infants, but also occurs when adults are in deep sleep, coma, or anesthesia
θ	4–7	Brain regions unrelated to hand function	Occurs in young children and adolescents, but also in adults who are tired but conscious
α	8–15	The back of the brain, the resting state is concentrated in the center	Relaxed/contemplative state with eyes closed
β	16–31	The brain is symmetrically distributed on both sides, with a prominent forehead	Positive thinking, focus, vigilance, anxiety
γ	32–45	Somatosensory cortex	Short-term memory, hearing, and touch, multisensory processing

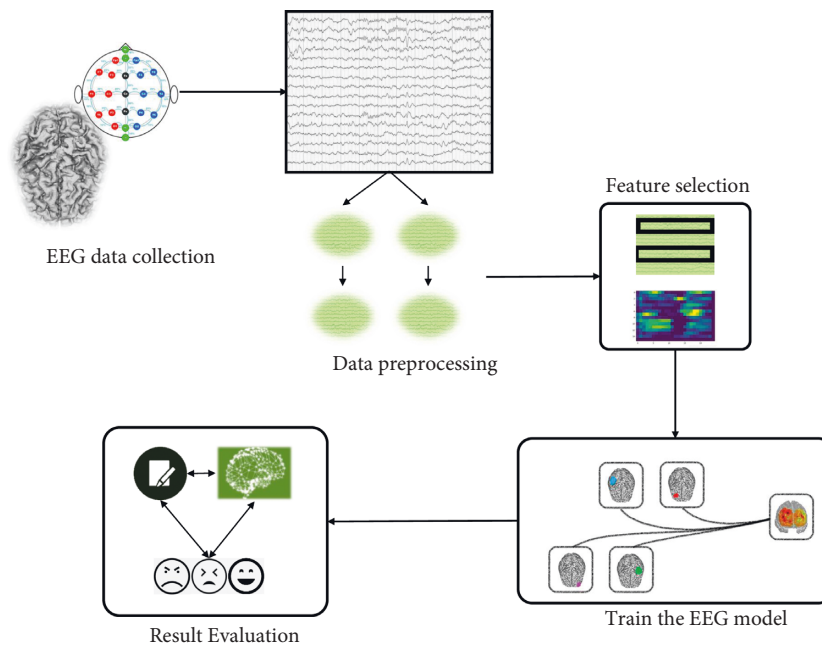


FIGURE 1: EEG-based stress assessment process.

2. EEG-Based Stress Detection

2.1. The Relationship between EEG and Students' Stress State.

Stress, pleasure, focus, etc., are all related to human emotional thinking. The brain controls human thoughts and emotions. When the brain deals with emotion-related issues, the amygdala, orbitofrontal cortex, anterior cingulate cortex, insula, nucleus accumbens, thalamus, and ventral tegmental area of the brain all respond differently to specific emotions. For example, the anterior cingulate cortex responds when we make a decision about something in a happy or sad emotional state. The nucleus accumbens becomes active when we anticipate a reward or something good will happen. When something disgusts us, the insula becomes very active. When these areas are stimulated and activated, neurons in the cerebral cortex are stimulated to generate action potentials. When the cells are stimulated by the impulse, the pyramidal cells will be depolarized, forming a potential difference and generating an electric current. After the electrodes are placed on the scalp, and after amplification, the EEG signal can be collected and an EEG can be drawn. People divide EEG

signals into δ , θ , α , β , and γ frequency bands according to different frequency bands. The descriptions of brain waves in each frequency band are shown in Table 1.

Research has shown that when we blink or think, α waves instantly disappear and reappear, and β waves become very active and have high amplitudes. γ is very sensitive to emotional changes and cognitive learning. When we are tired and sleepy, θ waves begin to appear and become active. To sum up, it can be found that the generation and activity of different brain waves can intuitively reflect the current activity state of students. Therefore, by studying the brain wave signals of these frequency bands, we can analyze the emotional state of college students in the process of study and life and then analyze the stress state of students.

2.2. EEG-Based Stress Assessment. Figure 1 shows the stress assessment process based on EEG signals.

As shown in Figure 1, the first step is to collect EEG data. EEG data collected are all from the freshman to third-year students of our school. Since EEG is a very weak

TABLE 2: Linear and nonlinear characteristics of EEG.

Feature type	Feature name
Linear feature	Full-band center frequency, Hjorth parameter, peak-to-peak, variance, slope, kurtosis
Nonlinear feature	CO complexity, correlation dimension, power spectral entropy, full-band power spectral entropy, Shannon entropy, Kolmogorov entropy

physiological electrical signal, its amplitude can generally only reach the order of microvolts. Therefore, during the acquisition process, various noises generated by different reasons such as the surrounding environment, eye movement signals, and EMG signals are easily mixed into the collected EEG, resulting in poor experimental results. Therefore, to ensure EEG's purity, it is often necessary to perform preprocessing such as denoising the EEG before analyzing the EEG data.

To reduce the data dimension and facilitate subsequent classification processing, feature extraction operations on the preprocessed data are usually required. At present, the research on EEG signal features mainly focuses on linear features, such as extracting the frequency, power, and other features of EEG signals by means of the Fourier transform. However, the structure of the human brain is more complex, and nonlinear features often have better performance. Generally, the linear and nonlinear characteristics are shown in Table 2.

A stress state recognition model is trained based on the EEG training dataset. The algorithm used is the IELM algorithm given in Section 3. The performance of the trained model is validated using the test dataset. The higher the classification accuracy of the sample, the better the model.

3. Imbalanced Extreme Learning Classification Algorithm

3.1. Weighted Extreme Learning Algorithm. For class-balanced tasks, weighted extreme learning is particularly successful. However, it has two drawbacks. First, as the size of the training set grows, so does the time complexity. The second issue is a lack of mistake compensation flexibility. Reference [20] developed a marker-weighted extreme learning machine based on the concept of cost-sensitive learning to address the drawbacks of this technique. By increasing the expected output value of minority class labels, label-weighted extreme learning machines improve the training error tolerance of minority class cases. Furthermore, because it does not use a weight matrix in the optimization method, it has the same time complexity as standard ELM.

$$\begin{aligned}
 L_{ELM} &= \frac{1}{2} \|\beta\|^2 + \gamma \frac{1}{2} W \sum_{i=1}^N \|\tau_i\|^2, \\
 h(x_i)\beta &= t_i^T - \tau_i^T, \\
 i &= 1, 2, \dots, n.
 \end{aligned} \tag{1}$$

Assume that the expected matrix T has m rows and n columns, where m is the number of categories and n is the number of training set samples. The penalty factor is γ . The

output layer weight to be solved is denoted by β . W is an n -dimensional diagonal matrix. Each diagonal element's value corresponds to the penalty factor regulation parameter of the corresponding sample.

$\tau_i = [\tau_{i1}, \tau_{i2}, \dots, \tau_{im}]$ represents the training error vector corresponding to the sample x_i on all output nodes. The expression for T is as follows:

$$T = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ t_{21} & t_{22} & \cdots & t_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mN} \end{bmatrix}. \tag{2}$$

For the setting of label weights in two-class and multi-class classification problems, [20] provides two weight distribution methods, as follows:

$$\begin{aligned}
 t_{ij} &= \begin{cases} \Delta_{\text{major}}(\text{num})/\Delta(\text{num}_i), & \text{if } x_j \in \text{num}_i, \\ -1, & \text{if } x_j \notin \text{num}_i, \end{cases} \\
 t_{ij} &= \begin{cases} \sqrt[2]{\Delta_{\text{major}}(\text{num})/\Delta(\text{num}_i)}, & \text{if } x_j \in \text{num}_i, \\ -1, & \text{if } x_j \notin \text{num}_i. \end{cases}
 \end{aligned} \tag{3}$$

where $\Delta(\text{num}_i)$ is the number of samples from the i th class in the training set and $\Delta_{\text{major}}(\text{num})$ is the number of samples from the majority class. Two weight distribution methods are provided:

$$\begin{aligned}
 t_{ij} &= \begin{cases} \sim \Delta_{\text{major}}(\text{num})/\Delta(\text{num}_i), & \text{if } x_j \in \text{num}_i, \\ -1, & \text{if } x_j \notin \text{num}_i, \end{cases} \\
 t_{ij} &= \begin{cases} \sqrt[2]{\sim \Delta_{\text{major}}(\text{num})/\Delta(\text{num}_i)}, & \text{if } x_j \in \text{num}_i, \\ -1, & \text{if } x_j \notin \text{num}_i, \end{cases}
 \end{aligned} \tag{4}$$

The weights of the majority class samples remain unchanged in the above weight distribution method, while the weights of the minority class samples are increased. The greater the class imbalance ratio, the greater the weight ratio between the minority and majority classes. The following steps are taken by the weighted extreme learning (Algorithm 1).

3.2. AdaBoost Algorithm. In ensemble learning, according to the different ways of generating base learners, it can be divided into serial ensemble learning algorithms represented by boosting [21] and parallel ensemble learning algorithms represented by bagging [22]. Among the boosting learning paradigms, the AdaBoost algorithm is the most famous, originally proposed by Freund and Schapirel in 1997.

Input: Training set X_1 , test set X_2
Output: Weight β
Step 1: Calculate the number of samples in each category in X_1
Step 2: Create the initial expectation matrix T
Step 3: For x_i in X_1
Step 4: Determine the corresponding t_{ij} in T
Step 5: Calculate t_{ij} according to the above formula
Step 6: Generate hidden layer parameters at random
Step 7: Using the hidden layer parameters and the training set X_1 , compute the hidden layer output matrix H
Step 8: Calculate the output weight β of the weighted extreme learning machine by equation (4)

ALGORITHM 1: LW-ELM algorithm.

Input: Training set X , number of base classifiers M , base classifier H , test sample X'
Output: Class label Y' of test sample X'
Step 1: For $i = 1 : M$
Step 2: Train the base classifier H_i on X_i
Step 3: Calculate the weighted error by equation (5)
Step 4: If the weighted error is greater than 0.5, update all sample weights through equation (7)
Step 5: Calculate the weights of the base classifier H_i by equation (6)
Step 6: Calculate the classification labels of the test samples: $y' = \arg \max \sum_{i=1}^n c_i F(H_i(x') == y)$, $y \in Y$

ALGORITHM 2: AdaBoost algorithm.

Suppose X is a set containing n training samples for training an AdaBoost classifier. In the AdaBoost algorithm, the importance of each base classifier H_i depends on its error rate ε_i :

$$\varepsilon_i = \frac{1}{n} \left[\sum_{j=1}^n w_j F(H_i(x_j) \neq y_j) \right], \quad (5)$$

where F is the indicator function. If the condition is true, it takes the value 1; otherwise, it is 0. Through the above formula, the importance weight of the base classifier H_i is given as follows:

$$c_i = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_i}{\varepsilon_i} \right). \quad (6)$$

It can be seen that when the error rate c_i is close to 0, $w_i^{(j)}$ corresponds to a large positive value. When the error rate is close to 1, $w_i^{(j)}$ corresponds to a large negative value.

$$\begin{aligned} w_i^{(j+1)} &= \frac{w_i^{(j)}}{S_j} \times e^{-c_j L_j(x_i) y_i} \\ &= \frac{w_i^{(j)}}{S_j} \times \begin{cases} e^{-c_j} & \text{if } H_j(x_i) = y_i, \\ e^{c_j} & \text{if } H_j(x_i) \neq y_i, \end{cases} \end{aligned} \quad (7)$$

where S_j is a normalization factor to ensure that $\sum_i w_i^{(j)} = 1$.

Through the above weight update formula, it is possible to increase the weight of the misclassified samples in the previous round and reduce the weight of those that have been correctly classified. The AdaBoost algorithm's execution steps are as follows (Algorithm 2).

3.3. *IELM*. Reference [23] weights the tokens to increase the expected output of minority class samples. By improving the training error tolerance of the minority class samples, the overall training error of the minority class is similar to that of the majority class in the global training error. In this way, the class imbalance problem is solved. Since the label weight only changes the expected output size and does not change the original optimization formula of the extreme learning machine, the time complexity does not change. However, this algorithm also has certain shortcomings. For example, the algorithm uses human experience to set the weights, which lacks flexibility. The core idea of the AdaBoost algorithm is to change the distribution of samples by modifying the weight of the samples, so that the classifier gradually focuses on those samples that are easy to be misclassified. In this way, the quality of the classification model is maximized. In the label weighting process, larger weights are often assigned to those important samples to avoid misclassification. The above two strategies have the same idea. Based on the inspiration of this idea, this study combines AdaBoost with a labeled weighted extreme learning machine. To efficiently deal with multi-label imbalanced data, the algorithm used improves the two key aspects of AdaBoost's initial weight setting and weight distribution update. The improved model always considers the inherent imbalance characteristics of multi-label and adjusts the size of the weights directionally.

3.3.1. *Adjust the Initial Weight*. For single-label classification problems, the weight distribution reflects the relative importance of the samples. Training samples that are often misclassified tend to receive larger weights than correctly

classified samples. For multi-label classification problems, the weight distribution can reflect the relative importance of the labels. Therefore, the object of weight setting is the marker. The traditional weight setting method is to use the method of evenly distributing the weight. If the data distribution is unbalanced, the weight can be increased by updating the weight in the iterative process and adjusting the minority class adaptively. In fact, the unbalanced degree of the data can be considered. Giving the minority class a higher weight and the majority class a lower weight will inevitably cause the model to converge faster.

Assuming that the multi-label data $S = \{(x_i, Y_i) | i = 1, 2, \dots, n\}$ are a training set containing n samples, $V = \{v_j | 1, 2, \dots, c\}$ represents a label set with c categories, where x_i is the feature vector of the i th sample, and $Y_i \subseteq V$ is its associated label set.

Reference [24] proposed an intra-marker imbalance measure, which has been widely used. For the j th marker, $S_j^+ = \{(x_i, +1) | y_j \in Y_i, 1 \leq i \leq n\}$ represents the positive class sample, $S_j^- = \{(x_i, -1) | y_j \notin Y_i, 1 \leq i \leq n\}$ represents the negative class sample, and the imbalance rate of the j th marker is as follows:

$$R_j = \frac{\max(|S_j^+|, |S_j^-|)}{\min(|S_j^+|, |S_j^-|)}. \quad (8)$$

The initial weight is set as an asymmetric matrix W according to the imbalance ratio, and W contains n rows and c columns, representing n samples and c markers, respectively. The j th token value of the i th sample is as follows:

$$t_{ij} = \begin{cases} \sqrt{(R_j)}, & \text{if } x_i \in \text{num}_j, \\ 1, & \text{if } x_i \in \text{num}_j, \end{cases} \quad (9)$$

$$W_{ij} = \frac{t_{ij}}{p_i}, \quad (10)$$

where p_i is the normalization factor, and its function is to constrain the sum of the weights of the markers to be 1. The square root of R is chosen here because the imbalance ratio in the multi-label is relatively high, and direct use will cause the weight of positive samples to be much larger than the weight of negative samples, making the model fall into the other extreme.

3.3.2. Weight Update. In this study, the purpose of setting asymmetric weight distribution is to make AdaBoost always focus on the imbalance problem within the label. In the iterative process, if the weights are updated in the usual way, the model cannot keep focusing on the multi-label imbalance problem. Therefore, this study updates the weights separately for each category of each token. For the l th label, its j th class error rate is calculated as follows:

$$\varepsilon = \sum_{x_i \in \text{class } j: H(x_i) \neq Y_i^l} S_t(x_i), \quad (11)$$

where H represents the base classifier used by the model and $H(x_i)^l$ represents the output of the base classifier on the l th token of the sample instance x_i .

$$\zeta^{lj} = \frac{1}{2} \ln \left(\frac{\sum_{x_i \in \text{class } j: H(x_i)^l = Y_i^l} S_t(x_i)}{\sum_{x_i \in \text{class } j: H(x_i)^l \neq Y_i^l} S_t(x_i)} \right). \quad (12)$$

When the error rate is close to 0, ζ corresponds to a large positive value. When the error rate is close to 1, ζ corresponds to a large negative value. The calculation formula of the weight W of the $t+1$ th round is as follows:

$$W_{t+1}^{lj}(x_i) = \frac{W_t(x_i) \exp(-\zeta_t^{lj} F(H(x_i), j))}{p_t^{lj}}, \quad (13)$$

where $F(\cdot)$ is an indicator function, and the role of p_t^{lj} is to ensure $\sum W_{t+1}^{lj}(x_i) = 1/2q$ at this time. The weight calculation formula of the entire trainer is as follows:

$$\zeta = \frac{1}{2} \ln \left(\frac{\sum_l \sum_{i: H(x_i)^l = Y_i^l} S_t(x_i)}{\sum_l \sum_{i: H(x_i)^l \neq Y_i^l} S_t(x_i)} \right). \quad (14)$$

3.3.3. IELM Algorithm Execution Steps. The IELM algorithm's execution steps (Algorithm 3) are as follows.

4. Experiment

4.1. Experimental Data Collection. To ensure that the experiment is not affected by factors such as the subject's physical health, other disturbances are minimized as much as possible. Therefore, before this experiment, we learned about the physical condition and basic information of each subject by means of a questionnaire, including the subject's name, age, gender, physical condition, academic status, family status, and the relationship with classmates and teachers. In the questionnaire, subjects are required to check the stress self-evaluation items. The table design of the questionnaire is shown in Table 3.

To better collect EEG without external interference, subjects were first allowed to fall asleep in a quiet environment. When the subject entered deep sleep, the subject's EEG was collected. A total of 35 questionnaire data and EEG data were collected from freshmen to junior college students. These data are simply screened according to the questionnaires filled in by the subjects. For example, when there is a large deviation between the subjects' emotions and the expected emotions of the videos watched, or the subjects are not in the state at all when watching the videos, this set of data will be discarded. Finally, 90 valid samples and 360 EEG fragments were screened out.

4.2. Multifeature Combination Experiment. There are many features of EEG data. This study mainly extracts four features: Hurst index (P1), volatility index, sample entropy, and permutation entropy. The collected EEGs were classified

Input: Training set X , test set X'
Output: Classification evaluation index value
Step 1: Initialize the weight W_1 using equations (9)-(10)
Step 2: For $i = 1 : M$
Step 3: Train base classifier H_i on W_i
Step 4: Calculate the overall error rate for all markers using equation (11)
Step 5: When the weighted error is greater than 0.5, update all sample weights using equation (13), otherwise the loop stops
Step 6: Calculate the weights of the base classifier H_i using equation (14)
Step 7: Calculate the class label of the test sample $y' = \text{sign}(\sum_{i=1}^M c_i H_i(x'))$
Step 8: Calculate the evaluation index value according to the evaluation index calculation formula

ALGORITHM 3: IELM algorithm.

TABLE 3: Stress self-assessment form.

Subject number				
Gender	Male <input type="checkbox"/>			Female <input type="checkbox"/>
Grade	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	
Study stress	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	
Life pressure	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	
Family stress	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	
Overall pressure	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	

TABLE 4: Classification accuracy based on ELM under four features.

Feature	α	β	γ	δ	θ
Hurst index	0.7122	0.6945	0.7342	0.5898	0.6342
Fluctuation index	0.8098	0.8120	0.7788	0.7556	0.7861
Sample entropy	0.6541	0.5987	0.5062	0.6012	0.5142
Permutation entropy	0.7193	0.6298	0.5865	0.7181	0.5880

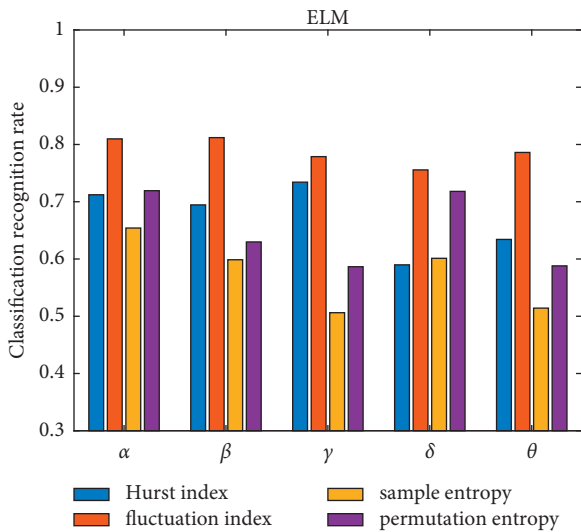


FIGURE 2: Comparison chart of classification accuracy based on ELM under four features.

using radial basis neural network (RBFNN) and IELM. The number of categories is 3, which are high stress, average stress, and low stress. The two algorithms run 10 times on each rhythm to get the average classification accuracy.

TABLE 5: Classification accuracy based on IELM under four features.

Feature	α	β	γ	δ	θ
Hurst index	0.7778	0.7624	0.7997	0.6175	0.7037
Fluctuation index	0.8134	0.8878	0.7959	0.7602	0.8178
Sample entropy	0.7190	0.6530	0.5368	0.6091	0.5392
Permutation entropy	0.7327	0.6490	0.5997	0.7204	0.5914

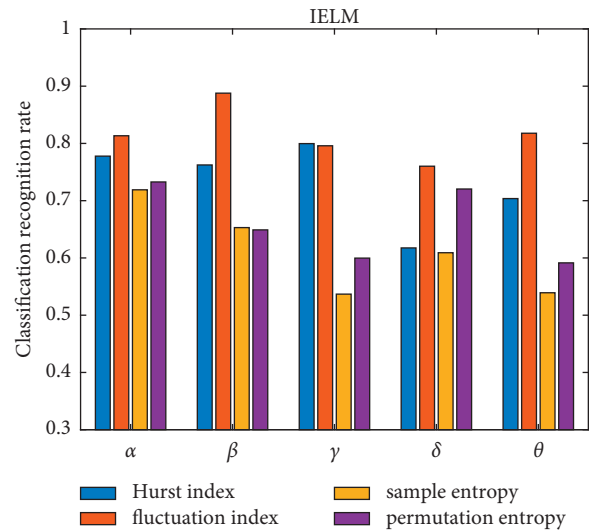


FIGURE 3: Comparison chart of classification accuracy based on IELM under four features.

Table 4 and Figure 2 show the experimental results based on the ELM algorithm. Table 5 and Figure 3 show the experimental results based on the IELM algorithm.

Figures 2 and 3 show that, regardless of classifier, the fluctuation index has the best classification effect, followed by the Hurst index and permutation entropy, and the worst is sample entropy. This demonstrates that different types of features contribute to the classification results at different rates. Figure 4 depicts the classification accuracy of various classifiers from four different perspectives.

As can be seen from Figure 4, under a single feature, the classification accuracy of the IELM algorithm under a single feature is higher than that of the ELM. To resolve the optimal combination of multiple features to improve the classification

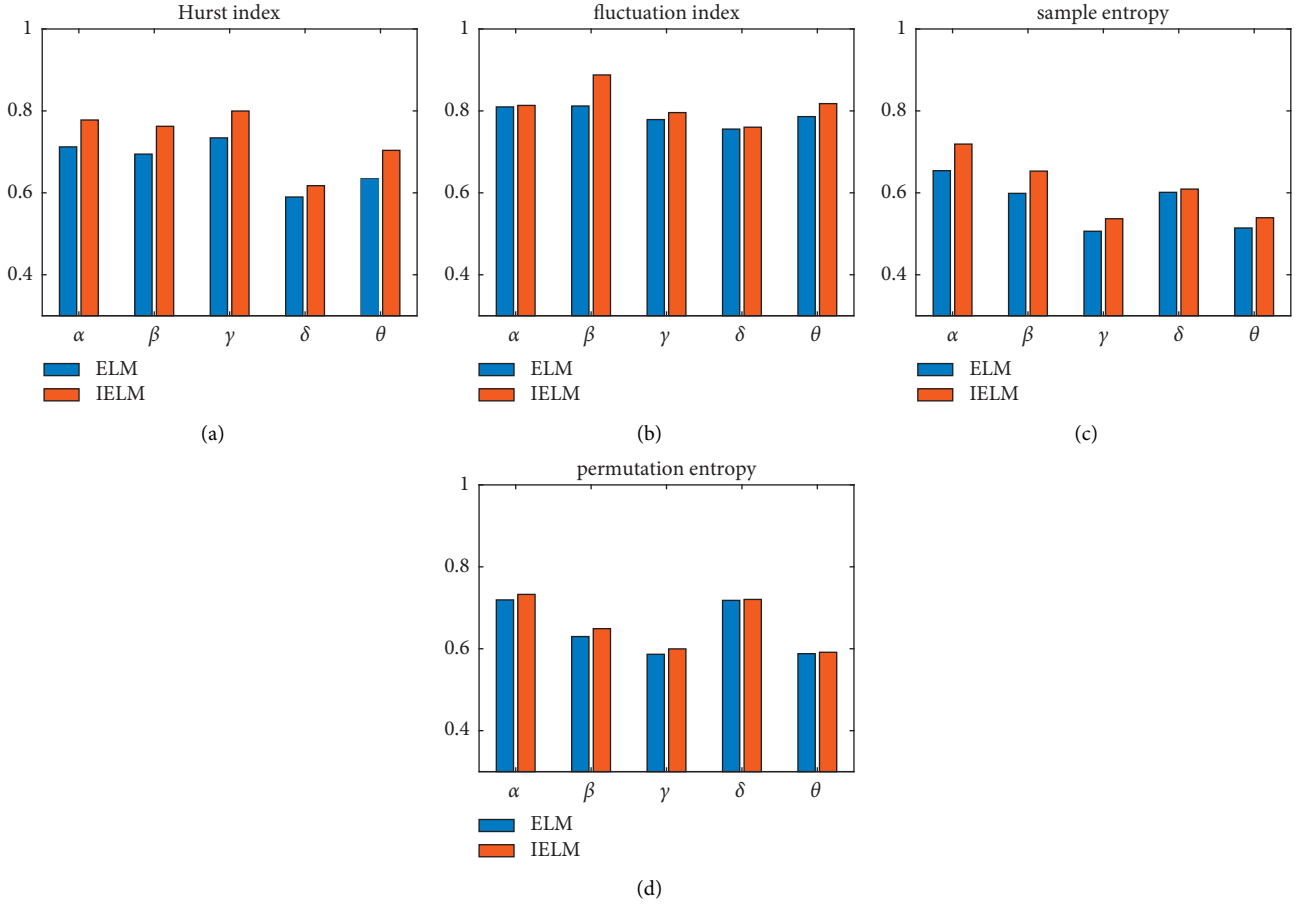


FIGURE 4: Classification accuracy of 4 features. (a) Hurst index. (b) Fluctuation index. (c) Sample entropy. (d) Permutation entropy.

recognition rate of EEG, in this study, a linear weighting formula is given to obtain the weight of each EEG feature. The weight calculation formula is as follows:

$$w_i = \frac{f_i}{f_1 + f_2 + f_3 + f_4}, \quad (15)$$

$$\text{s.t. } 1 \leq i \leq 4, 0 \leq w_i \leq 1,$$

where f_i is the classification recognition rate based on the i th EEG feature and w_i is the weight of the feature f_i . The data in Table 5 are substituted into equation (15), and the calculated feature weights are shown in Table 6.

The weights of the combined features under each frequency band can be obtained from the data in Table 6, and the feature combination formulas under 5 frequency bands are as follows:

$$F_\alpha = 0.2556P1 + 0.2673P2 + 0.2363P3 + 0.2408P4, \quad (16)$$

$$F_\beta = 0.2582P1 + 0.3007P2 + 0.2212P3 + 0.2198P4, \quad (17)$$

$$F_\gamma = 0.2927P1 + 0.2913P2 + 0.1965P3 + 0.2195P4, \quad (18)$$

$$F_\delta = 0.2281P1 + 0.2808P2 + 0.2250P3 + 0.2661P4, \quad (19)$$

TABLE 6: Details of each feature weight.

Feature	α	β	γ	δ	θ
Hurst index	0.2556	0.2582	0.2927	0.2281	0.2653
Fluctuation index	0.2673	0.3007	0.2913	0.2808	0.3084
Sample entropy	0.2363	0.2213	0.1965	0.2250	0.2033
Permutation entropy	0.2408	0.2198	0.2195	0.2661	0.2230

$$F_\theta = 0.2653P1 + 0.3084P2 + 0.2033P3 + 0.2230P4. \quad (20)$$

It can be seen from the experimental data in Table 6 that the weights of the Hurst index and fluctuation index are higher than the remaining two features in the five rhythm species. This shows that these two features show more superior performance in the task of stress assessment. The EEG data of the two feature combinations were classified separately using IELM. Table 7 and Figure 5 give the classification results.

It is clear from the experimental results shown in Table 7 and Figure 5 that for the same classification model, the accuracy of the weighted feature combination is significantly higher than that of the unweighted feature combination. This shows that the introduction of feature weighting strategy can improve the classification performance of EEG. The reason is that different types of features have different

TABLE 7: Classification accuracy under different feature combinations.

Feature combination	α	β	γ	δ	θ
Without weighted feature combination	0.8210	0.8932	0.8090	0.7881	0.8012
With weighted feature combination	0.8340	0.9001	0.8559	0.8293	0.8749

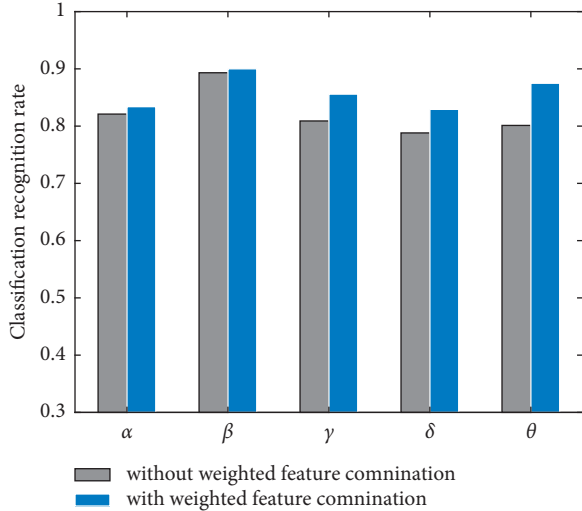


FIGURE 5: Comparison chart of classification accuracy under different feature combinations.

degrees of activity for classification. Based on the weighting formula given in this study, the features with high activity are given a large weight, and the features with low activity are given a small weight, so that the result is optimal.

4.3. Classification Model Experiment. Following the selection of the feature combination method, multiple comparison models are introduced to validate the performance of the classification model used in this study. They are support vector machine (SVM), linear support vector machine (linear SVM), radial basis neural network (RBFNN), random forest (RF), and ELM. Since a single indicator of recognition accuracy cannot fully characterize the performance of the recognition model, this study also introduces the recall indicator to evaluate the model. The experimental results obtained by different classification models are shown in Table 8 and Figure 6.

By observing the experimental results shown in Table 8 and Figure 6, the following experimental conclusions can be obtained:

- (1) Among various classification models, the experimental results obtained by the ELM model are relatively better, close to 0.8. This is one of the reasons why this study chooses ELM as the basic algorithm. When the label weighting and AdaBoost are introduced to optimize the traditional ELM, the experimental results obtained by IELM have been greatly improved. Compared with ELM, its accuracy is improved by 10.25%. This proves the effectiveness of the improved strategy in this study.

TABLE 8: Experimental results of different classification models.

Index/model	SVM	Linear SVM	RBFNN	RF	ELM	IELM
Accuracy	0.6275	0.6598	0.7609	0.7567	0.7912	0.8728
Recall	0.6515	0.6702	0.7122	0.6547	0.7687	0.8913

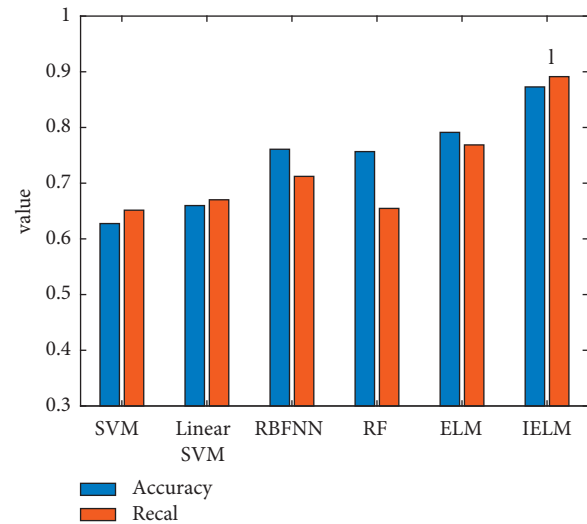


FIGURE 6: Accuracy and recall of different classification models.

- (2) A single indicator is not enough to illustrate the superiority of the model used, so the recall indicator is introduced in the experiment. Compared with SVM, linear SVM, RBFNN, RF, and ELM, the recall rate of the IELM algorithm used is increased by 36.8%, 32.99%, 25.25%, 36.14%, and 15.95%, respectively. The model in this study has been greatly improved in terms of recall rate. This demonstrates the robustness of the algorithm used in this study.

5. Conclusion

Psychological stress has a great impact on human health. The role of early detection and intervention of psychological stress in preventing mental diseases cannot be ignored. However, in terms of stress identification methods, traditional psychological tools such as self-rating scales are highly subjective, and hormone measurement cannot be widely promoted due to its invasiveness and other limitations. The current noninvasive objective evaluation method is still in the research stage. In this study, EEG was selected to assess the psychological stress state of college students. The experiment finally determined the EEG data of 90 students to be used in the experiment. There are 30 students in each grade of freshman, sophomore, and junior. According to the

stress assessment obtained from the questionnaire, 7% of the freshmen were under great stress, 43% were under average stress, and 50% were under little stress. In the sophomore year, 13% were under great stress, 47% were under average stress, and 40% were under less stress. In the third grade, 20% were under great stress, 53% were under average stress, and 27% were under less stress. In general, 14% were under high stress, 47% were under moderate stress, and 39% were under less stress. Among the results obtained by the stress state assessment method based on EEG identification, 12% were under high stress, 46% were under moderate stress, and 42% were under low stress. The stress assessment results obtained using the model used in this study are relatively close to the questionnaire results. This demonstrates that the method described in this study has a certain reference value. However, this study has some limitations. First, whether the self-made questionnaire can collect the most real stress state data needs further research. Second, the number of pressure states used in the experiment is 3, and the division range is relatively general. Future plans are to explore more objective and rigorous data collection methods and optimize the number of stress classifications.

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 open project from Jiangsu Key Laboratory of Media Design and Software Technology (Jiangnan University) under grant 21ST0203, Jiangsu Province Higher Vocational Education High-Level Professional Group Construction Project under grant Su Jiaozhi Letter (2021) no. 1, Jiangsu Province Higher Vocational Education Industry-Education Integration Platform Construction Project under grant Su Jiaozhi Letter (2019) no. 26, Natural Science Research Project of Jiangsu Province Colleges and Universities under grant 18KJD510011, Jiangsu Province High-Level Key Professional Construction Project Funding under grant Su Jiaogao (2017) no. 17, and Jiangsu Province Education Science "14th Five-Year Plan" 2021 Annual Project under grant D/2021/03/03.

References

- [1] T. H. Holmes and R. H. Rahe, "The social readjustment rating scale," *Journal of Psychosomatic Research*, vol. 11, no. 2, pp. 213–218, 1967.
- [2] B. M. Gadzella, "Student-life stress inventory," *College Students*, vol. 74, p. 12, San Antonio, TX, USA, November 1991.
- [3] I. A. Rochasingh, "Perceived stress among graduate students: development and validation of the graduate stress inventory," *Educational and Psychological Measurement*, vol. 54, no. 3, pp. 714–727, 1994.
- [4] P. D. Simon, "The 10-item Perceived Stress Scale as a valid measure of stress perception," *Asia-Pacific Psychiatry*, vol. 13, no. 2, Article ID e12420, 2021.
- [5] I. Ulstein, T. B. Wyller, and K. Engedal, "High score on the Relative Stress Scale, a marker of possible psychiatric disorder in family carers of patients with dementia," *International Journal of Geriatric Psychiatry*, vol. 22, no. 3, pp. 195–202, 2007.
- [6] L. Lemyre and R. Tessier, "Measuring psychological stress. Concept, model, and measurement instrument in primary care research," *Canadian family physician Medecin de famille canadien*, vol. 49, no. 9, pp. 1159–1160, 2003.
- [7] Z. Dharmawan, *Analysis of Computer Games Player Stress Level Using EEG data*, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Netherlands, 2007.
- [8] A. N. N. M. Yosi, K. A. Sidek, H. S. Yaacob, M. Othman, and A. Z. Jusoh, "Emotion recognition using electroencephalogram signal," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 2, pp. 786–793, 2019.
- [9] R. J. Davidson, D. Pizzagalli, J. B. Nitschke, and K. Putnam, "DEPRESSION: perspectives from affective neuroscience," *Annual Review of Psychology*, vol. 53, no. 1, pp. 545–574, 2002.
- [10] S. Richard, N. Y. W. Lewis, and T. H. Wang, "The effect of a naturalistic stressor on frontal EEG asymmetry, stress, and health," *Biological Psychology*, vol. 75, no. 3, pp. 239–247, 2007.
- [11] J. J. B. Allen, H. L. Urry, S. K. Hitt, and J. A. Coan, "The stability of resting frontal electroencephalographic asymmetry in depression," *Psychophysiology*, vol. 41, no. 2, pp. 269–280, 2004.
- [12] S. D. Ja, D. Posthuma, D. I. Boomsma, and E. J. C. de Geus, "The relation between frontal EEG asymmetry and the risk for anxiety and depression," *Biological Psychology*, vol. 74, no. 1, pp. 26–33, 2007.
- [13] J. A. Coan, *The Heritability of Trait Frontal EEG Asymmetry and Negative emotionality: Sex Differences and Genetic Non-additivity*, The University Of Arizona, Tucson, AZ, USA, 2003.
- [14] L. Liu, L. Kuang, and Y. Ji, "Multimodal MRI brain tumor image segmentation using sparse subspace clustering algorithm," *Computational and Mathematical Methods in Medicine*, vol. 2020, Article ID 8620403, 13 pages, 2020.
- [15] L. Liu, Y. Ji, Y. Gao, and L. Kuang, "Classification of epileptic electroencephalograms signal based on integrated radius-basis-function neural-network," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1462–1467, 2018.
- [16] A. Özbeyaz, "EEG-Based classification of branded and unbranded stimuli associating with smartphone products: comparison of several machine learning algorithms," *Neural Computing & Applications*, vol. 33, no. 9, pp. 4579–4593, 2021.
- [17] G. Altan, A. Yayık, and Y. Kutlu, "Deep learning with ConvNet predicts imagery tasks through EEG," *Neural Processing Letters*, vol. 53, no. 4, pp. 2917–2932, 2021.
- [18] J. R. McIntosh, J. Yao, L. Hong, J. Faller, and P. Sajda, "Ballistocardiogram artifact reduction in simultaneous EEG-fMRI using deep learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 1, pp. 78–89, 2021.
- [19] L. C. Sarmiento, S. Villamizar, O. López, A. C. Collazos, J. Sarmiento, and J. B. Rodríguez, "Recognition of EEG signals from imagined vowels using deep learning methods," *Sensors*, vol. 21, no. 19, p. 6503, 2021.
- [20] 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.
- [21] I. F. Kilincer, F. Ertam, and A. Sengur, “A comprehensive intrusion detection framework using boosting algorithms,” *Computers & Electrical Engineering*, vol. 100, Article ID 107869, 2022.
- [22] A. Zahirah, A. B. Selamat, H. Fujita, and O. Krejcar, “The best ensemble learner of bagged tree Algorithm for student performance prediction,” *Knowledge Innovation Through Intelligent Software Methodologies, Tools and Techniques*, vol. 327, pp. 55–64, 2020.
- [23] W. Samek, F. C. Meinecke, and K. R. Muller, “Transferring subspaces between subjects inBrain-computer interfacing,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, 2013.
- [24] M. L. Zhang, Y. K. Li, and X. Y. Liu, “Towards class-imbalance aware multi-label learning,” in *Proceedings of the International Conference on Artificial Intelligence*, pp. 4041–4047, Nanjing, China, August 2015.

Research Article

Assessing the Impact of Continuous Vaccination and Voluntary Isolation on the Dynamics of COVID-19: A Mathematical Optimal Control of SEIR Epidemic Model

Yue Yu, Manman Shi, 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 19 April 2022; Revised 10 May 2022; Accepted 12 May 2022; Published 2 June 2022

Academic Editor: Shengrong Gong

Copyright © 2022 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.

In order to study the impact of continuous vaccination and voluntary isolation for the COVID-19, a susceptible-exposed-infected-recovered-quarantine-vaccines (SEIR-QV) model is proposed. A basic regeneration number R_0 is defined to determine the extinction or persistence of the disease. We numerically analyze the impact of key parameters based on actual parameters of COVID-19, such as the vaccination rate, population importation rate, and natural (or causal) mortality transmission rate on the dynamics of disease transmission. Then we obtain sensitivity indices of some parameters on R_0 by sensitivity analysis. Finally, the stability of the system and the effectiveness of the optimal control strategy are verified by numerical simulation.

1. Introduction

Since the outbreak of the coronavirus disease 2019 (COVID-19), it has been significantly impacting the world economy and our lives. Globally, as of 4:00 pm CET, 28 February 2022, there have been 434,154,739 confirmed cases of COVID-19, including 5,944,342 deaths, reported to the WHO. As of 26 February 2022, a total of 10,585,766,316 vaccine doses have been administered.

With the emergence of multiple mutant strains of novel coronavirus, including Alpha, Beta, Gamma, Delta, and Omicron, many scholars further explore the transmission and epidemic pattern of COVID-19 from different view. These research results are helpful to learn about the transmission and mode of infectious diseases and provide reliable information for the prediction and control of infectious diseases.

In the study of the dynamics of infectious diseases, the susceptibility-infection-recovery (SIR) model was mainly established by Kermack and McKendrick in 1927 with the kinetic method. In order to better study the characteristics of infectious diseases and obtain the best prevention strategies, scholars propose many improved models based on the

traditional SIR model for different situations. By introducing time variation factors, Chen et al. [1] present a new SIR model, which has stronger adaptability and robustness in predicting the number of confirmed cases and inflection points of infectious diseases. Wang et al. [2] propose a SIR model with time-varying isolation protocol. In this model, the dynamic evolution of infectious diseases can be comprehensively analyzed and evaluated. In addition, Hota et al. [3] introduce a closed-loop framework combined with the SIR model and expound the significance of early detection through two feasible optimization questions.

In view of the latent characteristics of some infectious diseases, SEIR models are further proposed based on the SIR model. Li et al. [4] fit the basic regeneration number curve based on the SEIR model. Trends in COVID-19 outbreaks in China, the USA, India, and Iran are predicted and analyzed. The results show that the rapid and efficient isolation measures adopted in China are significant in suppressing COVID-19. In addition, some new models take into account the influence of different factors on infectious diseases, such as the presence of temperature and humidity [5, 6]. It is concluded that COVID-19 is more infectious and lethal at low temperatures and humidity. Besides, many research

studies show that the different epidemic prevention measures, such as mandatory isolation, wearing masks, vaccination, and government control policies, have significant effects on preventing the rapid spread of COVID-19 [7–9].

Due to the problem of asymptomatic infected persons in infectious diseases, Yu et al. [10] propose an SEIR-AQ model by considering factors including prevention and control efforts, isolation strategies, and asymptomatic infected persons. The SEIR-AQ model can well anticipate the spread trend of COVID-19, providing technical support for the scientific assessment of the infectious disease situation. Analysis from the SEIR-AQ model shows that prevention and control isolation, medical follow-up isolation, and other measures have a significant inhibitory impact on the spread of COVID-19. In addition, some scholars use deep learning methods to study the impact of the COVID-19 outbreak on human society [11]. Some other scholars have analyzed and studied the transmission trends of the COVID-19 outbreak. The spatiotemporal evolution and transmission trends of the COVID-19 epidemic are analyzed using an ontological modeling approach by Liu et al. [12].

Although control measures, vaccination, latent infectivity, and other factors are considered in these studies, the stability and existence of the model are not discussed from the mathematical view point. Therefore, we establish a new susceptible-exposed-infected-recovered-quarantine-vaccines (SEIR-QV) model based on the latent period and the influence of prevention and control measures of COVID-19. In mathematical theory, the extinction or persistence of a disease can be determined based on the basic reproduction number and the Lyapunov function. We numerically simulate and analyze the impact of key real-world parameters of COVID-19. What's more, our sensitivity analysis indicates that the related parameters have significant effects on the stability and existence of the SEIR-QV model.

2. The SEIR-QV Model

In this paper, the susceptible-exposed-infectious-recovered-asymptomatic-quarantine-vaccines (SEIR-AQV) model is developed based on the traditional SEIR model. The SEIR-AQV model takes into account the effect of isolation based on the traditional model and divides the population into unisolated susceptible people (S), unisolated exposed people (E), unisolated infected people (I), isolated susceptible people (S_q), isolated exposed people (E_q), asymptomatic infected people (A), hospitalized people (H), and recovered people (R).

The SEIR-AQV model has the following assumptions: first assumption: medical resources are adequate, i.e., isolated exposed people can be directly converted to hospitalized people after diagnosis; second assumption: once infected with COVID-19, one must undergo inpatient treatment to recover, i.e., infected people are unlikely to recover on their own; third assumption: both isolated exposed people and hospitalized people are isolated from the outside world, i.e., they are not infectious; fourth assumption: the infected people will be immune after recovery, i.e., he or she will not become susceptible again after recovery.

The bin transformation relationship is shown in Figure 1. The equations of the SEIR-AQV model are as follows:

$$\left\{ \begin{array}{l} \frac{dS}{dt} = \Lambda - \frac{c\beta}{\delta}S(I + \theta A + vE) + \lambda_2 S_q - (\lambda_1 + \eta)S, \\ \frac{dS_q}{dt} = \lambda_1 S - (\lambda_2 + \eta)S_q, \\ \frac{dE}{dt} = \frac{c(1-q)\beta}{\delta}S(I + \theta A + vE) - (\sigma + \eta)E, \\ \frac{dE_q}{dt} = \frac{cq\beta}{\delta}S(I + \theta A + vE) - (b_3 + \eta)E_q, \\ \frac{dI}{dt} = \sigma eE - (b_1 + \alpha_1 + \eta)I, \\ \frac{dA}{dt} = \sigma(1-e)E - (b_2 + \alpha_2 + \eta)A, \\ \frac{dH}{dt} = b_1 I + b_2 A + b_3 E_q - (r + \alpha_3 + \eta)H, \\ \frac{dR}{dt} = rH - \eta R, \end{array} \right. \quad (1)$$

where c means the rate of exposure; η means the natural mortality rate; q represents the proportion of isolated; δ represents the vaccine coverage rate; β represents the probability of transmission; θ stands for the ratio of A relative to the transmission capacity of I ; v stands for the ratio of E relative to the transmission capacity of I ; b_1, b_2, b_3 stand for the conversion of $I, A,$ and E_q , respectively to H ; λ_1 represent the rate of conversion of S into S_q ; λ_2 stands for the rate of conversion of S_q into S ; σ stands for the ratio of E to I conversion and $\sigma = 1/\text{Incubation period}$; e means the ratio of E to I conversion; $\alpha_1, \alpha_2, \alpha_3$ mean the rate of conversion of $I, A,$ and the rate of cause-specific death of H ; r means the rate of recovery of H ; and Λ means the rate of population importation.

However, given the reality of the COVID-19 epidemic, the status of isolated susceptible people is unlikely to persist over time, and isolated susceptible people have no translational relationship to hamlets other than unisolated susceptible people. Therefore, the population type of isolated susceptible people and the parameters involved in it are removed. Considering also that the proportion of asymptomatic infected people is very small and has minimal effect on the overall transmission trend of the COVID-19 outbreak, the population type of asymptomatic infected people and its involved parameters are also removed. Therefore, we improved the SEIR-AQV model in the context of the COVID-19 outbreak. We construct a new seasonal susceptible-exposed-infected-removed-quarantine-vaccines (SEIR-QV) model with the population transformation relationship bin view shown in Figure 2.

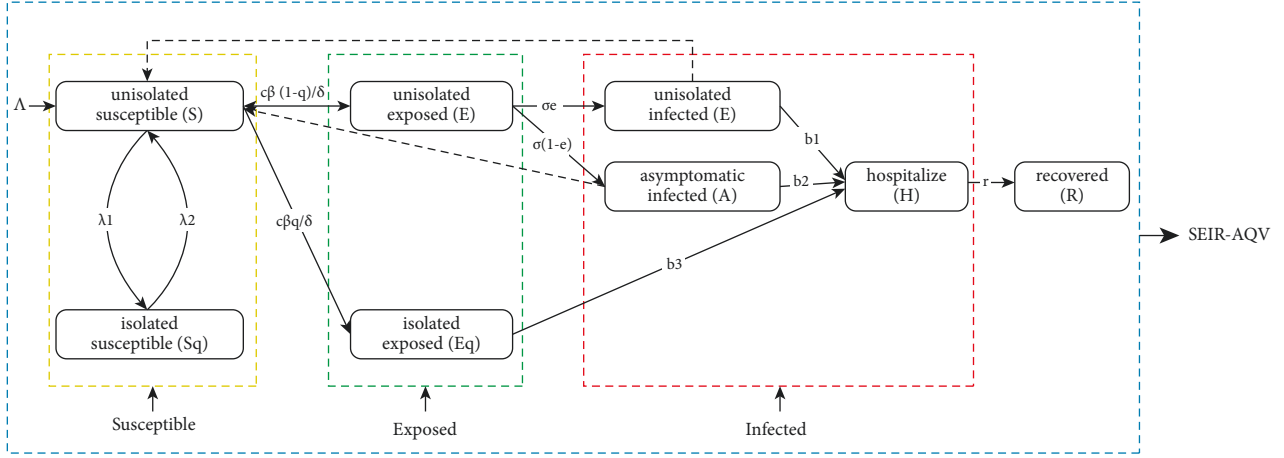


FIGURE 1: Bin view of the population transformation relationship of the SEIR-AQV model.

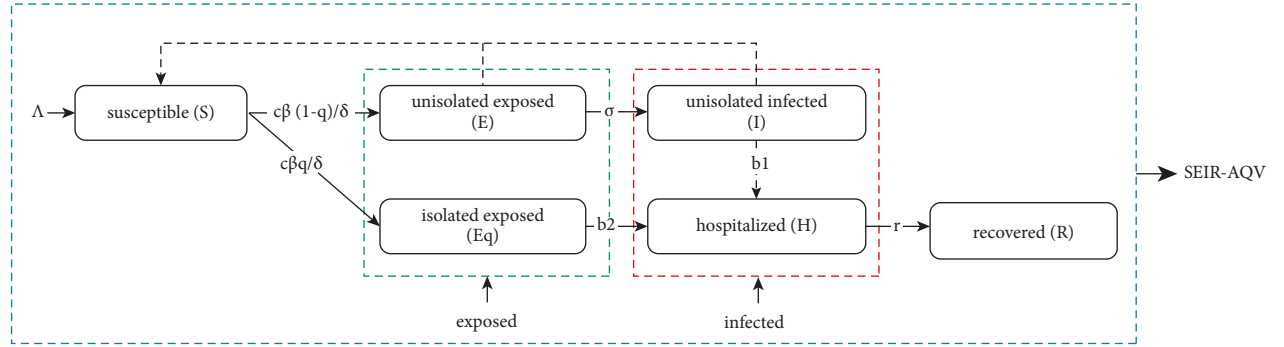


FIGURE 2: Bin view of the population transformation relationship of the SEIR-QV model.

The equations of the SEIR-QV model are as follows:

$$\begin{cases}
 \frac{dS}{dt} = \Lambda - \frac{c\beta}{\delta}S(I + vE) - \eta S, \\
 \frac{dE}{dt} = \frac{c(1-q)\beta}{\delta}S(I + vE) - (\sigma + \eta)E, \\
 \frac{dE_q}{dt} = \frac{cq\beta}{\delta}S(I + vE) - (b_2 + \eta)E_q, \\
 \frac{dI}{dt} = \sigma E - (b_1 + \alpha_1 + \eta)I, \\
 \frac{dH}{dt} = b_1 I + b_2 E_q - (r + \alpha_2 + \eta)H, \\
 \frac{dR}{dt} = rH - \eta R,
 \end{cases} \quad (2)$$

where α_1, α_2 represent the rate of cause-specific death of unisolated infected I and hospitalized H , respectively, and

b_1, b_2 represent the rate of conversion of unisolated infected I and isolated exposed E_q to hospitalized H , respectively.

2.1. Proof of Stability of the SEIR-QV Model. To study the stability of equation (2), it is sufficient to study the stability of the following three formulas in (2) [13, 14]:

$$\begin{cases}
 \frac{dS}{dt} = \Lambda - \frac{c\beta}{\delta}S(I + vE) - \eta S, \\
 \frac{dE}{dt} = \frac{c(1-q)\beta}{\delta}S(I + vE) - (\sigma + \eta)E, \\
 \frac{dI}{dt} = \sigma E - (b_1 + \alpha_1 + \eta)I.
 \end{cases} \quad (3)$$

Considering the biological significance of the system, the dynamical properties of equation (3) are discussed mainly in the closed set $\Omega = \{(S, E, I) \in R_+^3 | S \geq 0, E \geq 0, I \geq 0\}$, where R_+^3 denotes the first trigonometric limit of R^3 and contains the boundary.

2.1.1. *Existence of Equilibrium Point.* The equilibrium point of (3) satisfies

$$\begin{cases} \frac{dS}{dt} = 0, \\ \frac{dE}{dt} = 0, \\ \frac{dI}{dt} = 0. \end{cases} \quad (4)$$

When $I = 0$, (3) has the point of diseased equilibrium $P^0 = (S^0, E^0, I^0) = (\Lambda/\eta, 0, 0)$.

When $I \neq 0$, from $dI/dt = 0$, we get $I = \sigma E/\alpha_1 + b_1 + \eta$; from $dE/dt = 0$, we get $S(I + vE) = \delta(\sigma + \eta)E/c\beta(1 - q)$, from $dS/dt = 0$, we get $S(I + vE) = (\delta/c\beta)(\Lambda - \eta S)$, so $S = (\Lambda/\eta) - ((\sigma + \eta)E/(1 - q)\eta)$.

Bring $I = \sigma E/\alpha_1 + b_1 + \eta$ and $S = (\Lambda/\eta) - ((\sigma + \eta)E/(1 - q)\eta)$ into $S(I + vE) = \delta/c\beta(\Lambda - \eta S)$ to obtain the endemic equilibrium point $P^* = (S^*, E^*, I^*)$, where

$$\begin{aligned} S^* &= \frac{\delta(\sigma + \eta)(\alpha_1 + b_1 + \eta)}{c\beta(1 - q)[\sigma + v(\alpha_1 + b_1 + \eta)]}, \\ E^* &= \frac{\Lambda(1 - q)}{\sigma + \eta} - \frac{\delta\eta(\alpha_1 + b_1 + \eta)}{c\beta[\sigma + v(\alpha_1 + b_1 + \eta)]}, \\ I^* &= \frac{\sigma\Lambda(1 - q)}{(\sigma + \eta)(\alpha_1 + b_1 + \eta)} - \frac{\sigma\delta\eta}{c\beta[\sigma + v(\alpha_1 + b_1 + \eta)]}. \end{aligned} \quad (5)$$

2.1.2. *Basic Regeneration Number.* The basic regeneration number (R_0) indicates the number of people infected by a patient during the average disease period when all are susceptible at the beginning of the disease. $R_0 = 1$ can be used as a threshold to decide whether the disease is extinguished or not. When $R_0 < 1$, the disease will become extinct. When $R_0 > 1$, the disease will persist. The basic regeneration number is closely related to the stability of the endemic equilibrium point.

Next, we study the basic regeneration number R_0 of equation (3). Let $x = (E, I, S)^T$, then equation (3) can be rewritten as

$$\frac{dx}{dt} = F(x) - V(x), \quad (6)$$

where

$$F(x) = \begin{pmatrix} \frac{c(1 - q)\beta}{\delta} S(I + vE) \\ 0 \\ 0 \end{pmatrix}, \quad (7)$$

$$V(x) = \begin{pmatrix} (\sigma + \eta)E \\ (\alpha_1 + b_1 + \eta)I - \sigma E \\ \eta S + \frac{c\beta}{\delta} S(I + vE) - \Lambda \end{pmatrix}.$$

The Jacobi matrices of $F(x)$ and $V(x)$ at the disease-free equilibrium $P^0 = ((\Lambda/\eta), 0, 0)$ are, respectively,

$$\begin{aligned} DF(P^0) &= \begin{pmatrix} F_{2 \times 2} & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{c(1 - q)\beta v \Lambda}{\delta \eta} & \frac{c(1 - q)\beta \Lambda}{\delta \eta} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ DV(P^0) &= \begin{pmatrix} V_{2 \times 2} & 0 \\ \frac{c\beta v \Lambda}{\delta \eta} & \frac{c\beta \Lambda}{\delta \eta} & \eta \end{pmatrix} \end{aligned} \quad (8)$$

$$= \begin{pmatrix} \sigma + \eta & 0 & 0 \\ -\sigma & \alpha_1 + b_1 + \eta & 0 \\ \frac{c\beta v \Lambda}{\delta \eta} & \frac{c\beta \Lambda}{\delta \eta} & \eta \end{pmatrix},$$

where

$$F = \begin{pmatrix} \frac{c(1 - q)\beta v \Lambda}{\delta \eta} & \frac{c(1 - q)\beta \Lambda}{\delta \eta} \\ 0 & 0 \end{pmatrix}, \quad (9)$$

$$V = \begin{pmatrix} \sigma + \eta & 0 \\ -\sigma & \alpha_1 + b_1 + \eta \end{pmatrix}.$$

The basic regeneration number, denoted by R_0 , is thus given by the following equation:

$$R_0 = \rho(FV^{-1}) = \frac{c\Lambda\beta(1-q)}{\eta\delta(\eta+\sigma)} \left[v + \frac{\sigma}{\alpha_1 + b_1 + \eta} \right]. \quad (10)$$

2.1.3. Proof of Stability of the SEIR-QV Model

Theorem 1. For equation (3), the disease-free equilibrium $P^0 = ((\Lambda/\eta), 0, 0)$ is locally asymptotically stable if $R_0 < 1$.

Proof. Linearizing equation (3) at the disease-free equilibrium point $P^0 = ((\Lambda/\eta), 0, 0)$, we obtain the linearization matrix at point $P^0 = ((\Lambda/\eta), 0, 0)$ as the following equation:

$$J(P^0) = \begin{pmatrix} -\eta & \frac{c\beta v\Lambda}{\delta\eta} & \frac{c\beta\Lambda}{\delta\eta} \\ 0 & \frac{c(1-q)\beta v\Lambda}{\delta\eta} - (\sigma + \eta) & \frac{c(1-q)\beta\Lambda}{\delta\eta} \\ 0 & \sigma & -(\alpha_1 + b_1 + \eta) \end{pmatrix}. \quad (11)$$

The characteristic equation of this matrix is $\det(\lambda L - J(P^0))$, where L is a 3×3 unit matrix. Expanding it gives

$$(\lambda + \eta) \left(\lambda^2 + \left((\alpha_1 + b_1 + \eta) + (\sigma + \eta) - \frac{c(1-q)\beta v\Lambda}{\delta\eta} \right) \lambda + (\alpha_1 + b_1 + \eta)(\sigma + \eta)(1 - R_0) \right) = 0. \quad (12)$$

Obviously, this characteristic equation has a negative characteristic root $\lambda = -\eta$. The other characteristic roots satisfy the following equation:

$$\lambda^2 + \left((\alpha_1 + b_1 + \eta) + (\sigma + \eta) - \frac{c(1-q)\beta v\Lambda}{\delta\eta} \right) \lambda + (\alpha_1 + b_1 + \eta)(\sigma + \eta)(1 - R_0) = 0. \quad (13)$$

It is known that $R_0 = \rho(FV^{-1}) = (c\Lambda\beta(1-q)/\eta\delta(\eta+\sigma)) [v + (\sigma/\alpha_1 + b_1 + \eta)] \leq 1$.

Let $R_1 = R_0 - (c\Lambda\beta(1-q)/\eta\delta(\eta+\sigma)) \cdot (\sigma/\alpha_1 + b_1 + \eta) = (c\Lambda\beta(1-q)v/\eta\delta(\eta+\sigma)) < 1$, then

$$\begin{aligned} & (\alpha_1 + b_1 + \eta) + (\sigma + \eta) - \frac{c(1-q)\beta v\Lambda}{\delta\eta} \\ &= (\alpha_1 + b_1 + \eta) + (\sigma + \eta) \left[1 - \frac{c(1-q)\beta v\Lambda}{\delta\eta(\sigma + \eta)} \right] \\ &= (\alpha_1 + b_1 + \eta) + (\sigma + \eta)(1 - R_1). \end{aligned} \quad (14)$$

So we can obtain

$$\lambda^2 + ((\alpha_1 + b_1 + \eta) + (\sigma + \eta)(1 - R_1))\lambda + (\alpha_1 + b_1 + \eta)(\sigma + \eta)(1 - R_0) = 0. \quad (15)$$

It is easy to verify for $R_0 < 1$ and $R_1 < 1$

$$((\alpha_1 + b_1 + \eta) + (\sigma + \eta)(1 - R_1))(\alpha_1 + b_1 + \eta)(\sigma + \eta)(1 - R_0) > 0, \quad (16)$$

that the roots of the quadratic equation are as follows:

$$\lambda^2 + ((\alpha_1 + b_1 + \eta) + (\sigma + \eta)(1 - R_1))\lambda + (\alpha_1 + b_1 + \eta)(\sigma + \eta)(1 - R_0) = 0. \quad (17)$$

All have negative real parts, i.e., the disease-free equilibrium point $P^0 = ((\Lambda/\eta), 0, 0)$ of equation (3) is proved to be locally asymptotically stable. \square

Theorem 2. For equation (3), the disease-free equilibrium $P^0 = ((\Lambda/\eta), 0, 0)$ is globally asymptotically stable if $R_0 \leq 1$.

Proof. By considering the Lyapunov function,

$$V(t) = x_1 E(t) + x_2 I(t). \quad (18)$$

Clearly, the solution of $V(t)$ along equation (3) has $V(t) \geq 0$ and $V(t) = 0$ when and only when $E(t) = 0$ and $I(t) = 0$. The full derivative of the solution of the function $V(t)$ along equation (3) is as follows:

$$\begin{aligned}
\frac{dV(t)}{dt} &= x_1 \frac{dE(t)}{dt} + x_2 \frac{dI(t)}{dt} \\
&= x_1 \left[\frac{c(1-q)\beta}{\delta} S(I + vE) - (\sigma + \eta)E \right] + x_2 [\sigma E - (b_1 + \alpha_1 + \eta)I] \\
&= x_1 \frac{c(1-q)\beta}{\delta} SI + \left[\frac{x_1 c(1-q)\beta v}{\delta} S - x_1(\sigma + \eta) + x_2 \sigma \right] E - x_2(b_1 + \alpha_1 + \eta)I \\
&\leq \frac{c(1-q)\beta\Lambda}{\delta\eta} x_1 I - (b_1 + \alpha_1 + \eta)x_2 I + \left[\frac{c(1-q)\beta v\Lambda}{\delta\eta} x_1 - (\sigma + \eta)x_1 + \sigma x_2 \right] E.
\end{aligned} \tag{19}$$

Then,

Take $x_1 = \sigma$, $x_2 = \sigma + \eta(c(1-q)\beta v\Lambda/\delta\eta)$.

$$\begin{aligned}
\frac{dV(t)}{dt} &\leq \frac{c(1-q)\beta\Lambda\sigma}{\delta\eta} I - (b_1 + \alpha_1 + \eta) \left[(\sigma + \eta) - \frac{c(1-q)\beta v\Lambda}{\delta\eta} \right] I \\
&= \left\{ \frac{c(1-q)\beta\Lambda\sigma}{\delta\eta} + (b_1 + \alpha_1 + \eta) \frac{c(1-q)\beta v\Lambda}{\delta\eta} - (b_1 + \alpha_1 + \eta)(\sigma + \eta) \right\} I \\
&= \left\{ \frac{c(1-q)\beta\Lambda}{\delta\eta} [\sigma + (b_1 + \alpha_1 + \eta)v] - (b_1 + \alpha_1 + \eta)(\sigma + \eta) \right\} I \\
&= \left\{ (b_1 + \alpha_1 + \eta)(\sigma + \eta) \frac{c(1-q)\beta\Lambda}{\delta\eta(\sigma + \eta)} \left[v + \frac{\sigma}{b_1 + \alpha_1 + \eta} \right] - (b_1 + \alpha_1 + \eta)(\sigma + \eta) \right\} I, \\
&= (b_1 + \alpha_1 + \eta)(\sigma + \eta)(R_0 - 1)I,
\end{aligned} \tag{20}$$

where the inequality sign is obtained based on $S \leq S^0$. When $R_0 \leq 1$, there is $dV(t)/dt \leq 0$. Thus, $dV(t)/dt \leq 0$ when and only when $I(t) = 0$. Therefore, if $R_0 \leq 1$, the maximum tight invariant set $\{(S, E, I) \in \Omega | (dV(t)/dt) = 0\}$ is the single point set $\{P^0\}$. Therefore, if $R_0 \leq 1$, the disease-free equilibrium point P^0 of equation (3) is globally asymptotically stable in Ω . \square

Theorem 3. S^*, E^*, I^* are positive when $R_0 > 1$.

Proof

$$S^* = \frac{\delta(\sigma + \eta)(\alpha_1 + b_1 + \eta)}{c\beta(1-q)[\sigma + v(\alpha_1 + b_1 + \eta)]}. \tag{21}$$

It is known that the parameters are positive and $q < 1$, so $S^* > 0$ obviously holds the following equation:

$$\begin{aligned}
E^* &= \frac{\Lambda(1-q)}{(\sigma + \eta)} - \frac{\delta\eta(\alpha_1 + b_1 + \eta)}{c\beta[\sigma + v(\alpha_1 + b_1 + \eta)]} \\
&= \frac{c\beta\Lambda(1-q)[\sigma + v(\alpha_1 + b_1 + \eta)] - \delta\eta(\alpha_1 + b_1 + \eta)(\sigma + \eta)}{c\beta(\sigma + \eta)[\sigma + v(\alpha_1 + b_1 + \eta)]} \\
&= \frac{(c\beta\Lambda(1-q)/(\sigma + \eta))[\sigma + v(\alpha_1 + b_1 + \eta)] - \delta\eta(\alpha_1 + b_1 + \eta)}{c\beta[\sigma + v(\alpha_1 + b_1 + \eta)]} \\
&= \frac{(c\beta\Lambda(1-q)/\delta\eta(\sigma + \eta))[\sigma + v(\alpha_1 + b_1 + \eta)] - (\alpha_1 + b_1 + \eta)}{(c\beta/\delta\eta)[\sigma + v(\alpha_1 + b_1 + \eta)]} \\
&= \frac{(\alpha_1 + b_1 + \eta)\{(c\beta\Lambda(1-q)/\delta\eta(\sigma + \eta))[v + (\sigma/(\alpha_1 + b_1 + \eta))]\} - 1}{(c\beta/\delta\eta)[\sigma + v(\alpha_1 + b_1 + \eta)]} \\
&= \frac{(\alpha_1 + b_1 + \eta)(R_0 - 1)}{(c\beta/\delta\eta)[\sigma + v(\alpha_1 + b_1 + \eta)]}.
\end{aligned} \tag{22}$$

It is known that $R_0 > 1$, $q < 1$ and the parameters are positive, so $E^* > 0$.

$$\begin{aligned}
I^* &= \frac{\sigma\Lambda(1-q)}{(\sigma+\eta)(\alpha_1+b_1+\eta)} - \frac{\sigma\delta\eta}{c\beta[\sigma+v(\alpha_1+b_1+\eta)]} \\
&= \sigma \cdot \frac{c\beta\Lambda(1-q)[\sigma+v(\alpha_1+b_1+\eta)] - \delta\eta(\sigma+\eta)(\alpha_1+b_1+\eta)}{c\beta(\sigma+\eta)(\alpha_1+b_1+\eta)[\sigma+v(\alpha_1+b_1+\eta)]} \\
&= \sigma \cdot \frac{(c\beta\Lambda(1-q)/\delta\eta)[\sigma+v(\alpha_1+b_1+\eta)] - (\sigma+\eta)(\alpha_1+b_1+\eta)}{(c\beta/\delta\eta)(\sigma+\eta)(\alpha_1+b_1+\eta)[\sigma+v(\alpha_1+b_1+\eta)]} \\
&= \sigma \cdot \frac{(c\beta\Lambda(1-q)/\delta\eta(\sigma+\eta))(\alpha_1+b_1+\eta)[v+(\sigma/(\alpha_1+b_1+\eta))] - ((\alpha_1+b_1+\eta))}{(c\beta/\delta\eta)(\alpha_1+b_1+\eta)[\sigma+v(\alpha_1+b_1+\eta)]} \\
&= \sigma \cdot \frac{(c\beta\Lambda(1-q)/\delta\eta(\sigma+\eta))[v+(\sigma/(\alpha_1+b_1+\eta))] - 1}{(c\beta/\delta\eta)[\sigma+v(\alpha_1+b_1+\eta)]} \\
&= \sigma \cdot \frac{R_0 - 1}{(c\beta/\delta\eta)[\sigma+v(\alpha_1+b_1+\eta)]}.
\end{aligned} \tag{23}$$

It is known that $R_0 > 1$, $q < 1$ and the parameters are positive, so $I^* > 0$. \square

Theorem 4. For equation (3), the endemic equilibrium $P^* = (S^*, E^*, I^*)$ is locally asymptotically stable if $R_0 > 1$.

Proof. Linearize equation (3) at the endemic equilibrium point $P^* = (S^*, E^*, I^*)$ and obtain the linearization matrix at point $P^* = (S^*, E^*, I^*)$ as follows:

$$J(P^*) = \begin{pmatrix} \frac{c\beta}{\delta}(I^* + vE^*) - \eta & \frac{c\beta v}{\delta}S^* & \frac{c\beta}{\delta}S^* \\ \frac{c(1-q)\beta}{\delta}(I^* + vE^*) & \frac{c(1-q)\beta v}{\delta}S^* - (\sigma + \eta) & \frac{c(1-q)\beta}{\delta}S^* \\ 0 & \sigma & -(\alpha_1 + b_1 + \eta) \end{pmatrix}. \tag{24}$$

The characteristic equation of this matrix is $\det(\lambda L - J(P^*))$, where L is a 3×3 unit matrix. Expanding it gives $\lambda^3 + A\lambda^2 + B\lambda + C = 0$, where

$$\begin{aligned}
A &= -\frac{c(1-q)\beta v S^*}{\delta} + (\sigma + \eta) + (\alpha_1 + b_1 + \eta) + \eta + \frac{c\beta(I^* + vE^*)}{\delta}, \\
B &= -\frac{c\eta(1-q)\beta v S^*}{\delta} + \eta(\sigma + \eta) + \eta(\alpha_1 + b_1 + \eta) - \frac{c\sigma(1-q)\beta S^*}{\delta} \\
&\quad + (\alpha_1 + b_1 + \eta) \left[-\frac{c(1-q)\beta v S^*}{\delta} + (\sigma + \eta) \right] \\
&\quad + \frac{c\beta((\alpha_1 + b_1 + \eta) + (\sigma + \eta))(I^* + vE^*)}{\delta}, \\
C &= c\eta + \frac{c\beta(\alpha_1 + b_1 + \eta)(\sigma + \eta)(I^* + vE^*)}{\delta}.
\end{aligned} \tag{25}$$

TABLE 1: Values of disease-free equilibrium points q and δ .

q	δ	R_0
0.9	0.1	0.7289
0.9	0.5	0.1458
0.9	1	0.0729
0.1	1	0.6560
0.5	1	0.3645
0.9	1	0.5729

TABLE 2: Values of endemic equilibrium points q and δ .

q	δ	R_0
0.1	0.05	13.12
0.1	0.1	6.5602
0.1	0.5	1.3120
0.1	0.05	13.12
0.5	0.05	7.2891
0.9	0.05	1.4578

Then,

$$\begin{aligned}
AB - C \geq & \frac{c^2 \beta^2 (1-q)^2 v [v\eta + \sigma + v(\alpha_1 + b_1 + \eta)]}{\delta^2} S^{*2} + (f_1 + f_2 + \eta) [f_1 f_2 + \eta(f_1 + f_2)] \\
& - \frac{c^2 \beta^2 (1-q) [2v(\alpha_1 + b_1 + \eta) + 2v\eta + v\sigma + \sigma]}{\delta^2} - \frac{c\beta f_1 f_2}{\delta} - c\eta \\
& - \frac{c\beta(1-q)\Lambda \{v f_1 f_2 + v\eta(f_1 + f_2) + (v\eta + \sigma + v f_1)(f_1 + f_2 + \eta)\}}{\delta\eta},
\end{aligned} \tag{26}$$

where $f_1 = \alpha_1 + b_1 + \eta$ and $f_2 = \sigma + \eta$.

Due to $I^* + vE^* \geq I^0 + vE^0$ and $S^* \leq S^0$, $AB - C > 0$.

Therefore, the endemic equilibrium point $P^* = (S^*, E^*, I^*)$ of equation (3) is locally asymptotically stable on Ω^0 . \square

Theorem 5. For equation (3), the endemic equilibrium $P^* = (S^*, E^*, I^*)$ is globally asymptotically stable if $R_0 > 1$.

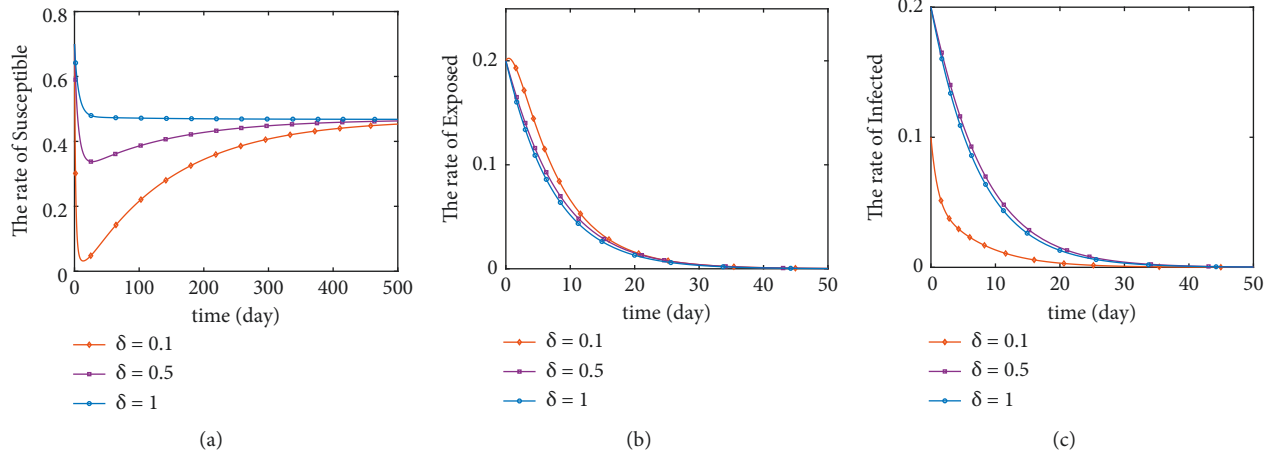
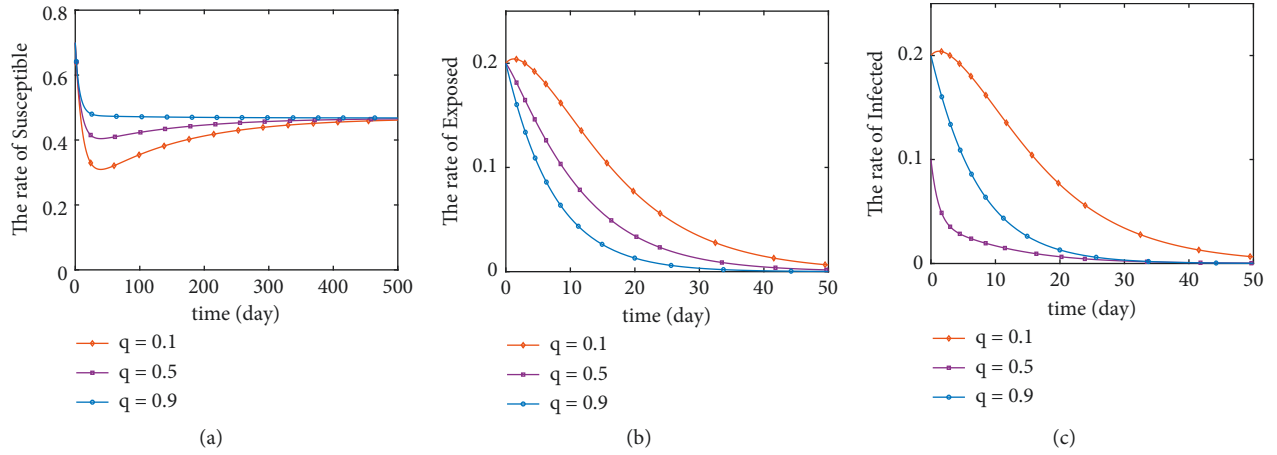
Proof. Considering the Lyapunov function, we can obtain

$$\begin{aligned}
V(t) = & S - S^* - S^* \ln \frac{S}{S^*} + x_1 \left(E - E^* - E^* \ln \frac{E}{E^*} \right) \\
& + x_2 \left(I - I^* - I^* \ln \frac{I}{I^*} \right).
\end{aligned} \tag{27}$$

The full derivative of the solution of the function $V(t)$ along equation (3) is as follows:

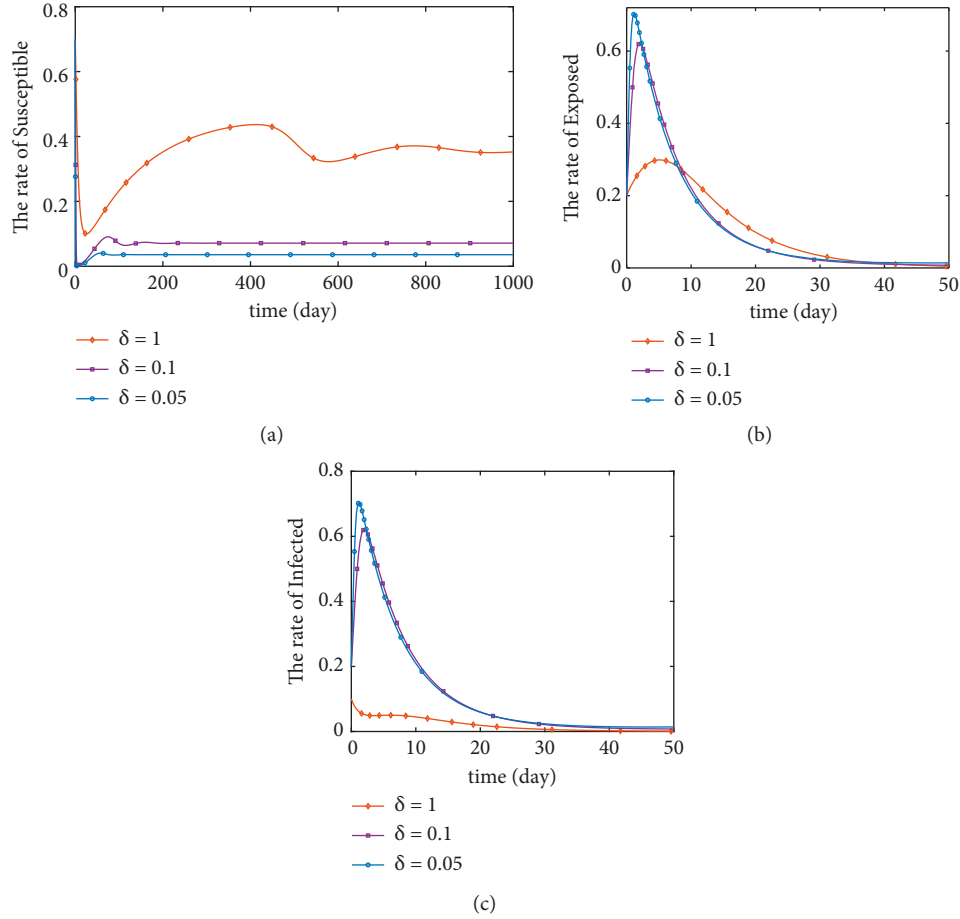
$$\begin{aligned}
\frac{dV(t)}{dt} = & \left(1 - \frac{S^*}{S} \right) \frac{dS}{dt} + x_1 \left(1 - \frac{E^*}{E} \right) \frac{dE}{dt} + x_2 \left(1 - \frac{I^*}{I} \right) \frac{dI}{dt} \\
= & \left(1 - \frac{S^*}{S} \right) \left[\Lambda - \frac{c\beta}{\delta} S(I + vE) - \eta S \right] + x_1 \left(1 - \frac{I^*}{I} \right) [\sigma E - (b_1 + \alpha_1 + \eta)I] \\
& + x_1 \left(1 - \frac{E^*}{E} \right) \left[\frac{c(1-q)\beta}{\delta} S(I + vE) - (\sigma + \eta)E \right] \\
= & (I + vE) \left[\frac{c\beta S^*}{\delta} + x_1 \frac{c(1-q)\beta S}{\delta} - \frac{c\beta S}{\delta} - x_1 \frac{c(1-q)\beta E^* S}{\delta E} \right] + \Lambda - \eta S - \frac{\Lambda S^*}{S} \\
& + \eta S^* - x_1 (\sigma + \eta)E + x_1 (\sigma + \eta)E^* + x_2 \sigma E - x_2 (b_1 + \alpha_1 + \eta)I - x_2 \frac{\sigma E I^*}{I} + x_2 (b_1 + \alpha_1 + \eta)I^*.
\end{aligned} \tag{28}$$

Bringing in $I^* = \sigma E^* / (\alpha_1 + b_1 + \eta)$, $S^* = (\Lambda / \eta) - ((\sigma + \eta)E^* / (1 - q)\eta)$ yields


 FIGURE 3: Stability of disease-free equilibrium point ($q = 0.9$).

 FIGURE 4: Stability of disease-free equilibrium point ($\delta = 1$).

$$\begin{aligned}
 \frac{dV(t)}{dt} &= (I + vE) \left\{ \frac{c\beta}{\delta} \left[\frac{\Lambda}{\eta} - \frac{(\sigma + \eta)E^*}{(1-q)\eta} \right] + x_1 \frac{c(1-q)\beta S}{\delta} - \frac{c\beta S}{\delta} - x_1 \frac{c(1-q)\beta E^* S}{\delta E} \right\} \\
 &\quad + \Lambda - \eta S + \left[\frac{\Lambda}{\eta} - \frac{(\sigma + \eta)E^*}{(1-q)\eta} \right] \left(\eta - \frac{\Lambda}{S} \right) - x_1 (\sigma + \eta)E + x_1 (\sigma + \eta)E^* + x_2 \sigma E \\
 &\quad - x_2 (b_1 + \alpha_1 + \eta)I - \frac{x_2 \sigma^2 EE^*}{I(b_1 + \alpha_1 + \eta)} + x_2 (b_1 + \alpha_1 + \eta) \frac{\sigma E^*}{(b_1 + \alpha_1 + \eta)} \\
 &= \frac{c\beta}{\delta} (I + vE) \left\{ S^* + [x_1(1-q) - 1]S - x_1 \frac{(1-q)E^* S}{E} \right\} - \frac{(\Lambda - \eta S)^2}{\eta S} \\
 &\quad + \frac{(\Lambda - \eta S)(\sigma + \eta)E^*}{\eta S(1-q)} + [x_2 \sigma - x_1(\sigma + \eta)]E + [x_2 \sigma + x_1(\sigma + \eta)]E^* \\
 &\quad - x_2 (b_1 + \alpha_1 + \eta)I - \frac{x_2 \sigma^2 EE^*}{I(b_1 + \alpha_1 + \eta)}.
 \end{aligned} \tag{29}$$

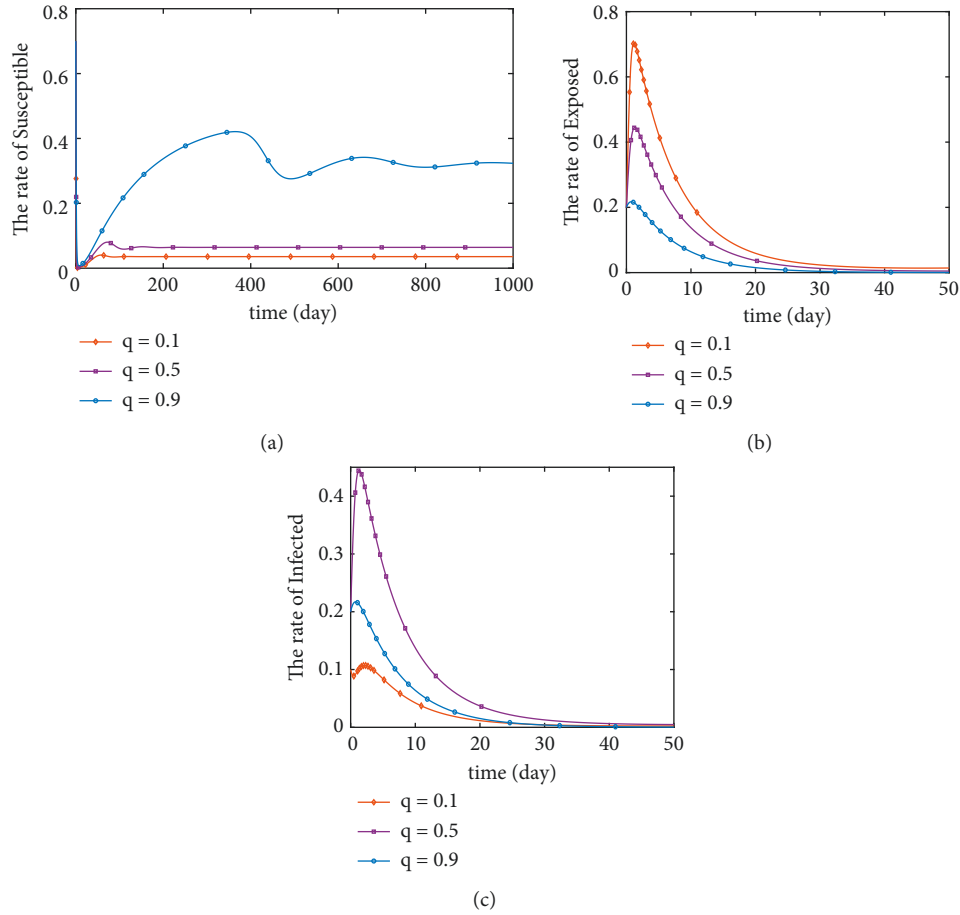
Take $x_1 = 2/(1-q)$ and $x_2 = 2(\sigma + \eta)/\sigma(1-q)$.

FIGURE 5: Stability of disease-free equilibrium point ($q = 0.1$).

Then,

$$\begin{aligned}
 \frac{dV(t)}{dt} &= \frac{c\beta}{\delta} (I + vE) \left[S^* + S - \frac{2E^*S}{E} \right] - \frac{(\Lambda - \eta S)^2}{\eta S} + \frac{(\Lambda - \eta S)(\sigma + \eta)E^*}{\eta S(1 - q)} \\
 &\quad + \frac{4(\sigma + \eta)E^*}{(1 - q)} - \frac{2(\sigma + \eta)(b_1 + \alpha_1 + \eta)I}{\sigma(1 - q)} - \frac{2\sigma(\sigma + \eta)EE^*}{I(b_1 + \alpha_1 + \eta)(1 - q)} \\
 &= \frac{c\beta}{\delta} (I + vE) \left[S^* + \left(1 - \frac{2E^*}{E}\right)S \right] - \frac{(\Lambda - \eta S)}{S} \left[\frac{(\Lambda - \eta S)}{\eta} - \frac{(\sigma + \eta)E^*}{\eta(1 - q)} \right] \\
 &\quad + 2 \frac{2\sigma(b_1 + \alpha_1 + \eta)(\sigma + \eta)E^*I - (\sigma + \eta)(b_1 + \alpha_1 + \eta)I^2 - \sigma^2(\sigma + \eta)EE^*}{\sigma(1 - q)(b_1 + \alpha_1 + \eta)I}.
 \end{aligned} \tag{30}$$

Since $\Lambda - \eta S < \Lambda$ and when $E > (\alpha_1 + b_1 + \eta)$, $\sigma^2 E^2 < (\sigma^2 EE^* / \alpha_1 + b_1 + \eta)$.


 FIGURE 6: Stability of disease-free equilibrium point ($\delta = 0.05$).

At this time,

$$\frac{dV(t)}{dt} \geq \frac{c\beta}{\delta} (I + vE) \left[S^* + \left(1 - \frac{2E^*}{E} \right) S \right] + \frac{(\eta S - \Lambda)}{S} S^* + 2(\sigma + \eta) \frac{(I - \sigma E)^2}{\sigma(1-q)I}, \quad (31)$$

and when $E > 2E^*$ and $S > \Lambda/\eta$, $1 - 2E^*/E > 0$, and $\eta S - \Lambda > 0$ holds, so $dV(t)/dt \geq 0$. Therefore, if $R_0 > 1$, the endemic equilibrium point $P^* = (S^*, E^*, I^*)$ of equation (3) is globally asymptotically stable on Ω^0 . \square

3. Numerical Experiments

3.1. Numerical Simulation. The stability of disease-free and endemic equilibrium points is discussed in this section, and this section uses the software Matlab to numerically simulate equation (3) to verify the above conclusions.

The initial values of the system are $(0.7, 0.2, 0.1)$, and the parameters are assigned as $c = 2$, $\eta = 0.00714$, $\beta = 0.1$, $v = 1$, $b_1 = 0.8$, $\sigma = 1/7$, $\alpha_1 = 0.04$, and $\Lambda = 0.00334$, where q and δ are two variables. In this paper, we will take some random values to simulate the change of S , E , and I ratio when $R_0 < 1$ and $R_0 > 1$, respectively. When $q = 0.9$, δ takes

a random value between 0 and 1 so that $R_0 < 1$ holds. When $\delta = 1$, q takes a random value between 0 and 1 so that $R_0 < 1$ holds. Its value of R_0 is less than 1 ($R_0 < 1$) when q and delta take the values in Table 1. Its value of R_0 is greater than 1 ($R_0 > 1$) when q and delta take the values in Table 2.

The global asymptotic stability of both disease-free and endemic equilibrium points can be seen from Theorems 1–5. Figures 3–6 illustrates the correctness of the results obtained from the above theorem.

If $R_0 < 1$, by numerically simulating the stability of the disease-free equilibrium point, we can analytically conclude the following points. Figure 3 shows the variation curves of S , E , and I as the parameter δ varies when q is constant. The variation of S increases, and the rate of decrease and the rate of increase both increase with the increase of δ . The rate of decrease of E and I decreases with the increase of δ . Finally,

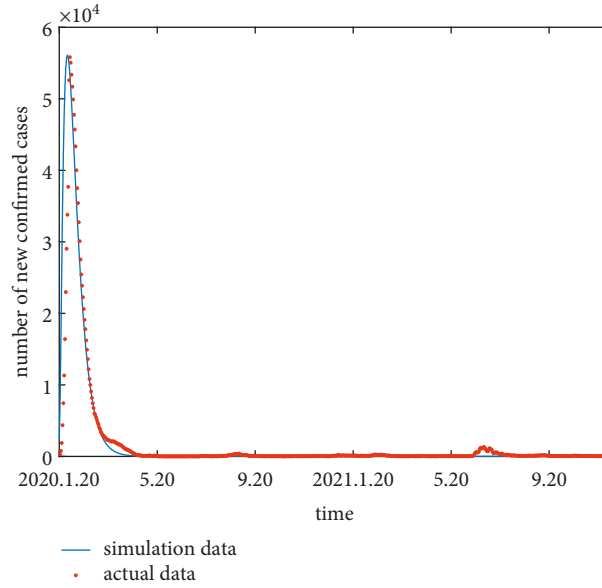


FIGURE 7: Model of the amount of confirmed cases in China per day.

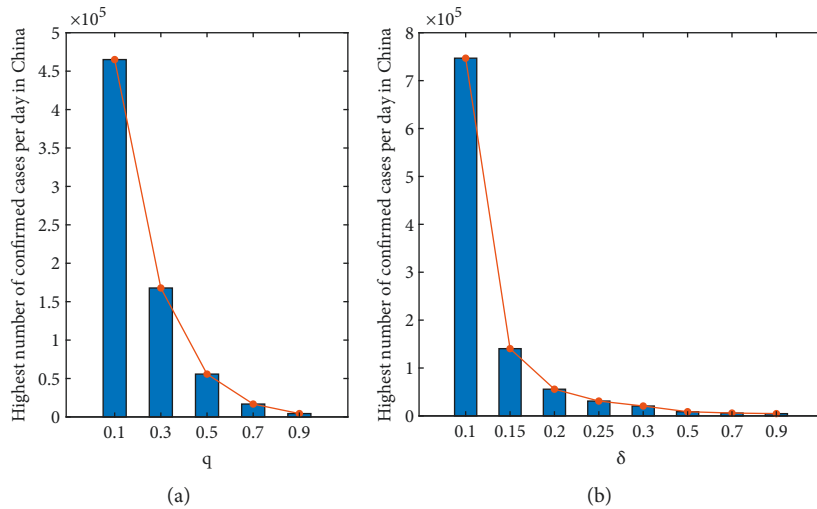


FIGURE 8: Impact of control strategies on the maximum number of daily confirmed cases of the epidemic in China: (a) $\delta = 0.2$ and (b) $q = 0.5$.

S , E , and I all stabilized (see Figures 3(a)–3(c)). The change of S decreases, and the rate of decrease of E decreases with the increase of q . But the change of I has no obvious pattern with the increase of q . Finally, S , E , and I all tend to be stable (see Figures 4(a)–4(c)). Figures 3 and 4 illustrate that when $R_0 < 1$, the changes of q and δ have significant effects on the trends of S , E , and I . And when $R_0 < 1$, the disease-free equilibrium point is tending to be stable.

In Figures 5 and 6, the curve of S , E , and I variation with q and δ is simulated. By comparing the trend of S , E , and I with q in Figure 5 and the trend of S , E , and I with δ in Figure 6, we can draw the following conclusions. Increases in both q and δ lead to larger fluctuations in S and delayed stabilization times. It will lead to an increase in the value of S as it stabilizes. What is more, as δ increases, the rate of

increase and decrease of E decreases. The monotonically increasing interval of I decreases and the rate of decrease of the monotonically decreasing interval decreases with the increase of δ . The maximum peak value of both E and I decreases with the increase of δ . And the time of reaching the maximum peak value of E is delayed and the time of reaching the maximum peak value of I is advanced. Furthermore, as q increases, the rate of increase and decrease of E decreases and E reaches its maximum peak earlier and at a lower value. As q changes, there is no obvious pattern of change in I , but it can be concluded that q has a significant effect on the change in the value of I . Figures 5 and 6 illustrate that when $R_0 > 1$, the changes of q and δ have significant effects on the trends of S , E , and I . And when $R_0 > 1$, the endemic equilibrium point is tending to be stable.

3.2. Model Simulation. The real data of China in this paper are from the official website of the WHO. We preprocess the data, remove some bad points, and make the time continuous so that the data can be better adapted to the SEIR-QV model. The Chinese COVID-19 outbreak is fitted based on the SEIR-QV model. The parameters used in the model are listed as follows: $c = 0.219$, $\eta = 0.00714$, $q = 0.5$, $\delta = 0.2$, $\beta = 1 \times 10^{-8}$, $v = 0.275$, $b_1 = 0.05$, $\sigma = 1/7$, $\alpha_1 = 0.035$, $\Lambda = 0.00334$, $b_2 = 0.05$, $\alpha_2 = 0.035$, and $r = 0.9493$. The simulation results and the actual data are shown in Figure 7.

The Chinese government has decisively taken strict preventive and control measures following the emergence of the COVID-19 outbreak. These prevention and control measures, along with vigorous advocacy by the Chinese government, have reduced exposure rates c and increased isolation rates q and minimized the development of aggregated outbreaks. The COVID-19 outbreak has become manageable and stable. As can be seen in Figure 7, the simulated data fit better with the real data from January 20, 2020 to May 20, 2021. However, there is a small outbreak of the epidemic in China in June 2021. The main reason is that the epidemic is out of control in many places abroad, and people at home are taking it lightly and weakening the efforts of prevention and control, thus giving the opportunity for the epidemic to spread.

As can be seen in Figure 8, both the isolation ratio and vaccine coverage play an important role in the control of the outbreak. The highest number of confirmed cases per day in China decreases as both q and δ increase.

The correlation coefficient is used to measure the accuracy of the fit, and the value of R is found to be 0.9394. Therefore, the results show that the SEIR-QV model can be used for COVID-19 epidemic development status assessment and has important implications.

4. Conclusion

An SEIR-QV model based on vaccination, isolation strategies, and the impact of different parameters on the development of infectious diseases is developed. The equilibrium point and stability of the new model are proved by using the basic regeneration number and the Lyapunov function theory. Simulation experiments show that the new method has certain theoretical value for analyzing and predicting the development of the COVID-19 epidemic. We can get the conclusion that there is a significant impact on the development of infectious diseases by different vaccination and isolation strategies.

Data Availability

The dataset used to support the findings of this study is 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 National Natural Science Foundation of China (no. 62171203).

References

- [1] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A time-dependent SIR model for COVID-19 with undetectable infected persons," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3279–3294, 2020.
- [2] L. Wang, Y. Zhou, J. He et al., "An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China," *Journal of Data Science*, vol. 18, pp. 409–432, 2021.
- [3] A. R. Hota, J. Godbole, and P. E. Pare, "A closed-loop framework for inference, prediction, and control of SIR epidemics on networks," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 3, pp. 2262–2278, 2021.
- [4] W. Li and R. Du, "Analysis of transmission characteristics of novel coronavirus pneumonia and prediction of epidemic development trend," *Journal of Xiamen University*, vol. 59, pp. 1025–1033, 2020.
- [5] Z. Fareed, N. Iqbal, F. Shahzad et al., "Co-variance nexus between COVID-19 mortality, humidity, and air quality index in Wuhan, China: new insights from partial and multiple wavelet coherence," *Air quality, atmosphere, & health*, vol. 13, pp. 1–10, 2020.
- [6] A. Ahlawat, A. Wiedensohler, and S. K. Mishra, "An overview on the role of relative humidity in airborne transmission of SARS-CoV-2 in indoor environments," *Aerosol and Air Quality Research*, vol. 20, no. 9, pp. 1856–1861, 2020.
- [7] Z. Gu, L. Wang, X. Chen et al., "Epidemic risk assessment by a novel communication station based method," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 332–344, 2022.
- [8] C. Hou, J. Chen, Y. Zhou et al., "The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): a well-mixed SEIR model analysis," *Journal of Medical Virology*, vol. 92, no. 7, pp. 841–848, 2020.
- [9] J. Medina, R. Cessa-Rojas, and V. Umpaichitra, "Reducing COVID-19 cases and deaths by applying blockchain in vaccination rollout management," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 2, pp. 249–255, 2021.
- [10] Y. Yu, Y. Zhou, X. Meng et al., "Evaluation and prediction of COVID-19 prevention and control strategy based on the SEIR-AQ infectious disease model," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 1981388, 12 pages, 2021.
- [11] C. Sitaula, A. Basnet, A. Mainali, and T. B. Shahi, "Deep learning-based methods for sentiment analysis on Nepali COVID-19-related tweets," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 2158184, 2021.
- [12] J. Liu and C. Li, "Analysis of the spatial and temporal evolution and transmission trends of the COVID-19 epidemic in ChangZhuTan," *Geospatial Information*, vol. 20, pp. 7–13, 2022.
- [13] H. Zhang and J. Xv, "Stability analysis of SIRS model with vertical infection and vaccination," *Natural Science Journal of Hunan Normal University*, vol. 34, pp. 20–25, 2018.
- [14] H. Jiao and Q. Shen, "Dynamics analysis and vaccination-based sliding mode control of a more generalized SEIR epidemic model," *IEEE Access*, vol. 8, pp. 174507–174515, 2020.

Research Article

Combining BERT Model with Semi-Supervised Incremental Learning for Heterogeneous Knowledge Fusion of High-Speed Railway On-Board System

Lu-jie Zhou ^{1,2}, Zhi-peng Zhao,³ and Jian-wu Dang¹

¹School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

²Key Laboratory of Railway Industry of BIM Engineering and Intelligent for Electric Power, Traction Power Supply, Communication and Signaling, Lanzhou Jiaotong University, Lanzhou 730070, China

³Signal & Communication Research Institute, China Academy of Railway Sciences, Beijing 100081, China

Correspondence should be addressed to Lu-jie Zhou; 792321186@qq.com

Received 30 March 2022; Revised 17 April 2022; Accepted 19 April 2022; Published 31 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Lu-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.

On-board system fault knowledge base (KB) is a collection of fault causes, maintenance methods, and interrelationships among on-board modules and components of high-speed railways, which plays a crucial role in knowledge-driven dynamic operation and maintenance (O&M) decisions for on-board systems. To solve the problem of multi-source heterogeneity of on-board system O&M data, an entity matching (EM) approach using the BERT model and semi-supervised incremental learning is proposed. The heterogeneous knowledge fusion task is formulated as a pairwise binary classification task of entities in the knowledge units. Firstly, the deep semantic features of fault knowledge units are obtained by BERT. We also investigate the effectiveness of knowledge unit features extracted from different hidden layers of the model on heterogeneous knowledge fusion during model fine-tuning. To further improve the utilization of unlabeled test samples, a semi-supervised incremental learning strategy based on pseudo labels is devised. By selecting entity pairs with high confidence to generate pseudo labels, the label sample set is expanded to realize incremental learning and enhance the knowledge fusion ability of the model. Furthermore, the model's robustness is strengthened by embedding-based adversarial training in the fine-tuning stage. Based on the on-board system's O&M data, this paper constructs the fault KB and compares the model with other solutions developed for related matching tasks, which verifies the effectiveness of this model in the heterogeneous knowledge fusion task of the on-board system.

1. Introduction

The train control system incorporates various technologies, such as computers, control, and communication. It is the essential technical equipment for controlling train operations, ensuring operational safety, and improving operational efficiency. As an important component of the train control system, the on-board system plays a major role in operating and controlling the train. However, the on-board system works continuously for a long time, and the fault is inevitable. The fault has the characteristics of concealment and burst. At the same time, the monitoring data and maintenance data supporting the safe and reliable operation of the train control on-board system have the problems of

multi-source heterogeneity and incomplete information. The main reason for these problems is that the sources of O&M data are complex, such as the manual records of drivers, railway experts, and maintenance personnel, as well as the equipment records of on-board vital computers, judicial recorder unit (JRU), and dynamic monitoring system (DMS), which constitute multi-source O&M record information [1]. At present, the maintenance of train control on-board system mainly depends on the technical staff to find out the causes of faults, formulate fault disposal measures, and complete on-board equipment fault information records or analysis reports. This method requires technical staff to repeatedly query and remember a large amount of fault information and maintenance knowledge. The fragmented

and unorganized records make it difficult to share and inherit the fault analysis experience. Therefore, for complex and isolated data, it is necessary to use intelligent technology to extract knowledge from on-board fault maintenance records or reports to form a fault KB to provide support for the transformation of knowledge into a structured and visual knowledge graph. This method is of great significance for comprehensively mastering the key information of faults, realizing dynamic O&M auxiliary decision-making, and improving emergency response-ability. Due to the different data sources, forms, and publishers of on-board system fault KB, the expression forms of fault information are also different, resulting in semantic heterogeneity problems such as homonyms and homographs between knowledge units. The fault KB contains a lot of fuzzy and redundant information, which seriously affects the fusion of fault knowledge of the on-board system. Therefore, it is necessary to match the knowledge units in the KB, establish synonymous entity associations, and eliminate the inconsistency of knowledge expression to ensure the quality of the knowledge in the KB.a.

The fundamental strategy for dealing with the multi-source heterogeneity of knowledge units in on-board fault KB is to match the entities and realize knowledge fusion by judging whether different entities belong to identical objects in reality. The semantic ambiguity makes EM challenging due to the sparse knowledge unit representation and lack of context. Measuring the similarity of knowledge units is the key to the task of EM. The existing research on EM mainly adopts three methods: string similarity [2–4], structure similarity [5–8], and semantic similarity [9–13]. Traditional string-similarity-based methods focus on feature engineering and necessitate extensive theoretical knowledge from experts in order to accomplish EM by mining the similarity rules of knowledge units, which is hard to migrate to another domain. Structure and semantic similarity-based methods typically use knowledge embeddings and word embeddings to map entities in knowledge units into low-dimensional vectors by embedding representations, so that the semantic relevance of entities can be represented by the geometric structure of vector space. After that, various deep learning-based neural network models are used to complete the matching of heterogeneous knowledge units. Deep learning models can extract important features from embeddings automatically, thus avoiding complex feature construction. However, the network structure of neural network-based EM models should be carefully designed to capture the deep semantic or syntactic features of knowledge units to achieve optimal matching performance. In general, neural network performance is influenced by the training corpus, and the cost of corpus construction is very high in railway domain knowledge fusion tasks, where the scale of the domain corpus limits models' performance.

Although the early word embedding is trained on the corpus, it obeys the assumption of context independence, so each word has the same embedding after training. Pretrained models have recently become a research hotspot, such as BERT, OpenAI-GPT, ULM-FiT, etc. [14–17]. Such methods not only substantially improve the text's semantic

representation ability but also facilitate model transfer applications, avoiding the burden of restarting the training after the model has been initialized. With relatively limited training data, BERT achieves competitive results in 11 natural language processing (NLP) tasks and significantly outperforms most embedding-based representations, such as word2vec and Glove. With intensive research on pre-trained language models, these methods have achieved good performance in sentence matching, question answering, classification, etc. However, their potential in knowledge fusion tasks has not been fully explored. It is critical to make sufficient use of the limited supervised data in heterogeneous knowledge fusion tasks of the railway domain to fine-tune the BERT for task awareness and exact matching.

Aiming at the problem that the multi-source heterogeneity of on-board system O&M data affects the construction quality of fault KB, an EM model based on the combination of BERT and semi-supervised incremental learning is proposed, which formulates the fusion of on-board fault knowledge units as a task of pairwise binary classification of entities, to realize the fusion of multi-source heterogeneous knowledge. We investigate the advantages demonstrated by the proposed model for the task of heterogeneous knowledge fusion in high-speed railway on-board systems and design exhaustive experiments to assess our model's performance. This work consists of the following contributions:

- (1) For the multi-source heterogeneity of O&M data in high-speed railway on-board systems, a BERT-based EM model is proposed to extract deep semantic features from data-sparse and context-constrained knowledge units, and the impact of feature selection at different layers in the BERT model on the effect of heterogeneous knowledge fusion is explored.
- (2) To improve the utilization of unlabeled and limited-labeled samples, we propose a pseudo-label-based semi-supervised incremental training strategy that allows the model to collaboratively utilize pseudo-labeled samples to obtain higher knowledge fusion accuracy.
- (3) To strengthen the model's robustness to outliers and noisy data, we utilize an embedding-based adversarial training algorithm in the fine-tuning phase to update the model parameters by adversarial training on noisy data and clean data.

2. Related Work

2.1. Entity Matching. The existing research on EM mainly adopts three methods: string similarity, structure similarity, and semantic similarity.

The string-similarity-based method depends on complex feature engineering, such as entity character matching, attribute matching, or rule mining of knowledge units, and the effect of heterogeneous knowledge fusion is improved by carefully designing features. Such algorithms are effective at matching, but they are unable to handle textual heterogeneity, that is, knowledge units with different forms of

expression [2]. In the task of constructing the KB of railway signal equipment, Li [3] proposed calculating the similarity between words in the railway domain through HowNet and then matching the knowledge units according to the set threshold and the similarity combination of words. In the multi-source heterogeneous knowledge fusion of steam turbines, Yan et al. [4] realized knowledge unit matching by combining character similarity and attribute similarity. This method has high requirements for knowledge unit standardization and requires the participation of a large number of domain experts, which is insufficient in universality.

The structure-similarity-based method relies on knowledge graph (KG) structural information to judge the equivalence of knowledge units. KG can be stored by $\langle \text{head entity, relation, tail entity} \rangle$ triples. These kinds of methods assume that entities in knowledge units representing the same real object in the knowledge graph have similar internal structural information. Through knowledge representation learning, entities or relationships in the KG are encoded to vector spaces, known as knowledge embeddings [5]. In this semantic space, those entities with identical or related meanings tend to be close to each other, and we can use knowledge embeddings to implement knowledge fusion. Existing KG embedding models include translational models [6] and deep models [7]. A recent work [8] points out that the artificially constructed KGs are denser than the real-world KGs. In the entity distribution of real knowledge graphs, most entities are only connected to one or two other entities, known as “long-tail entities.” In the structure-similarity-based knowledge fusion method, long-tailed entities have trouble drawing the attention of the model, so the method is not satisfactory in practice.

The semantic-similarity-based method is to transform the multi-source heterogeneity knowledge fusion task into a pairwise binary classification task of entities in the knowledge units and to realize the fusion of entity pairs with similar semantics through EM. Through the distributed representation technology, that is, word embedding, the entities of knowledge units are mapped from vocabulary to real-number vectors, which represent the semantic features of entities. Logical loss is then used to determine whether the word embeddings of the entity pairs match. For example, Kang et al. [9] first find the possible matching entity pairs by training knowledge embeddings using the KG structural information and use the Word2vec model to obtain word embeddings to select the final matching knowledge units based on the semantic-similarity model. In the literature [10], for the entity category matching task in geographic KB, the entities’ semantic information is learned and represented as semantic vectors by word embedding methods, and the similarity in the entity categories is determined by calculating vectors so that geographic knowledge units can be fused. By training the fastText model, Zeng et al. [11] received independent word embeddings from the information about entity names, which were then combined with the information about entity structures to achieve knowledge fusion via iterative learning. Deep learning (DL) models are a focus of research in the current state of knowledge fusion based on EM. On-board system O&M data as text data can

be converted into chain sequences, and the key to accurately extracting features from these chains is to capture the dependencies between adjacent elements, which often requires complex feature preprocessing or domain knowledge [12]. Mudgal et al. [13] reviewed and validated a variety of DL models for EM, including attention networks, recurrent neural networks, and smooth inverse frequency, as well as their variants, demonstrating the advantages of DL models in heterogeneous knowledge fusion tasks. Constructing entities of semantic matching models based on semantic similarity is an important method to realize multi-source heterogeneous knowledge fusion and has no special requirements for the scale of KB, which is more suitable for the fusion task of domain-specific KB. However, the knowledge units have a deficiency of lack of context, and it is difficult to represent the text features. In existing studies, word embeddings are usually obtained in an unsupervised way to represent the characteristics of knowledge units. This method does not consider the change in vocabulary context and maps entities into fixed vectors. The quality of vectors also directly affects the effect of downstream tasks.

2.2. Pretraining Models. Recent studies have proposed a method to pretrain the model through large unlabeled corpora and fine-tune it to implement the specific task to avoid complex task-specific model structure design and reduce the burden of learning parameters from scratch for the model in NLP tasks [14]. A three-stage model training approach is proposed in the transfer learning method ULM-FiT, consisting of a pretraining model, fine-tuning model, and fine-tuning classifier [15]. Some scholars have proposed the OpenAI-GPT model, which uses a large unlabeled corpus to train the model and learn general language representations, but the model is only based on unidirectional prediction [16]. Using past studies as a foundation, a BERT model based on deep self-attention was developed by Google scholars [17]. It is pretrained on the corpus by masked language model (MLM) and next sentence prediction (NSP), and then the model is fine-tuned using task-specific datasets so that no specific model structure is required, and the model performs well across in language processing tasks. BERT is built using deep bi-directional Transformer networks and pretrains the model with a large corpus, thus providing a deeper structural hierarchy and good parallelism. BERT can be used to better encode contextual representations and fine-tune specific downstream tasks.

Many innovative studies have been conducted using the BERT model. Tenney et al. [18] explored the syntactic and semantic structure within a sentence resolved by each layer of the BERT network. Based on their work, it has been demonstrated that basic syntactic features are typically extracted from the shallow structure of the model, and advanced semantic features are extracted from deep structures, and the use of advanced features helps disambiguate low-level decisions. Considering that each layer of BERT captures the different features of the input text, Sun et al. [19] investigated the effectiveness of features from different layers. By fine-tuning the different layers of BERT in the

classification task, they observed that the feature from BERT's last layer gives the greatest results. At the same time, the last four-layer connection maximizes the collection of BERT information and achieves good performance. In GitHub's open-source bert-as-service [20] project, Dr. Xiao Han proposed that during BERT pretraining, the model's last layer will be closer to the predicted targets and the extracted features will be more skewed toward these targets. Therefore, this service works on the second-to-last layer of BERT. In this work, we combine high-speed railway knowledge and further apply BERT to the heterogeneous knowledge fusion task of on-board system.

3. Task Description of Heterogeneous Knowledge Fusion

To guarantee that the high-speed railway train control on-board system operates efficiently and reliably, the technicians of the railway electricity section and other relevant departments must monitor, analyze, and maintain the on-board equipment in time. Technicians locate the fault according to the operation data of the on-board system and formulate fault handling measures according to maintenance experience and form fault records. Therefore, during the O&M of the on-board system, a large number of unstructured records are accumulated, such as on-board system operation abnormality information analysis reports or tables, which record the operation status and maintenance of the on-board system in detail. Due to the lack of a uniform record format and data standard for on-board system O&M data, the data structure of fault records varies greatly. The problem of multi-source heterogeneity of data not only poses a challenge to the successful sharing of information but also results in an abundance of isolated data. Integrating multi-source heterogeneous on-board system Q&M data and establishing a uniform and standardized fault knowledge system for on-board equipment is the key to achieving intelligent Q&M decisions for the on-board systems. Therefore, we explore the construction of fault KB and multi-source heterogeneous knowledge fusion. Firstly, the key fault knowledge units are extracted from the multi-source heterogeneous data of the unstructured train control on-board system to construct the KB. Knowledge units primarily involve knowledge elements such as entities and relationships. Secondly, the heterogeneous knowledge in the KB is fused to ensure the quality of knowledge and provide support for the organization of knowledge units into a structured and visual KG. KG describes the relationship between faults and maintenance measures, which can help on-site technicians analyze the causes of faults and put forward maintenance suggestions. The construction of KB is a crucial step. The fault KB is mainly composed of entities such as fault modules, fault types, fault causes, fault analysis, equipment phenomena, treatment measures, and maintenance measures, as well as the relationship between entities. The data structure of the on-board system fault KB is shown in Figure 1.

The fault knowledge of the on-board system contains a large number of professional terms related to the railway field. Due to the different data sources, forms, and recorders

of the fault KB, there are semantic heterogeneity problems of homonyms and homographs among entities, and knowledge units contain a large amount of fuzzy or redundant information. For further explanation, we present in Table 1 some examples of entities with the same meaning, i.e., entity 1 and entity 2 represent the identical real object. When combined with the characteristics of the entities in the O&M knowledge unit of the high-speed railway domain, the task of heterogeneous knowledge fusion will have the following three challenges:

Firstly, the challenge of a lack of context. It can be seen from the data types contained in the on-board system fault KB that the length of entities in the on-board knowledge units is short and the context is missing. It is difficult to obtain word embedding of entities in the case of sparse data and a lack of rich context. At the same time, there are too many similar characters between short entities in knowledge units, and the characters with differences are difficult to recognize. For example, when describing equipment maintenance measures, "Replace emergency brake relay" and "Check emergency brake relay" are two different maintenance measures, but they are only different in two Chinese characters. When describing equipment treatment measure, "Switching and restarting the system" can also be written as "Change system and reboot." These two entities represent the same semantics but have more different Chinese characters. Therefore, it is difficult to judge the semantic difference between on-board fault knowledge only by the difference between characters.

Secondly, the challenge of data quality. The on-board system fault knowledge base contains a large number of professional terms. Due to the problems of format, unit, case, space, abbreviation nouns, typing errors, and so on, it will cause a lot of difficulties and interference in knowledge fusion. For example, "CTCS-3 exception downgraded to CTCS-2" can also be written as "C3 \rightarrow C2" when describing equipment phenomena. In addition to the great differences in characters between the two entities, the problems of abbreviation and symbol substitution also affect knowledge fusion. Due to the different writing habits of technicians, some fault records are written in two versions, i.e., Chinese or English. When recording "wireless communication connection timeout" in the fault cause, "wireless" is mistakenly written as "infinite," and this kind of homonym miswriting that leads to semantic ambiguity often occurs.

Thirdly, the challenge of obtaining a priori matching data. A priori matching data are also called training data. When the on-board fault knowledge fusion task is realized by the supervised method, a certain volume of labeled data is required. The number of on-board fault knowledge units is huge. If all entities in the fault KB are matched and labeled in pairs, the cost of construction is relatively high. Moreover, the limited-labeled data also limits the effect of fault knowledge fusion, and the unlabeled samples have not been fully utilized.

To eliminate the semantic conflict of knowledge units in the on-board fault KB and solve the problems of semantic heterogeneity of homonyms and homographs among entities, combined with the requirements of multi-source heterogeneous knowledge fusion in the railway field, the fault

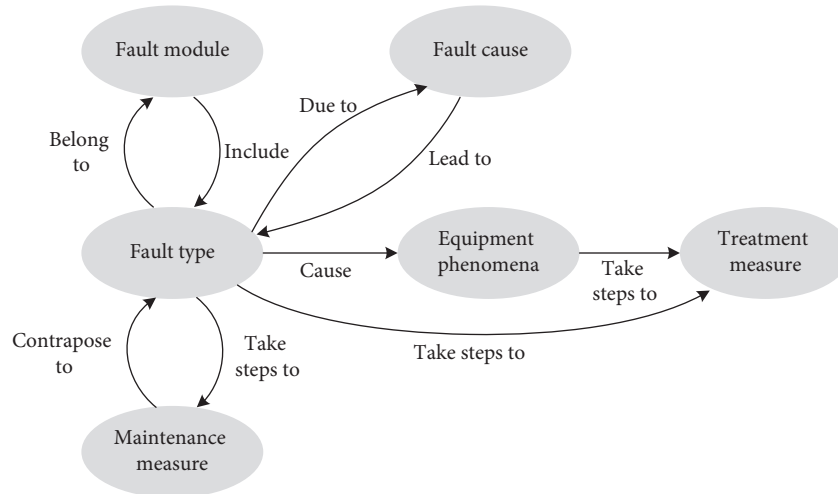


FIGURE 1: Fault knowledge data structure of on-board system.

TABLE 1: Examples of entities in the on-board system fault KB.

Data type	Entity 1	Entity 2
Fault type	A-Kernel mode transition invalid	A mode transition invalid
Equipment phenomena	CTCS-3 exception downgraded to CTCS-2	C3 \rightarrow C2
Fault cause	Wireless communication connection timeout	Infinite communication connection timeout
Treatment measure	Switching and restarting the system	Change system and reboot
Maintenance measure	Replace emergency brake relay	Check emergency brake relay

knowledge units are fused by constructing an EM model to ensure the quality of the on-board fault KB.

4. BERT-Based Model Knowledge Fusion

4.1. Data Preprocessing. In the fault KB of the train control on-board system, $K = (E, R, T)$ is used to represent the knowledge, where E represents the entity, R denotes the relationship between entities, T represents triplet of the relationship facts of the entities in the fault KB. After defining the definition of the KB, we formally define the knowledge fusion task in the KB. Entities in the knowledge units can include any element in the fault KB. In particular, heterogeneous knowledge fusion in fault KB of train control on-board system can be defined as the matching task of entity pairs, and formally defined as:

$$\text{Match}(K_1, K_2) = \{(e_1, e_2) \mid e_1 \in K_1, e_2 \in K_2\}. \quad (1)$$

The EM task is to find all similar entities and generate the matching result $S = \{(e_1, e_2) \mid e_1 = e_2, e_1 \in K_1, e_2 \in K_2\}$, with the equal sign indicating that the two entities represent the identical real object.

The knowledge fusion task of train control on-board system fault KB is formulated into a binary classification problem of entity pairs in our work, and the heterogeneous knowledge fusion is completed by matching the semantic similarity between entities in the fault knowledge units. When constructing the entity pairs data set of on-board fault knowledge, according to the binary classification principle of fault knowledge matching, the matched entity pair is labeled

as 1 and the unmatched entity pair is labeled as 0. We construct the data set according to the sample similarity transfer. To maintain the balance of the data set categories and reduce the influence of background samples, the symmetric expansion method is adopted in constructing the positive samples and the under-sampling method is adopted in constructing the negative samples. Samples that are positive will be labeled with 1, while samples that are negative will be labeled with 0. As shown in Figure 2, the construction method and scale of the on-board entity pair data set is illustrated. Entity1 and Entity2 represent the entities in two fault knowledge units, and Label is the tag of the entity pair.

4.2. Model Structure. We propose an EM model for train control on-board system fault knowledge based on BERT, which is used to realize the task of multi-source heterogeneous knowledge fusion. Figure 3 describes an overall structure of BERT-based EM model for on-board system fault knowledge. The model components include the entity pair input layer for fault knowledge units, a BERT encoder, and the output layer for fault entity pair matching results. In the input layer, the entity pairs of fault knowledge units are used to construct the input sequence for the matching model. Then, the sequence of the input fault entity pair is encoded into a specific hidden state vector containing semantic information by the BERT encoder. Finally, the vector is transferred to the binary classifier to calculate the conditional probability distributions on the predefined fault entity pair categorical labels.

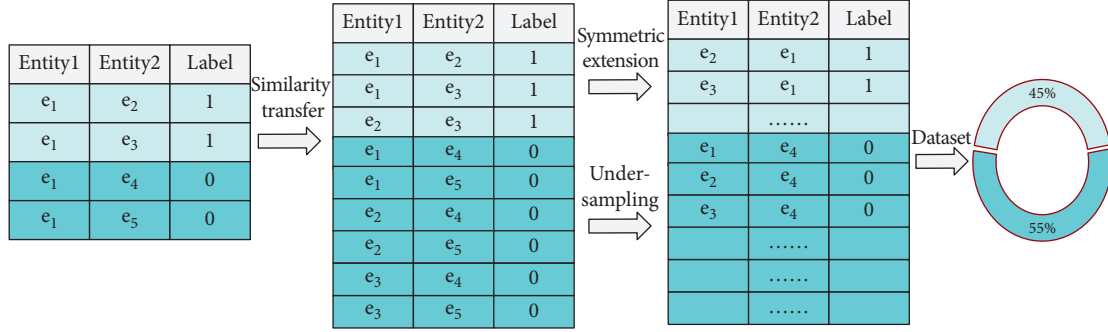


FIGURE 2: Construction method of the entity pairs data set of on-board fault knowledge.

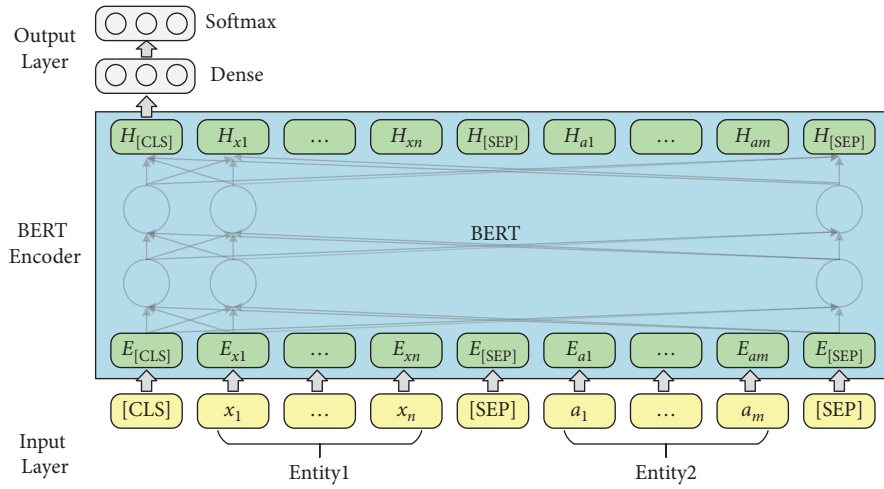


FIGURE 3: Structure of BERT-based model for EM for on-board systems fault knowledge.

- (1) Input layer: The input sequence of tokens are entity pairs constructed using the entities in the on-board fault knowledge units, as shown in Figure 3. For an input entity pair $x_{1:n} = \{x_1, x_2, \dots, x_n\}$ and $a_{1:m} = \{a_1, a_2, \dots, a_m\}$, we add a special token [CLS] to the beginning of the input sequences, and a special token [SEP] to the bottom of the input sequences and to the split between the two entities. The input sequence of on-board fault entity pair tokens is denoted as follows:

$$I = \{[\text{CLS}], x_1, x_2, \dots, x_n, [\text{SEP}], a_1, a_2, \dots, a_m, [\text{SEP}]\}. \quad (2)$$

For the input sequence I , the BERT model constructs token representations E as shown in Figure 4, primarily by the summation of the embeddings of the tokens W , positions W , and segments S . The token $a_{1:m} = (a_1, a_2, \dots, a_m)$ representations of fault entity pair tokens are denoted as:

$$E = \{E_{[\text{CLS}]}, E_{x_1}, E_{x_2}, \dots, E_{x_n}, E_{[\text{SEP}]}, E_{a_1}, E_{a_2}, \dots, E_{a_m}, E_{[\text{SEP}]}\}. \quad (3)$$

- (2) BERT Encoder: BERT (Chinese version) has been pretrained on the Chinese Wikipedia corpora and is

capable of extracting deep semantic features of common Chinese words. However, the on-board fault knowledge text for train control systems is very diverse compared to the common Chinese vocabulary, which contains special professional terms, language conversions, and abbreviations in the railway field. Therefore, we need to continue fine-tuning the BERT using the entity pair data set of on-board fault knowledge. BERT has two-parameter intensive settings, BERTlarge and BERTbase. The BERTlarge requires more memory than the BERTbase [17]. Therefore, we use BERTbase as the basic model for further processing. With a stack of 12 Transformer encoders [21], BERTbase contains 768 hidden layers and 12 self-attention heads, and the basic structure of the Transformer encoder is shown in Figure 5. The Transformer encoder is composed by two parts: a multi-headed self-attention machine and a fully connected layer in which residual connections and layer normalization operations enhance the extraction and retention of features. Due to the limited contextual information of the on-board system fault entities, it is necessary to rely on BERT's self-attention mechanism to capture the global dependencies of the sequences and learn the deep internal features of the on-board fault entity pair sequence.

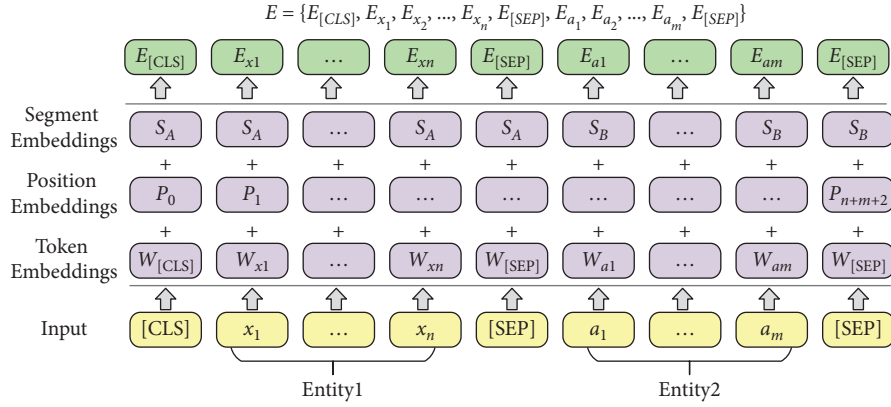


FIGURE 4: Construction of input sequence representations of the fault entity pair for BERT.

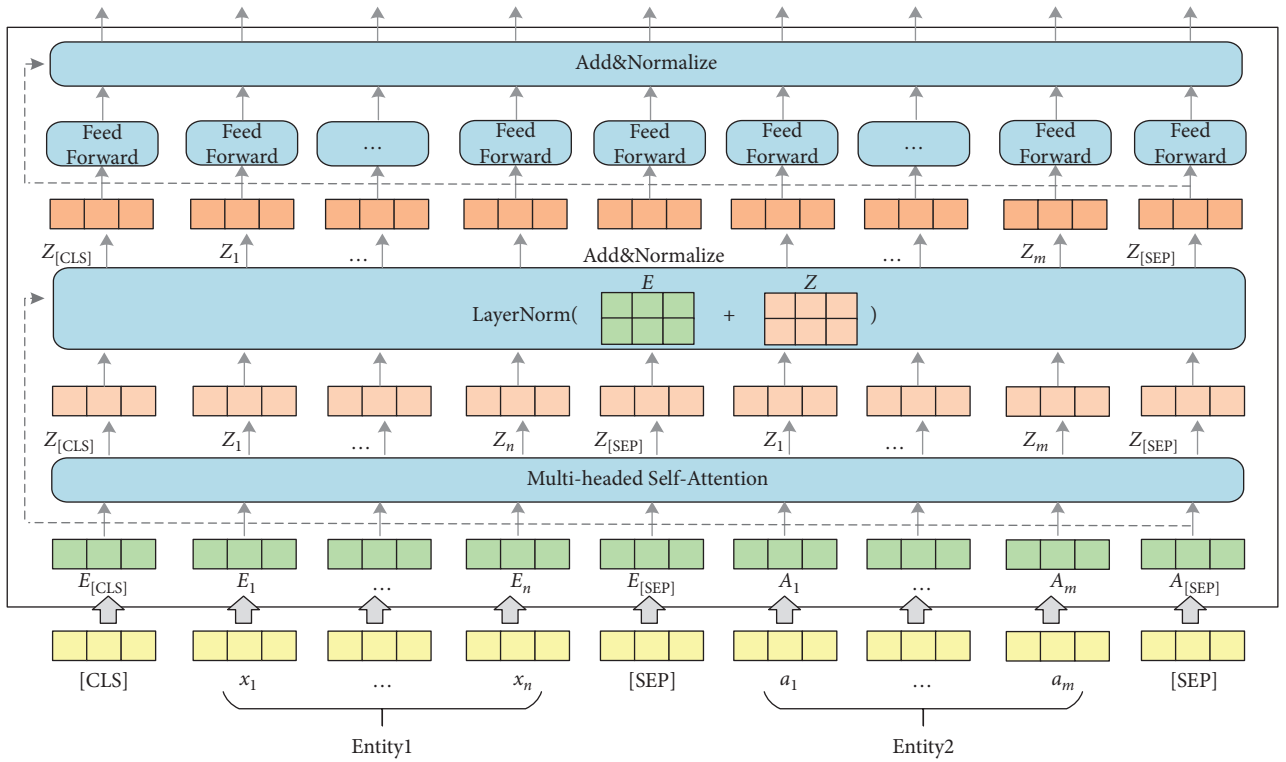


FIGURE 5: Basic structure of Transformer encoder.

The token representations of input on-board fault entity pair are first through the self-attention layer, making the model more focused on the semantic features of fault entities in various subspaces. For a sequence of fault entity pair token representations $E = \{E_1, E_2, \dots, E_N\}$, the attention sub-layer's output $Z = \{Z_1, Z_2, \dots, Z_N\}$ can be calculated as [19]:

$$a_{ij}^{(k)} = \text{softmax} \left(\frac{1}{\sqrt{d_Z}} \cdot (W_Q^{(k)} \cdot E_i)^T \cdot (W_K^{(k)} \cdot E_j) \right),$$

$$Z_i^{(k)} = \sum_{j=1}^N a_{ij}^{(k)} \cdot (W_V^{(k)} \cdot E_j), \quad (4)$$

$$Z_i = W_O [Z_i^{(1)}, Z_i^{(2)}, \dots, Z_i^{(k)}],$$

where $d_Z = d_E/K$, d_E and K denote the hidden states' dimension and self-attention heads' number, respectively. $W_Q^{(k)} \in R^{d_Z \times d_E}$, $W_K^{(k)} \in R^{d_Z \times d_E}$, $W_V^{(k)} \in R^{d_Z \times d_E}$, and $W_O^{(k)} \in R^{d_Z \times d_E}$ are the parameter matrices.

The transformer encoder contains optimization operations such as residual structure and layer normalization to reduce the risk of vanishing gradient and weight matrix degradation caused by the increase in depth of neural networks. Equation (5) is used to calculate the optimized output $Z' = \{Z'_1, Z'_2, \dots, Z'_N\}$.

$$Z' = \text{layerNorm}(E + Z). \quad (5)$$

Input Z' to the fully connected network, $H = \{H_1, H_2, \dots, H_N\}$ can be output after calculation:

$$H_i = W_2 \cdot \text{RELU}(W_1 \cdot Z'_i + b_1) + b_2, \quad (6)$$

where $W_1 \in \mathbb{R}^{d_E \times d_E}$, $b_1 \in \mathbb{R}^{d_E}$, $W_2 \in \mathbb{R}^{d_E \times d_E}$, $b_2 \in \mathbb{R}^{d_E}$ are the parameters. The output $H' = \{H'_1, H'_2, \dots, H'_N\}$ can be calculated following residual connection and layer normalization operation.

$$H' = \text{layerNorm}(Z' + H). \quad (7)$$

Finally, the contextual representation of the on-board system fault entity pair sequence is generated.

- (3) Output layer: Given an input on-board fault entity pair sequence $X = \{x_1, x_2, \dots, x_N\}$, BERT outputs the representations $H' = \{H'_1, H'_2, \dots, H'_N\}$ of each token through the transformer encoder. In existing studies, the common approach to solving sentence matching tasks utilizing BERT models is to add a token [CLS] at the beginning of the input sequence and use the state vector of [CLS] at model's last layer to represent the features of the whole input sequence [16, 22]. The fully connected network is usually joined after the last layer of tokens [CLS], and then Softmax is used to combine all the extracted features for sentence pair matching. The entity pair matching of train control on-board fault knowledge is a pairwise binary classification task. The key information of the on-board fault entity needs to be captured by BERT, so the feature extraction ability of fault knowledge is very important. To investigate the effectiveness of the features selected by BERT from different layers in the on-board fault knowledge fusion task, three output structures present for fault EM are proposed, inspired by the work of Refs. [18–20]. Figure 6 illustrates the three models' output structures. By incorporating the Softmax layer on each output structure's bottom, we can calculate conditional probability distributions on the pre-defined fault entity pair categorical labels.

- (i) The token representations that correspond to [CLS] of BERT's last layer are used to represent the whole sequence and connect it to the Softmax classifier, which is recorded as $\text{BERT}_{\text{LAST}}$.
- (ii) Connect the token representations corresponding to [CLS] of each layer in the last four layers of BERT as the representation of the whole sequence, and then sent to the Softmax classifier, which is recorded as BERT_{CON} .
- (iii) Take the token representations corresponding to [CLS] in the second last layer of BERT used to represent the whole sequence, and connect it to the Softmax classifier to predict the conditional probability distributions of labels, which is recorded as BERT_{SEC} .

4.3. Model Pretraining and Fine-Tuning Strategy. The training of the on-board system fault knowledge fusion model on BERT is divided into two stages: first, it is

pretrained using large unlabeled corpora, and then it is fine-tuned by supervised learning to achieve entity pairs matching of the train control on-board KB. The unsupervised pretraining process is implemented by BERT using the MLM and NSP methods. The former uses the MLM method to predict arbitrarily masked words, while the latter is used to judge if the input sentence is consecutive. Meanwhile, BERT provides a pretrained model for Chinese.

Once it has been pretrained, BERT can be fine-tuned for downstream tasks using supervised learning once it has been pretrained to make it more suitable for the specific domain task of matching for train control on-board system fault knowledge units. Therefore, this paper will investigate the fine-tuning methods of the BERT-based model for EM of train control on-board system fault knowledge, including output structure selection, semi-supervised incremental training strategy based on pseudo-label, and the adversarial training based on embedding.

The first is the output structure selection of the BERT-based model for EM. Since different hierarchical structures in BERT extract different syntactic or semantic features for on-board fault knowledge entity pairs, in order to make BERT adapt to our task, three different output structures are designed to obtain the output features of BERT in the fine-tuning, so as to select the most effective hidden layer features for heterogeneous knowledge fusion. For the EM task oriented to knowledge fusion, it is necessary to judge whether the input two entities point to the same object. In this paper, the matched entity-pair is labeled as 1, and the unmatched entity-pair is labeled as 0. All parameters in BERT are jointly updated by maximizing the conditional probability of correct labels.

In the EM task for fault knowledge fusion of high-speed railway on-board system, the training, validation, and test sets are constructed. Since there is a huge volume of on-board fault knowledge, if all entities in the fault KB are matched and labeled in pairs, the cost of construction is relatively high. The test set includes a great number of unlabeled samples, the hidden information is not fully utilized by the model. In most supervised learning methods, a considerable amount of labeled data is required to train a model. However, semi-supervised learning methods can utilize unlabeled data to train the model. In semi-supervised learning, graph-based and pseudo-label-based methods are typically included according to the theoretical basis. The graph-based methods involve building a nearest-neighbor graph, assuming that the connected nodes are similar and have the same labels, and learning about the distributions of data structures and categories from the unlabeled samples, which is a feature ranking algorithm [23]. Pseudo-label-based methods utilize labeled data to train single or integrated classifiers, which expand the supervised data set with the pseudo label generated by the classifier, and it is an iterative "prediction-selection" process [24]. To decrease the construction burden of labeled data sets, maximize the advantages of unlabeled and limited-labeled samples, and rapidly expand the size of the training set, we propose a pseudo-label-based semi-supervised incremental (SSI) learning strategy, which enables the model to collaboratively utilize pseudo-labeled samples to further optimize the heterogeneous knowledge fusion effect.

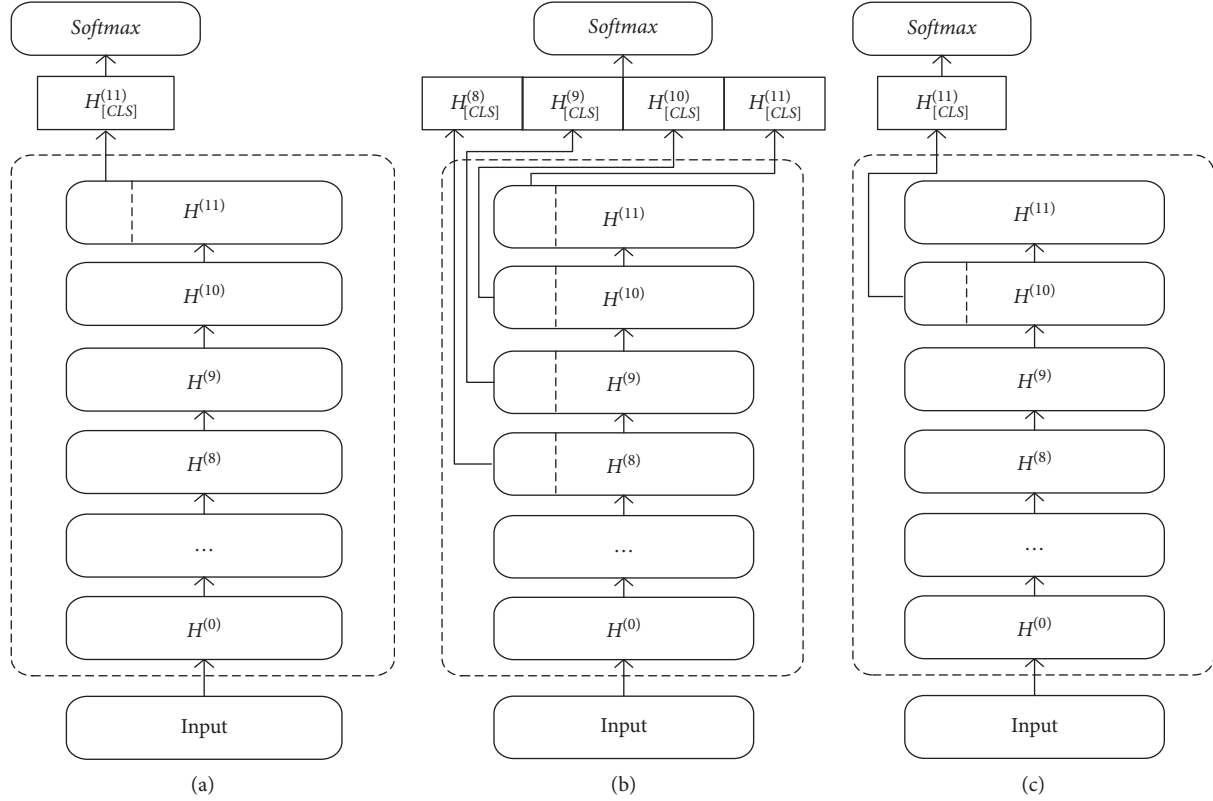


FIGURE 6: The output structure of the model (a) $BERT_{LAST}$. (b) $BERT_{CON}$. (c) $BERT_{SEC}$.

Although unlabeled test data does not have the label information, they are the same as labeled data, which are obtained from the same data source and meet the assumption of independent and identically distributed [25]. Therefore, the information they contain is very beneficial to the optimization model. The SSI learning strategy is shown in Figure 7. In the first stage, the EM model is fine-tuned with the labeled training set until the training result reaches the expected accuracy, and the supervised training is stopped. Then pseudo labels are generated. The unlabeled fault entity pairs set is predicted using the EM model created in the first stage, and the probability value output by the Softmax layer is used as the evaluation index to judge the confidence of the fault EM model to the prediction labels. The prediction label of fault entity pairs with high confidence is taken as the pseudo label, and the pseudo label data are added to the training set for incremental learning. In this step, the labeled data and pseudo labeled data are combined to train the model for the second time. In the process of supervised training and incremental training, the validation set remains unchanged. When finally predicting the test set, the prediction label can be jointly output according to the output results of the first fine-tuning model and the output results of the second training model to complete the final entity pair matching of on-board heterogeneous knowledge units.

Studies have shown that most neural networks are very sensitive to changes in input, and small input disturbances may cause large output differences [26]. In semi-supervised learning, which is prone to the issue of insufficient

generalization ability caused by random noise, some wrong entity pairs will inevitably be introduced when adding pseudo-label samples to the on-board system fault entity pair training set. The input of noise data will cause the model to output incorrect answers with high credibility, which is not conducive for the learning quality of the EM model. To overcome the negative effect of random noise similar to adversarial samples on the model performance and further enhance the generalization ability and robustness of domain EM models for on-board system fault knowledge, this work introduces adversarial training based on embedding as a regularization strategy. The basic idea of adversarial training is to construct adversarial samples to attack the trained network in order to adjust the network parameters to improve robustness so that the network can resist these attacks [27]. We add a small disturbance to the input embedding to generate antagonistic data, and the antagonistic data and the original sample are used as the inputs of the fault knowledge EM model. Adversarial training refers to calculating the disturbance in maximizing the loss and minimizing the loss of the model after increasing the disturbance. When perturbations are introduced into the input embedding of the EM model, it is necessary to introduce disturbance parameters into the loss functions [28], that is:

$$L = -\log p(y | x + r_{adv}; \theta), \quad (8)$$

$$r_{adv} = \arg \min_{r, \|r\| \leq \epsilon} \log p(y | x + r; \hat{\theta}), \quad (9)$$

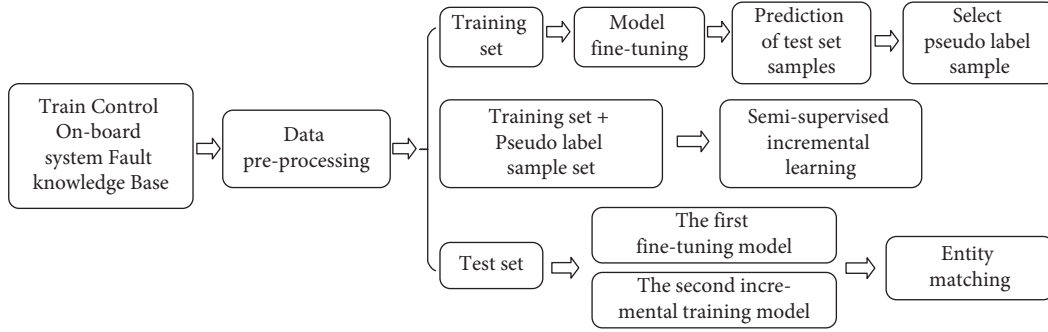


FIGURE 7: SSI learning strategy based on pseudo label.

where x is the input and θ are the parameters of the model, r is a perturbation on the input embedding, θ is a constant set to the current parameters. In each step of training, the most influential perturbations r_{adv} is generated for the current model in (9), and then the model is trained by minimizing (8) with respect to θ to defend against error perturbations and finding the model parameters with the highest robustness. However, we cannot accurately calculate the input perturbations r_{adv} in (9). Therefore, Goodfellow et al. [27] proposed a linear approximation method shown in (10) and (11) with norm constraints to obtain the adversarial perturbation.

$$r_{adv} = \frac{\varepsilon \cdot g}{\|g\|_2}, \quad (10)$$

$$g = \nabla_x \log p(y | x; \hat{\theta}). \quad (11)$$

This approximate method can obtain the input perturbations more easily, and then optimize the model parameters through (8) to strengthen the EM model's robustness and optimize the effect of heterogeneous knowledge fusion.

5. Experiments

5.1. Data Sets and Experimental Settings. Based on the fault knowledge base of the on-board system, the goal of this paper is to realize the fusion of heterogeneous knowledge units, the elimination of semantic conflict, and the unified integration of multi-source information through EM. To examine our model's performance in the heterogeneous knowledge fusion task of the train control on-board system, this paper constructs the fault KB based on the fault maintenance data accumulated from 2017 to 2021 of the CTCS-3 train control on-board system operating on the Xuzhou-Lanzhou High-speed Railway line. On the basis of fault KB, the entity pairs data set of on-board fault knowledge is constructed according to the method proposed by data preprocessing. We chose six types of entities for the knowledge fusion task in this experiment, including fault type, fault cause, fault analysis, equipment phenomenon, treatment measures, and maintenance measures. In the sample set, there are 26000 samples of which the ratio of positive samples (labeled as 1) to negative samples (labeled as 0) is approximately 45% to 55%. Count the character length of entity 1 and entity 2 contained in the entity pair.

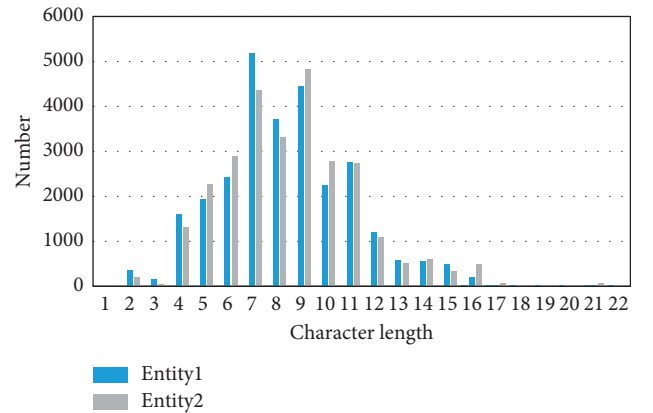


FIGURE 8: Character length distribution of entities.

Figure 8 depicts the entities' character length distribution. The length is mainly between 2 and 22, which is a typical sparse short text. A ratio of 6 : 2 : 2 is used to separate the data set into three parts for training, validation, and testing.

The computer configuration for this experiment is an Intel(R) Core(TM) i7-10750H processor with 2.6 GHz and an NVIDIA GeForce RTX 2060 GPU. All models are implemented in the Python programming language. The BERT-based EM model is encoded using BERTbase, and the dropout rate has been set to 0.2 on the fully connected layer to prevent overfitting of the model. By adjusting the parameters of the model, we adjust the batch size and epochs to 24 and 10, respectively, and select an Adam optimizer with a learning rate of $1e-5$ to train the model and introduce early stops during the training process.

5.2. Contrast Model. To comprehensively evaluate the effectiveness of the EM model in the heterogeneous knowledge fusion task of the on-board system, three metrics including macro-precision (Marco-P), macro-recall (Marco-R), and macro-F (Marco-F) are used for experimental evaluation. And compare our model with other solutions developed for related matching tasks:

LSTM-Siamese [29]: The model incorporates a character-level bi-directional LSTMs with a Siamese network to match semantics between text pairs. We use the Manhattan distance to calculate the semantic similarity between strings to increase the matching accuracy.

TABLE 2: Experimental results under different EM models.

	Entity representation	Models	Marco-P	Marco-R	Marco-F
I	Embedding-based models	ABCNN-3	0.8782	0.8762	0.8765
		LSTM-Siamese	0.8553	0.8219	0.8374
		BIMPM	0.9078	0.8998	0.8985
		ESIM	0.9127	0.9093	0.9086
II	BERT-based models	ALBERT	0.9553	0.9511	0.9529
		ERNIE	0.9729	0.9711	0.9720
III		AD-BERT _{LAST} -SSI	0.9887	0.9883	0.9885
		AD-BERT _{CON} -SSI	0.9859	0.9856	0.9857
		AD-BERT _{SEC} -SSI	0.9871	0.9863	0.9867

ABCNN-3 [30]: In this work, the attention structure is introduced into a convolutional neural network (ABCNN) to construct a sentence pair matching model. ABCNN-3 constructs the attention matrix on both the convolution and pooling parts to establish the connection between sentence pairs.

BIMPM [31]: This paper presents a BiMPPM model in which two sentences are encoded by a BiLSTM encoder, and the encoded sentences are matched in two directions. Then, another BiLSTM can produce matched results.

ESIM [32]: A chain LSTM-based inference model has been designed, and the recursive architecture in both local inference modeling and inference composition is considered.

BERT [17]: This is a basic BERT model. The token representations that correspond to [CLS] of BERT’s last layer are used to represent the whole sequence.

ALBERT [33]: It is a lightweight variant of the BERT structure. To reduce memory consumption and improve training speed, two-parameter reduction technologies are proposed, namely factorized embedding parameterization and self-supervised loss for sentence-order prediction.

ERNIE [34]: Based on the idea of token-entity alignments masking, an enhanced model is trained using a large-scale text corpus and knowledge graph.

5.3. *Model Comparison Results.* Table 2 summarizes each model’s results for the heterogeneous knowledge matching task of the high-speed railway on-board system. Part I are semantic similarity methods, which focus on binary similar entity pair matching by embedding-based models. Part II and III focus on constructing knowledge matching models through various network designs based on BERT, with Part II involving earlier BERT-based variants, and Part III involving different output structures.

Each model’s test results are presented in Table 2. The AD-BERT_{LAST}-SSI model achieves the optimal effect in knowledge fusion of train control on-board system fault KB, and outperforms the strongest baseline model ERNIE by 1.72% for the Marco-R and 1.65% for the Macro-F. According to the experimental results, it can be found that the enhancement of the knowledge fusion effect is primarily for several reasons:

Encoder: In Part I of the experiment, ESIM obtained the best performance in this part, and the Macro-F reached 0.9086. This model makes use of the good sequence modeling ability of LSTM, obtains the local semantic relationship between entity pairs in on-board knowledge units through an attention mechanism, and integrates the local information to construct global reasoning. The methods in Part II and III are better than all embedding-based models in heterogeneous knowledge fusion, and ALBERT is 4.43% higher than the Macro-F of ESIM. The models in Part II and Part III make full use of the advantages of BERT, which constructs a deep network by stacking bi-directional transformer encoders. In comparison with LSTM, this construction method is able to extract local and global dependencies between the on-board entity vocabulary and the context more deeply through self-attention techniques. The multi-level structure of BERT can obtain higher quality semantic and syntactic features, which makes a good foundation for semantic matching of on-board heterogeneous knowledge.

Language Model Pretraining: In the task of heterogeneous knowledge fusion in the on-board system, the BERT-based EM models fully utilize the pretrained model and thus outperform the embedding-based models in fusion. In the knowledge fusion method based on semantic similarity, words need to be converted into low-dimensional vector space using an embedding matrix, and then features are extracted from the on-board knowledge units by deep networks. The parameters of the model are randomly initialized and need to be learned from scratch, so such models do not benefit from pretraining. The model based on pretraining combined with fine-tuning is more flexible. The BERT can be fine-tuned directly to complete the fusion of heterogeneous knowledge.

Fine-tuning Strategies: The experimental results of Part III are better than those of Part II. After the adversarial training and SSI learning strategies are integrated into the BERT model with three output structures, the Macro-F of knowledge fusion is higher than that of ALBERT and ERNIE. By contrast, the combination of SSI learning in BERT can effectively use the semantic feature information of on-board entities contained in unlabeled samples to enhance the model’s generalization ability, as well as by using adversarial training to enhance the model’s robustness, so as to improve the effect of EM. Through the experimental verification, it can also be seen that the feature from BERT’s last layer gives

the best performance, which is 0.18% higher than the Macro-F of AD-BERT_{SEC}-SSI and is more useful for heterogeneous knowledge fusion.

5.4. Ablation Study. We conducted ablation experiments from the perspective of adversarial training and SSI learning strategies to remove some components from the AD-BERT_{LAST}-SSI, AD-BERT_{CON}-SSI, and AD-BERT_{SEC}-SSI proposed in this paper, and judged the contribution of each component in the heterogeneous knowledge fusion task of the on-board system according to the performance (Tables 3 to 5).

The core part of the proposed method is a BERT-based heterogeneous knowledge fusion model, which introduces adversarial training and SSI learning strategies based on the three output structures of the model. Firstly, it can be seen from the experimental results that the BERT-based knowledge fusion model achieves excellent results after removing adversarial training and SSI learning in the presence of abundant data in the training set. Even the basic BERT_{CON} has a Marco-F of 0.9784. Through the ablation experiment, we further observe the general improvement effect of each optimization strategy on the BERT-based model, so as to judge its contribution.

When removing the adversarial training from the three EM models of AD-BERT_{LAST}-SSI, AD-BERT_{CON}-SSI, and AD-BERT_{SEC}-SSI, the Marco-F of on-board system heterogeneous knowledge fusion of the three models showed an overall decrease of some magnitude, and the maximum decreased by 0.54%, which appears on AD-BERT_{CON}-SSI. The general degradation of model performance is due to the existence of unlabeled samples in incremental learning, and the data enhancement method with random noise is not conducive to semi-supervised learning. After using adversarial training to generate adversarial noise that is more relevant to the on-board knowledge, the ability of the model to identify input disturbance is improved, which can facilitate the model to achieve the purpose of identifying unknown samples. Therefore, the experimental results show that after removing the adversarial training, the fusion effect of the models on heterogeneous knowledge all showed a decrease. The adversarial training based on embedding strengthens the EM model’s resolution ability by input disturbances, as well as its generalization ability and robustness in the on-board system’s heterogeneous knowledge fusion.

When removing the SSI, the EM models cannot fully utilize the unlabeled samples, resulting in the decline of knowledge fusion performance of the three models to a certain extent, and the precision and recall are generally decreased. Among them, the performance decreases the most after AD-BERT_{SEC}-SSI removes SSI, with a decrease of 0.23% in Marco-F. In this work, a combination of the BERT and semi-supervised learning is used. The former has a deeper network structure, and can obtain more detailed feature information of entity pairs in on-board system knowledge units through pretraining and fine-tuning. This model has achieved high accuracy in the supervised training

TABLE 3: Ablation study on AD-BERT_{LAST}-SSI Model.

Model	Marco-P	Marco-R	Marco-F
AD-BERT _{LAST} -SSI	0.9887	0.9883	0.9885
<i>Remove components</i>			
adversarial training	0.9874	0.9868	0.9871
SSI learning	0.9879	0.9875	0.9877
adversarial training & SSI learning	0.9860	0.9857	0.9859

TABLE 4: Ablation study on AD-BERT_{CON}-SSI Model.

Model	Marco-P	Marco-R	Marco-F
AD-BERT _{CON} -SSI	0.9859	0.9856	0.9857
<i>Remove components</i>			
adversarial training	0.9812	0.9796	0.9803
SSI learning	0.9851	0.9850	0.9851
adversarial training & SSI learning	0.9794	0.9776	0.9784

TABLE 5: Ablation study on AD-BERT_{SEC}-SSI Model.

Model	Marco-P	Marco-R	Marco-F
AD-BERT _{SEC} -SSI	0.9863	0.9871	0.9867
<i>Remove components</i>			
adversarial training	0.9871	0.9861	0.9866
SSI learning	0.9845	0.9842	0.9844
adversarial training & SSI learning	0.9839	0.9838	0.9838

stage, which also provides a reliable basis for the generation of pseudo labels. The latter increases the incremental learning based on unlabeled test samples, expands the training set’s scale, and makes full use of pseudo-label samples to further enhance the ability of the EM model in knowledge fusion. Certainly, the number of labeled samples also directly affects the fusion effect of the supervised model, and we continue to verify it in the next section of the experiment.

When both adversarial training and SSI learning are removed, the maximum decrease of Marco-F is 0.73%, which appears on AD-BERT_{CON}-SSI. The heterogeneous knowledge fusion of the basic BERT model is already powerful after fine-tuning the model through supervised learning using the knowledge base. Among the proposed optimization strategies, adversarial training and SSI learning are synergistic with each other. When the two optimization strategies are incorporated into the models separately, the Marco-P, Marco-R, and Marco-F of the three models on the on-board system heterogeneous knowledge fusion task are further improved to various extents, and the improvement is stable, so the addition of these components is necessary.

5.5. Analysis of SSI Learning Strategy. Using a limited number of supervised on-board fault entity pairs data set to fine-tune the BERT in downstream tasks is crucial. To investigate the influence of the proposed model and semi-supervised incremental learning strategy on on-board

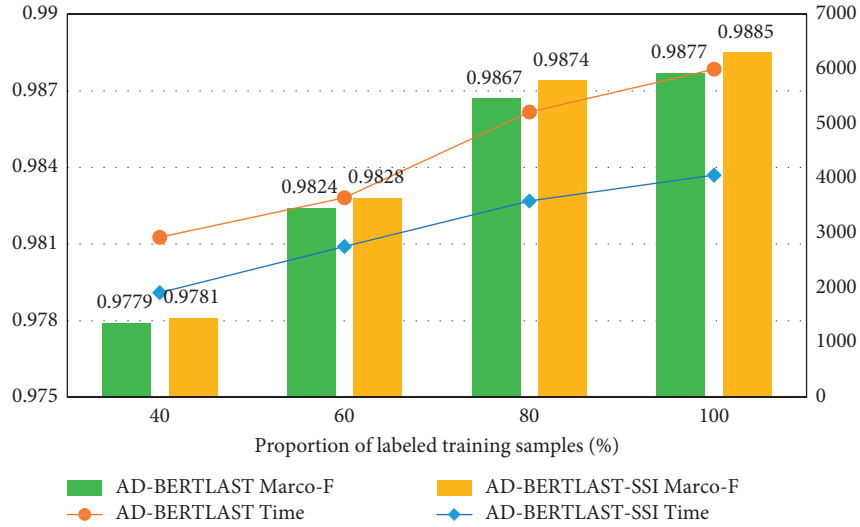


FIGURE 9: Experimental results under different labeled sample numbers.

heterogeneous knowledge fusion under different numbers of labeled samples, we carried out experimental verification under varying numbers of labeled data. Some data from the training data are selected as labeled training data in proportions of 40%, 60%, 80%, and 100% to produce four separate training sample sets, while the validation set and test set remain unchanged. Because AD-BERT_{LAST}-SSI has the best performance among the three EA models proposed, take this model as an example, take AD-BERT_{LAST} as the comparison model to verify the effect of on-board system heterogeneous knowledge fusion. Figure 9 displays the Marco-F values and each Epoch's training time(s) for the two models on different scales of the training set, where the AD-BERT_{LAST} shows the average fine-tuning time of each Epoch. Since AD-BERT_{LAST}-SSI is incremental learning on AD-BERT_{LAST}, the model shows the training time of the first round of incremental learning.

The BERT base model includes a huge number of parameters, about 110M, so the fine-tuning process of the AD-model is also very time-consuming. As illustrated in Figure 9, the model fine-tuning time and incremental learning time grow as the data set gradually expands. The SSI strategy requires several rounds of iterative learning, and as the number of iterative rounds increases, the total training time also increases accordingly round by round.

From the Marco-F value, the fusion effect of the on-board fault knowledge of the two models gradually improves with the growth of the labeled sample size. As the training sample size grows, more prior information is contained in the training set, and the Marco-F of the two EM models will also increase.

AD-BERT_{LAST} is a supervised training model. When using only 40% of the labeled data for training, the Marco-F for the on-board system heterogeneous knowledge fusion task is 0.9779, which is 0.98% lower than when using 100% of the data for training. It shows that the supervised learning method is very dependent on the label information in the training samples, and when the label information is insufficient, the effect of knowledge fusion will be significantly reduced.

AD-BERT_{LAST}-SSI is a semi-supervised incremental learning method. Under the proportion of four kinds of label samples, the result of on-board fault knowledge fusion of AD-BERT_{LAST}-SSI is better than that of AD-BERT_{LAST}. Even when the number of labeled samples is only 40%, the AD-BERT_{LAST}-SSI model can obtain a better matching effect for entity pairs than the supervised training model, and the Marco-F value is increased by 0.02%.

It shows that when the supervised training model can better fuse the heterogeneous knowledge of the on-board system, combined with the semi-supervised incremental learning method, it can accurately expand the labeled samples, and reduce the dependence of the DL model on the label data to a certain extent. When labeled data are limited, the SSI method also has good advantages in the heterogeneous knowledge fusion task of the on-board system.

In practical applications, the scale of labeled data can be controlled to make the DL model have sufficient sample size support. On the other hand, supervised models can be used as pseudo-labeled sample generators to promote data utilization and knowledge fusion by incremental learning of unlabeled samples. This means that SSI learning is capable of reducing the construction cost of the entity pairs data set of on-board fault knowledge, improving the fusion effect of heterogeneous knowledge units, and has strong application value.

6. Conclusion

Processing multi-source heterogeneous O&M data of high-speed railway on-board system and building a high-quality fault knowledge base are important foundations to support knowledge-driven dynamic O&M decision-making of the on-board system. To solve the problem of multi-source heterogeneity of fault knowledge, this work proposes a semi-supervised incremental learning EM model based on BERT to complete the knowledge fusion task. After detailed experimental verification, the results showed that:

- (1) The BERT-based EM model can effectively capture the deep semantic information hidden by knowledge units in the absence of contextual information. The feature from the BERT's last layer gives the best performance, which is more useful for the heterogeneous knowledge fusion task of the on-board system.
- (2) A pseudo-label-based SSI learning strategy is introduced based on the EM model to enable the model to collaboratively utilize pseudo-labeled samples to strengthen the model's generalization capability. This method reduces the dependence of the model on labeled data and has advantages when labeled data are limited.
- (3) Combining the model with adversarial training, adversarial training over noise data and clean data is adopted to update the parameters of the model, guiding the model to learn noise-independent hidden representations, boosting the model's robustness and enhancing the effect of heterogeneous knowledge fusion of the on-board 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 was supported by the Opening Foundation of Key Laboratory of Railway Industry of BIM Engineering and Intelligent for Electric Power, Traction Power Supply, Communication and Signaling (no. BIMKF-2021-01), Fund Project of China Academy of Railway Sciences Corporation Limited (no. 2021YJ099).

References

- [1] T. H. Xu, L. B. Yang, and H. L. Hu, "Heterogeneous data fusion and intelligent maintenance decision for high speed railway signaling systems," *Journal of Xi'an Jiaotong University*, vol. 49, no. 1, pp. 72–78, 2015.
- [2] T. Jiang, C. Bu, and Y. Zhu, "Two-stage entity alignment: combining hybrid knowledge graph embedding with similarity-based relation alignment," in *Proceedings of the Pacific Rim International Conference on Artificial Intelligence (PRICAI) 2019: Trends in Artificial Intelligence*, pp. 162–175, Yanuca Island, Fiji, August 2019.
- [3] X. Q. Li, *Research on Text Big Data Analysis Method of High Speed Railway Safety*, China Academy of Railway Sciences, 2020.
- [4] C. F. Yan, Y. M. Zhang, and K. Y. Ai, "Ontology-based heterogeneous knowledge modeling and fusion of steam turbine," *Journal of Lanzhou University of Technology*, vol. 47, no. 5, pp. 30–37, 2021.
- [5] H. Zhu, R. Xi, and Z. Liu, "Iterative entity alignment via joint knowledge embeddings," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4258–4264, Melbourne, Australia, August 2017.
- [6] A. Bordes, N. Usunier, and A. Garcia-Duran, "Translating embeddings for modeling multi-relational data," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 2787–2795, Lake Tahoe, Nevada, December 2013.
- [7] T. Dettmers, P. Minervini, and P. Stenetorp, "Convolutional 2D knowledge graph embeddings," in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 1811–1818, New Orleans, Louisiana, February 2018.
- [8] L. Guo, Z. Sun, and W. Hu, "Learning to exploit long-term relational dependencies in knowledge graphs," in *Proceedings of the 36th International Conference on Machine Learning*, pp. 1–10, Long Beach, CA, USA, June 2019.
- [9] S. Kang, L. Ji, S. Liu, and Y. Ding, "Cross-Lingual entity alignment model based on the similarities of entity descriptions and knowledge embeddings," *Acta Electronica Sinica*, vol. 47, no. 9, pp. 1841–1847, 2019.
- [10] Z. H. Xu, Y. Q. Zhu, and J. Song, "Word embedding-based method for entity category alignment of geographic knowledge base," *Journal of Geo-information Science*, vol. 23, no. 8, pp. 1372–1381, 2021.
- [11] W. Zeng, X. Zhao, and J. Tang, "Iterative entity alignment via re-ranking," *Journal of Computer Research and Development*, vol. 57, no. 7, pp. 1460–1471, 2020.
- [12] D. Zhang, L. Yao, K. Chen, and S. X. Y. Wang, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3033–3044, 2020.
- [13] S. Mudgal, H. Li, and T. Rekatsinas, "Deep learning for entity matching: a design space exploration," in *Proceedings of the 2018 International Conference on Management of Data*, pp. 10–15, Houston, TX, USA, May 2018.
- [14] S. Yu, J. Su, and D. Luo, "Improving BERT-based text classification with auxiliary sentence and domain knowledge," *IEEE Access*, vol. 7, pp. 176600–176612, 2019.
- [15] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pp. 15–20, Melbourne, Australia, July 2018.
- [16] A. Radford, K. Narasimhan, and T. Salimans, *Improving Language Understanding with Unsupervised Learning*, Technical report, June 2018.
- [17] J. Devlin, M. W. Chang, and K. Lee, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, Minneapolis, Minnesota, June 2019.
- [18] I. Tenney, D. Das, and E. Pavlick, "BERT rediscovers the classical NLP pipeline," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4593–4601, Florence, Italy, July 2019.
- [19] C. Sun, X. Qiu, and Y. Xu, "How to fine-tune BERT for text classification?" in *Proceedings of the China National Conference on Chinese Computational Linguistics*, pp. 194–206, Kunming, China, October 2019.
- [20] H. Xiao, "Bert-as-service," 2018, <http://github.com/hanxiao/bert-as-service>.
- [21] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017)*, pp. 5998–6008, Long Beach, CA, USA, December 2017.

- [22] F. Chen, Z. Yuan, and Y. Huang, "Multi-source data fusion for aspect-level sentiment classification," *Knowledge-Based Systems*, vol. 187, p. 104831, 2020.
- [23] M. Luo, X. Chang, L. Nie, and Y. A. G. Q. Yang, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 648–660, Feb 2018.
- [24] K. Chen, L. Yao, D. Zhang, and X. X. F. Wang, "A semi-supervised recurrent convolutional attention model for human activity recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1747–1756, May 2020.
- [25] Y. Shi, H. Xu, and Y. Liu, "A few-shot modulation recognition method based on pseudo-label semi-supervised learning," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, vol. 38, no. 5, pp. 1074–1083, 2020.
- [26] L. Xiang, J. Zhu, Y. Zhao, Y. Zhou, and C. Zong, "Robust cross-lingual task-oriented dialogue," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 6, pp. 1–24, 2021.
- [27] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proceedings of the International Conference on Learning Representations*, pp. 1–10, San Diego, CA, USA, May 2015.
- [28] T. Miyato, A. M. Dai, and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," in *Proceedings of the International Conference on Learning Representations*, pp. 1–11, Neptune conference house, Toulon, France, April 2017.
- [29] P. Neculoiu, M. Versteegh, and M. Rotaru, "Learning text similarity with siamese recurrent networks," in *Proceedings of the 1st Workshop on Representation Learning for NLP*, pp. 148–157, Boston, MA, USA, August 2016.
- [30] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: attention-based convolutional neural network for modeling sentence pairs," *Transactions of the Association for Computational Linguistics*, vol. 4, no. 1, pp. 259–272, 2016.
- [31] Z. Wang, W. Hamza, and R. Florian, "Bilateral multi-perspective matching for natural language sentences," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 4144–4150, Melbourne, Australia, August 2017.
- [32] Q. Chen, X. Zhu, and Z. Ling, "Enhanced LSTM for natural language inference," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1657–1668, Vancouver, Canada, January 2017.
- [33] Z. Lan, M. Chen, and S. Goodman, "ALBERT: a lite BERT for self-supervised learning of language representations," in *Proceedings of the International Conference on Learning Representations*, pp. 1–17, Addis Ababa, Ethiopia, 2020.
- [34] Z. Zhang, X. Han, and Z. Liu, "ERNIE: enhanced language representation with informative entities," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, May 2019.

Research Article

Dance Performance in New Rural Areas Based on 3D Image Reconstruction Technology

Li Xie 

College of Music and Dance, Yulin Normal University, Yulin 537000, Guangxi, China

Correspondence should be addressed to Li Xie; 20040643@ylnu.edu.cn

Received 14 March 2022; Revised 1 April 2022; Accepted 6 April 2022; Published 27 May 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Li Xie. 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.

Because of the special ecological environment and humanistic atmosphere in new rural areas, excellent regional dance art has been created. Through computer-aided technology, the essence of dance art in rural areas can be reconstructed and displayed. Therefore, based on 3D image reconstruction technology, this paper obtains the dance data of southeast Guangxi and puts forward the dance display scheme of new rural areas. Through acquisition of image information and image matching algorithm, the dance pose is estimated, and the extracted dance sequence is simplified by 3D reconstruction and mapped by texture. In addition, extraction effect of data set, comparison of dance similarity, and user authenticity score were used to test the five types of dance, which provides ideas for the inheritance and development of traditional folk dance culture.

1. Introduction

With the acceleration of the process of socialist modernization, the dance culture is gradually changing. If the development of traditional folk dance stagnates, it will inevitably be eliminated by society and the times, thus being marginalized and disappearing, especially for new rural areas. At the same time, the system of traditional dance culture has attribute of geographical structure, cultural, and social value which directly influence the development direction and mode of dance culture [1, 2].

At present, computer-aided technology already exists to help ethnic dance artists to carry out dance training and inheritance. Researchers express people's behavior, thoughts, and feelings through extracting, organizing, and artistic processing of human actions, but the movements of human limbs move in complex 3D space which is difficult to express it as simply and intuitively as the score of recorded music [3, 4]. Therefore, dance has always been regarded as an art that can only be understood but difficult to express. Computer-aided technology on dance refers to the form of capturing dance movements and abstracting them into models by means of 3D visual sensors and behavioral sensors. Through analysis, processing, and research, we

finally create a dance form that is suitable for the preservation of movement information, so as to make scientific analysis and research on dance, sum up experience rules, and improve the skills of dance art [5, 6]. More importantly, it can record the characteristics of national dance art and contribute to the inheritance of cultural heritage.

As of late, with the turn of events and advocacy of profound learning, artificial brain networks have been effectively applied to the age of dance developments [7, 8]. The conspicuous benefit of embracing profound learning for dance age is that they can straightforwardly remove progressed highlights from crude information (sound, movement catch, and so forth). In addition, deep neural network can create new dance movements but with some problems in the dance generation algorithm based on deep learning. For example, because of the end-to-end model, the front and back frames of the generated dance may not be smooth, which will make the visualization effect of the generated dance worse. Dance data often come from the real world, so it is necessary to extract continuous data of dance pose by using human pose estimation technology [9, 10]. Therefore, applying 3D reconstruction technology to the dance field in new rural areas can promote the spread and protection of Chinese dance culture to a certain extent.

2. Scheme of Dance Display in New Rural Areas

2.1. 3D Image Reconstruction Technology. At present, computer graphics has entered the 3D era, and 3D graphics are everywhere. Virtual reality, scientific visualization, and 3D animation have become the three main research directions of computer graphics in recent years, and their technical core is 3D modeling and 3D reconstruction, which is the reverse process of camera imaging process. In other words, the central issue of 3D remaking innovation from a specific pixel point in the direction framework to a spatial point on the planet coordinate framework is the means to get the profundity data of the objective scene or article. Under the condition that the profundity data of the scene is known, the 3D remaking of the scene can be acknowledged distinctly through the enrollment and combination of point cloud information [11, 12].

2.1.1. Introduction of Process. The acquisition of the depth information of the target object can be divided into passive measurement and active measurement [13, 14]. Active measurement refers to the use of 3D modeling software (such as Maya, 3D MAX, and CAD) and devices (such as scanners) to realize 3D reconstruction. The technology of reconstructing 3D models of real objects by this method is mature, but active operation is complicated. And there seems to be poor reconstruction effect and low efficiency of complex objects. Passive measurement is commonly referred to as image-based method, whose accuracy is relatively low, and the algorithm implementation is relatively complex, but it only needs less equipment and has less restrictions on the reconstructed objects. Image-based 3D reconstruction generally utilizes the impression of the general climate; for example, normal light utilizes the camera to get the picture and afterward works out the three-layered spatial data of the item through a particular calculation. It mainly consists of three phases as shown in Figure 1.

It includes three stages, that is, camera calibration, 3D information extraction, and texture mapping, where the most important part in the reconstruction is the 3D information extraction stage.

2.1.2. 3D Coordinate Extraction. For the computer, the change of angle is particularly complicated in the geometric calculation. The simplest geometric situation is that the camera image is located in the same plane, and those that are not in the same plane can be converted into the same image plane by linear transformation and reprojection, which is called image correction [15]. Stereo vision with multiple cameras under fixed illumination is also called motion recovery structure, and its basic principle is shown in Figure 2.

The purpose of 3D reconstruction is to uniquely determine the 3D coordinates (X, Y, Z) of point P by (x_1, y_1) and (x_2, y_2) . So how to choose the most suitable matching point is tricky. Although the parallax results reflect the 3D positional relationship of the scene, there is still a slight gap between the parallax of some pixels and the standard value [16].

2.1.3. Scheme of Dance Display in New Rural Areas. Considering the complexity of the character model, the multiview stereo vision reconstruction algorithm based on feature points is used to reconstruct the character model. Firstly, the anthropomorphic model is used to experiment, and the rules of reconstructing the character model are explored. Then, the existing reconstruction algorithms are studied and improved, and a relatively perfect reconstruction algorithm is generated. According to experiences, the standard of obtaining data is determined through many comparative experiments [17]. After multiview stereo reconstruction, the anthropomorphic model is generated, the reconstruction effect of the anthropomorphic model is observed, and the model is modified appropriately to complete the establishment of the role model. In addition, texture mapping is carried out after the 3D character model is established to form a realistic model. The specific scheme is shown in Figure 3.

3. Implementation Process

3.1. Estimation Dance Pose

3.1.1. Image Acquisition. In order to reconstruct the character model more easily, many factors should be paid attention in the process of photographing. In terms of light, try not to use flash and choose a room or passage with suitable light. If the light is insufficient, flash can be used to hit the ceiling or the back, so that the light can be evenly distributed. If the flash hits the object or the sunlight shines on one side of the object, the reflection will be formed on the surface of the object, which will easily lead to less matching points of the feature points of the photographed object, and the reconstruction cannot be completed.

During shooting, it is necessary to avoid blurring the photos. The quality of the photos determines the number of point clouds in the reconstructed model and the workload to be modified after reconstruction. Moreover, it is forbidden to focus and change the exposure when taking the same group of photos, that is, choose the same camera to take pictures or take pictures by camera array. Figure 4 shows the shooting of folk dances in southeast Guangxi. Each photo should be completely captured for objects. If there are incomplete characters or objects in the photos, it may lead to less matching of feature points, which may lead to the invalidation.

For the choice of the number of photos, we take the method of multiple photos to shoot each group of objects. The first angle was the flat angle, where one photo was taken every 18° on the flat angle and a total of 20 photos were collected. The second angle was looking down at 30° , where one photo was taken every 30° , and a total of 12 photos were collected. The third angle was looking down at 45° , where one photo is taken every 45° , and a total of 40 photos of 8 kinds are collected, which constitute a set of experimental image data. The angle moved by the photos can be roughly estimated without accurate measurement. In addition, when the photographed object is complex or the photographed image is reconstructed and some objects are missing, the number of photos can be added appropriately.

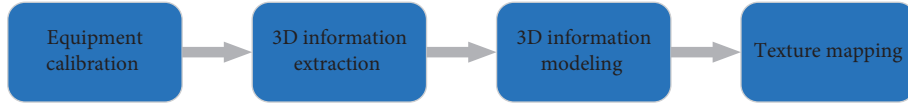


FIGURE 1: Process of image-based 3D reconstruction.

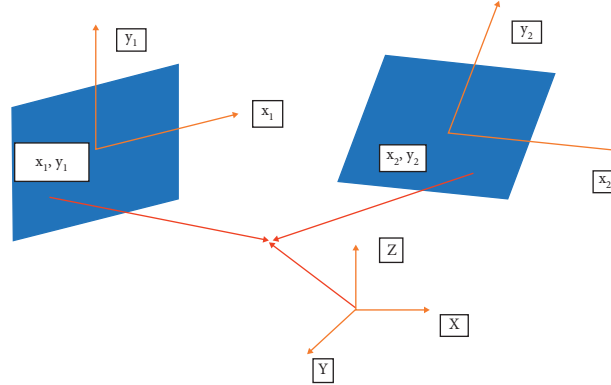


FIGURE 2: Principle of coordinate extraction.

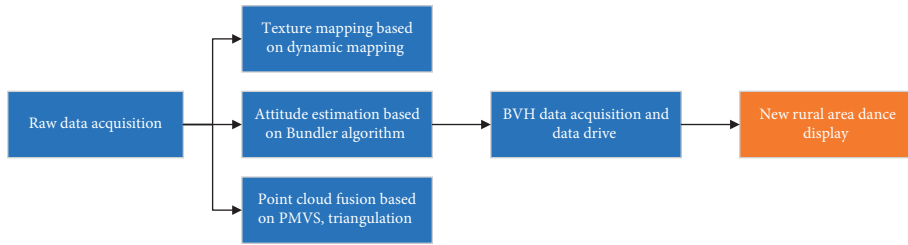


FIGURE 3: Scheme dance display in new rural areas.

3.1.2. Image Matching. In this paper, the SIFT-based image matching method is used, and the RANSAC algorithm is used to eliminate noise. In the field of target detection, the objects in the image are usually different in distance and size. The purpose of scale theory is to build a space where many details can be clearly seen at low scale and outlines can be seen at high scale. The scale space of the image is defined as follows:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y). \quad (1)$$

Among them, $G(x, y, \sigma)$ is a Gaussian function with variable scale.

$$G(x, y, \sigma) = \left(\frac{1}{2\pi\sigma^2} \right) e^{-(x^2+y^2)/2\sigma^2}, \quad (2)$$

where (x, y) represents the spatial coordinates and σ represents a scale coordinate. The blur degree of the image is determined by σ , high value of σ corresponds to the general feature of the image, that is, the low resolution of the image; and low value of σ corresponds to the detail feature of the image, that is, the high resolution of the image.

SIFT takes image pyramid to build scale space, and DOG is a special image pyramid as follows:

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned} \quad (3)$$

In order to find out the extreme point of scale space, it is necessary to compare the sample point with all points of the same scale (i.e., image space) and adjacent scale (i.e., scale space). If the sample point is the maximum or minimum of all points, then the sample point is the extreme point of scale space. By marking points with 8 points of the same scale and 18 points of adjacent scale, a total of 26 points are compared. If the marked point is the maximum or minimum value, it can be guaranteed that it is the extreme point in both image space and scale space, and then the marked point is one of the feature points in this scale. In the process of extreme value comparison, it is impossible for the first and last two layers of each group of images to compare extreme values.

Using DOG pyramid will produce strong edge correspondence, and some points obtained by extreme value



FIGURE 4: Images of folk dance in southeast Guangxi.

comparison may be edge response points. Therefore, in order to select feature points more accurately, it is necessary to use the derivative of spatial scale function and Hessian matrix to achieve subpixel accuracy, which can further eliminate feature points and edge response points that do not meet the contrast requirements.

The spatial scale function is shown in the following formula:

$$D(x, y, \sigma) = D(x, y, \sigma) + \left(\frac{\partial D^T}{\partial x}\right)x + \frac{1}{2x^T} \left(\frac{\partial^2 D}{\partial x^2}\right)x. \quad (4)$$

Deriving it and making it 0, the accurate position \hat{x} can be obtained.

$$\hat{x} = -\left(\frac{\partial^2 D}{\partial x^2}\right)^{-1} \left(\frac{\partial D}{\partial x}\right). \quad (5)$$

In the feature points that have been compared with extreme values in scale space before, the feature points and

edge response points that do not meet the contrast requirements are eliminated.

Introduce \hat{x} into $D(x, y, \sigma)$ and take the first two terms of the space scale function:

$$D(\hat{x}) = D(x, y, \sigma) + \frac{1}{2} \left(\frac{\partial D^T}{\partial x}\right)\hat{x}. \quad (6)$$

If $|D(\hat{x})| < 0.03$, the feature point is discarded.

The edge response points are eliminated by the principal curvature. If the feature points compared by extreme values have a larger principal curvature in the direction parallel to the edge and a smaller principal curvature in the direction perpendicular to the edge, the feature points should be discarded. A 2×2 Hessian matrix of H is defined as

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}.$$

The principal curvature of the spatial scale function is proportional to the eigenvalue of the Hessian matrix, so that α is the maximum characteristic value, β is the smallest eigenvalue, and then

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta, \quad (7)$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$$

Make $\sigma = \gamma\beta^2$, then

$$Det(H) = \frac{(\alpha + \beta^2)}{\alpha\beta} = \frac{(\gamma\beta + \beta)^2}{\gamma\beta^2} = \frac{(\gamma + 1)^2}{\gamma}. \quad (8)$$

If the ratio is greater than $(\gamma + 1)^2/\gamma$, remove it and make $\gamma = 10$ in SIFT. The eliminated points are the detected SIFT feature points, and their positions have been determined. The direction parameters of each SIFT feature point are determined by the gradient direction.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}. \quad (9)$$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}. \quad (10)$$

Use (9) and (10) to find the modulus and direction of gradient in feature points (x, y) , respectively.

Through the above calculation, all SIFT feature points in the image have been detected, and the feature area of each feature point can be determined by the location, scale, and direction of each feature point.

3.1.3. Attitude Estimation. In computer vision, motion and structure reconstruction refer to recovering corresponding 3D information from 2D images or videos, including camera motion parameters and scene structure information. In this paper, we use Bundler based on the Levenberg–Marquardt algorithm. After detection and match of features, Bundle

Adjustment can be based on the projection of all points in the image as the standard and the relative motion parameters of 3D points and the parameters of the camera are obtained at the same time [18]. It can match the observed image position with the predicted image position by minimum error that is expressed by the sum of squares of nonlinear functions:

$$x^* = \arg \min_x \sum_{i=1}^k f_i(x)^2. \quad (11)$$

Minimizing error is realized by the nonlinear least square method, and formula (12) can be used to express the working process of Bundle Adjustment:

$$\min_{a_j, b_i} \sum_{i=1}^n \sum_{j=1}^m v_{ij} d(Q(a_j, b_i), x_{ij})^2, \quad (12)$$

assuming that there are n 3D points in m shooting scenes. V_{ij} represents the mapping relation of point i on image j . If there exists mapping relation, V_{ij} is taken as 1; otherwise, it is 0. a_j is the vector parameterization of image j , and b_i is the parameterization of 3D points. Q is the predicted image bit of point i on image j , and x_{ij} is the observed image bit. The process of minimization is to find the minimum error match between predicted image bits and observed image bits.

3.2. Simplification of 3D Model. Mesh simplification is to reduce the precision of the model surface, reduce the number of triangular surfaces in the model as much as possible, and keep the shape of the original model as much as possible in the simplified model, which finally generate an approximate model of the original model. The approximate model basically keeps the features of the original model, including geometric features and visual features. However, the number of vertices and triangles is less than that of the original mesh. There are usually two kinds of constraints when generating the approximate model, namely, the triangle number constraint and the simplified error constraint [19, 20]. Through mesh simplification, a number of simplified models with different resolutions can be generated, and the real-time rendering can be accelerated. The technology of model simplification can make the storage, transmission, calculation, and display rendering of the model more effective, which can reduce the consumption of hard disk and memory and make the computer more effective, so as to achieve faster speed of rendering and execution.

In this paper, we use the subdivision level function provided by Zbrush to change the number of triangular faces, and the subdivision level function adopts the method of triangle folding. When the model is rendered, advanced subdivision is adopted, so that the rendering effect is better and more realistic. When the model is output, it is simplified and low-level subdivision is adopted, which can speed up the operation efficiency and facilitate the subsequent modification.

3.3. Texture Mapping. Texture mapping can be realized by scanning or digital photography, such as mapping with image processing software Photoshop or directly drawing texture on 3D surface in 3D drawing tools. This process specifically assigns a texture coordinate to each vertex on the polygon mesh, which is known as UV coordinates in the case of 2D; this process can be completed by explicitly assigning vertex attributes and manually editing in 3D modeling package, where the conversion from 3D space to texture space can also be realized by plane projection cylindrical mapping or spherical mapping application.

The dynamic mapping technology in Zbrush is also selected, which benefits from Zbrush's powerful drawing function. We can bake the texture in the photo directly to the 3D character model, draw the texture color on the character model through photos from different angles, then spread it

through UV (equivalent to cutting the surface of the 3D model), and finally integrate the finished texture into a 2D texture map.

4. Test Results and Analysis

All our model experiments were trained on a NVIDIA 1080tigpu. In the generator training stage, we set the number of training rounds of the model to 10,000 rounds. The input dimension of the model is 35, the number of convolution layers of the encoder is 3, the maximum length of each convolution kernel is 5, the dimension of the decoder RNN is 1024, the dimension of Prenet is 256, the learning rate is set to 0.001, the gradient clipping threshold is set to 1, and the weight attenuation is set to $1e^{-6}$. Batchsize is set to 40, and seqlen is set to 125. In addition, the learning rate of the discriminator is 0.001, the weight attenuation is $1e^{-6}$. The training set takes 80% of dance data, and the rest are adopted by the test set, and the dance data were taken from Tea things, Qianbian Dance, Cup Dance, and other forms in southeast Guangxi.

4.1. Results Feature Extraction. In the stage of dance feature extraction, the influence of different data processing methods on the final loss function is analyzed. Without filtering out the wrong data, the loss value using the model characteristics will be greater than the loss value without using the model characteristics. As the dimension of dance data increases, the noise may also increase, but the motion features may be useful for the final generation effect of dance, so it is better to use interpolation function to supplement the missing values of dance data than not. It can be seen from Table 1 that whether to filter out the wrong data has the most important influence on the final result. As long as the key points of the wrong human posture are removed, the final result will be significantly enhanced.

4.2. Results of Dance Sequence Generation. The training set of the dance data set is preprocessed, and the dance sequence is segmented according to the settings of seqlen = 125 and batchsize = 40. The similarity between the corresponding dance sequence and the generated dance sequence is calculated. If a piece of dance has appeared in the previous training data, then directly calculate the similarity between the dance sequence and the real dance sequence to measure the actual generation effect of dance. The similarity of five dances in southeast Guangxi is compared, and the results are shown in Table 2.

As can be seen from Table 2, the similarity of dance display in southeast Guangxi after image-based 3D reconstruction is high, among which the Money whip dance is the highest, reaching 93.29%. However, the similarity of reconstructed Chikuma dance is relatively low, which may be due to the fact that it pays more attention to the simulation of animal behavior, and its recognition degree in dancers' modeling and movement arrangement is low.

TABLE 1: Extraction effect of dance data set.

Method	Train	Valid	Test
Frames not dropped	20.39	22.28	32.89
Interpolation	17.23	19.32	25.83
Frames dropped	11.09	12.20	16.36
Interpolation	9.89	11.90	14.23

TABLE 2: Comparison of dance similarity in southeast Guangxi.

Name of the dance	Similarity (%)
Tea-picking dance	87.49
Money whip dance	93.29
Playing cup dance	92.20
Cardano ring dance	89.20
Chikuma dance	68.76

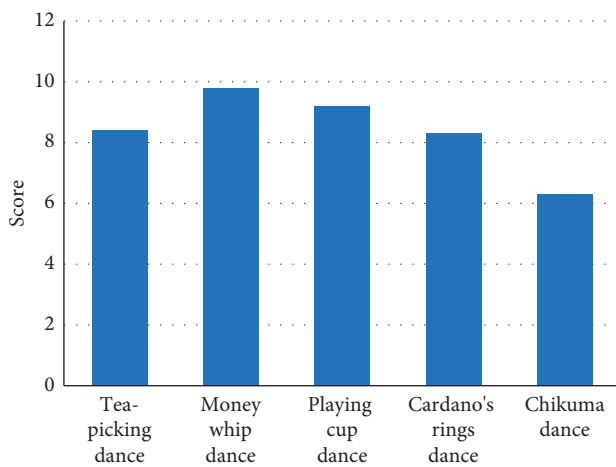


FIGURE 5: User's score of dance authenticity.

4.3. Authenticity Score Results. The degree of reduction of the reconstructed dance was investigated. We invited 20 local observers to conduct scoring experiments and showed each observer 15 pieces of five different kinds of reconstructed dances. Each observer scores according to the fidelity of the dance where the highest score is 10 points and the lowest score is 0 points. The scores of each model are averaged according to the raters' scores of 15 videos, and finally the scores of all raters are averaged, so that the trueness scores of each model can be obtained.

As can be seen from Figure 5, the score of users on the authenticity of dance in southeast Guangxi is quite consistent with the result of the similarity of dance sequences, with Tea-picking dance, Money whip dance, Playing cup dance, Cardano ring dance, Chikuma dance scoring 8.4, 9.8, 9.2, 8.3, and 6.3, respectively. Therefore, the dance display in new rural areas based on 3D image reconstruction technology has gained excellent user evaluation.

5. Conclusion

In this paper, aiming at the dance display in new rural areas based on 3D reconstruction, the 3 character reconstruction and texture mapping based on images are deeply studied, the

process of dance display in southeast Guangxi under 3D reconstruction technology is introduced, and five types of local dance are selected for test and analysis. The results show that the pose sequence generated by the restructured dance model is reasonable, where there are fewer incorrect and unreasonable pose frames, and the similarity with the dance in the original area is high. In terms of users' evaluation, the dance generated by 3D reconstruction technology has a good effect in restoring the authenticity. Among the five dance types displayed, the reorganization effect of Qianbian Dance is the best, with a similarity of 93.29% and a user authenticity score of 9.8.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by "Research on art and Culture Construction of New Rural Communities in Guangxi, taking southeast Guangxi as an example." The project number is 2020YJJD0009.

References

- [1] J. Rett and J. Dias, "Laban movement analysis using a bayesian model and perspective projections," *Frontiers in Brain, Vision and AI*, vol. 1, no. 10, pp. 183–211, 2008.
- [2] J. Wang, *Overview of Contemporary New media Dance Development*, p. 59, Nanjing Art Institute: Nanjing Art Institute, China, 2011, in Chinese.
- [3] L. Long, "Research on live transmission and green development of culture in ethnic areas in the new era," *Journal of Baoji University of Arts and Sciences: Social Science Edition*, vol. 33, no. 6, pp. 71–76, 2019, in Chinese.
- [4] L. Xue, "Research on the cultural inheritance system of minority dance in the new era," *Guizhou Ethnic Studies*, vol. 40, no. 3, pp. 116–119, 2019, in Chinese.
- [5] Y. Zhao and J. Wang, "Research on the construction of traditional ethnic dance cultural system in the new era," *Sichuan Drama*, vol. 48, no. 9, pp. 157–159, 2019, in Chinese.
- [6] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura, "A virtual reality dance training system using motion capture technology," *IEEE Transactions on Learning Technologies*, vol. 4, no. 2, pp. 187–195, 2011.
- [7] H. Zhao, H. Zhu, X. Mu, and M. Zhu, "3D folk dance display method driven by motion capture data," *Application of Computer System*, vol. 25, no. 6, pp. 136–140, 2016, in Chinese.
- [8] Z. Wang and A. A. R. O. N. Li, "Comparative study on digital rescue and reconstruction methods of minority heritage at home and abroad," *Journal of Mianyang Normal University*, vol. 35, no. 3, pp. 126–128, 2016, in Chinese.
- [9] Y. Huang and G. Tan, "Research on digital protection and development of intangible cultural heritage in China," *Journal of Huazhong Normal University*, vol. 29, no. 3, pp. 49–55, 2012, in Chinese.

- [10] S. Lai, "Multidimensional speculation on the application of digital means in the protection of intangible cultural heritage," *Research on Design Art*, vol. 31, no. 1, 35 pages, 2014, in Chinese.
- [11] Z. Liu, H. Qin, S. Bu et al., "3D real human reconstruction via multiple low-cost depth cameras," *Signal Processing*, vol. 112, no. 1, pp. 162–179, 2015.
- [12] O. Vogel, M. Breu, and J. Weickert, "Perspective shape from shading with non-lambertian reflectance," in *Proceedings of the 30th DAGM Symposium on Pattern Recognition*, Berlin, Heidelberg, June 2008.
- [13] Z. Cao, T. Simon, and S. E. Wei, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, Hawaii, HI, USA, July 2017.
- [14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural networks," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 3104–3112, Montreal, Canada, November 2014.
- [15] C. Chan, S. Ginosar, and T. Zhou, "Everybody dance now," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5933–5942, Seoul, Korea, October 2019.
- [16] M. Yu, F. Su, and Y. Yang, "3D reconstruction algorithm based on stereo vision," *Computer Engineering and Design*, vol. 34, no. 2, pp. 730–734, 2013, in Chinese.
- [17] S. Zhu, B. Zhou, and Z. Liu, "A stereo matching algorithm based on phase," *Industrial Instruments and Automation*, vol. 34, no. 2, pp. 101–105, 2013, in Chinese.
- [18] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–8, IEEE, America, Minnesota, USA, June 2007.
- [19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [20] G. Ziegler, "GPU point list generation through histogram pyramids," *Technical Report*, vol. 4, no. 6, pp. 85–97, 2006.

Research Article

Hotspots and Cutting-Edge Visual Analysis of Digital Museum in China Using Data Mining Technology

Yajing Hou , Lijun Xu , and Lu Chen

Institute of Art and Design, Nanjing Institute of Technology, Nanjing, China

Correspondence should be addressed to Yajing Hou; hoyajing@njit.edu.cn

Received 25 April 2022; Revised 9 May 2022; Accepted 11 May 2022; Published 25 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yajing Hou 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.

During the last several years, the building and development of digital museums has grown in importance as a study issue of increasing importance. On the other hand, systematic and extensive literature study on digital museums is rare in the academic community throughout the world. This paper employs data mining technology to conduct a comprehensive analysis of the total amount of academic literature, research hotspots, frontiers, and trends in the field of digital museums in China since the beginning of the twenty-first century, including both historical and contemporary data. In this research, the CNIK database and the CiteSpace program are utilized. The findings revealed that the quantity of published literature expanded significantly between 2000 and 2021, with some variations along the way, but that the general growth rate remained consistent. Colleges and universities are the driving force behind academic research in the field of digital museums; research institutes and big museums play a key part in the academic research that is being conducted by digital museums. Cooperation between research institutes, on the other hand, is severely lacking. Furthermore, the advancement of digital technology is an unavoidable byproduct of the efforts to transform the digital museum into a smart museum, as previously said. When it comes to digital museum development in the postepidemic period, the optimization and updating of a user-centered information service platform is the most important step toward long-term success. In order to maintain the richness of Chinese traditional culture while also meeting the expanding cultural requirements of the general public, China's digital museum research has as its ultimate objective the construction of sustainable digital museums that are appropriate for the country's national conditions. The findings also demonstrate that the construction of a Chinese Digital Museum is a study issue with distinct Chinese features that has the potential to contribute to the preservation of Chinese cultural heritage, both tangible and intangible. This research gives insights into the following aspects: researchers and practitioners from across the world will work together to promote a better knowledge of the building and growth of the digital museum in China, among other things.

1. Introduction

Since its inception in the 1990s, the digital museum has amassed more than two decades of experience. A global perspective reveals that foreign development is generally quick and comprehensive and that there are several eye-catching building projects that should be studied and promoted throughout the world. Despite the fact that China got off to a late start, the country has achieved many incredible things.

The Library of Congress sponsored the “American Memory” initiative in 1990, which served as the official launchpad for the “Digital Museum” movement. It was in

1995 that the Museum Internet System was officially established in the United States, which integrated many collection information data into the network system, allowing the information of museum collections to transcend the limitations of time and space, and thus signaling the maturation of the digital construction of American museums. Since 2010, several museums have begun to seek the refined services of the target audience, subdivide the target population on the museum website, and adapt popular science information for them in order to continually extend their impact and enhance the attention of users. Since 2011, with the help of powerful digital technology and Internet technology, Google has collaborated with museums

around the world to launch the Google art project, which uses Google Street View technology to shoot the real scene inside the museum and equipment with an accuracy of up to 7 billion pixels to shoot famous historical paintings in the museum for global users to enjoy [1]. Google art project uses Google Street View technology to shoot the real scene inside the museum and equipment with an accuracy of up to 7 billion pixels to shoot famous historical paintings in the museum for global users to enjoy. Visitors may choose from more than 18 languages, including English, French, and Japanese, as part of the program. The Canadian Heritage Information Network, or CHIN, has collaborated with major institutions to create the Virtual Museum of Canada, which is now open to the public (VMC). Beyond virtual exhibitions, the website also includes special columns such as interesting games, a teacher center, community memory activities, and online shopping to help bring the digital museum even closer to people's lives. This position in social and cultural service, integrating education and entertainment, has been described as follows: social and cultural service position integrating education and entertainment [2].

The European virtual museum was formally opened to the public in 2008, marking the first anniversary of the museum's creation. It is based on a real museum and incorporates digital assets from the digital library, museum, and archives into one cohesive whole. More than 2000 years of human history have been preserved in digital archives in Europe [3], making them the world's largest repository of historical information. In terms of digital construction, the Louvre Museum in France has broken through the static Internet page and is reshaping the museum website through the use of virtual panoramic technology. Visitors to the website may engage in "virtual tourism" and visit various locations around the museum by using the mouse to move up, down, left, and right through the many rooms [4]. In 2013, the British Museum utilized the crowdsourcing platform to digitize its collections, which was a first for the institution. It completed the tasks that previously required a large number of professionals' long-term efforts and even commercial solutions and it established a new way of working in the field of culture and the Museum [5]. This was accomplished through the joint efforts and cooperation of offline and online, professionals, and volunteers.

Despite the fact that Asian nations implemented museum digitization later than western countries, they were fast to identify a route for developing the digitalization of local museums in accordance with their own growth and national requirements. During 2006, Japan established the "Digital Museum Research Association," which investigated a mode of collaboration between the government and large corporations that combined the scientific and technological R&D capabilities of large corporations with the cultural relics resource capabilities of museums [6]. As a result of its "borderless" Digital Art Museum, which completely relies on modern digital technology to show art works, it has wowed visitors from all over the world in 2018, providing ample evidence of the success of its digital development strategy.

Taiwan has a head start on the Chinese mainland when it comes to the digital creation of museums. Taiwan's "scientific committee" started the "Digital Museum" initiative in 1998, with the goal of integrating museum materials and creating a museum with local features in order to broaden the meaning of online education [7] in the country. The Natural Science Museum of Taiwan, in collaboration with Jinan University of Taiwan, is putting the idea into action. The "Taiwan collection digitization project" was established in 2001 with the goal of digitizing Taiwan's collections, creating a database, and making it available to the public over the Internet [7]. As a result, the well-known "Imperial Palace digital museum" project in Taipei has been progressing at a quick pace as well. The accomplishments in the areas of collecting, application, and instruction offered useful experience for the Chinese mainland in the area of museum digitization.

The project of "cultural relics survey and database management system" was begun by the China National Cultural Relics Bureau in September 2001, marking the first time that the Chinese mainland has carried out the National Museum digitization effort [8]. The computer network information center of the Chinese Academy of Sciences has built a virtual museum and a China popular scientific Expo based on the database information resources of the Chinese Academy of Sciences, using web technology and multimedia technology. Founded in 1998, it is the first virtual museum group in China, with six display spaces spanning the domains of natural science and social science, respectively. Its digital building knowledge and experience has been widely shared and studied by others. Following that, the Beijing Palace Museum, the Nanjing Museum, and the Dunhuang research center all began working on the digital creation of their own museums. A number of them, such as the Palace Museum and the Dunhuang research institute, have chosen to collaborate with internationally renowned cultural and museum organizations in order to develop, apply, and show Chinese cultural heritage to a wider audience throughout the world. Furthermore, the digital development of Chinese museums does not overlook the importance of education as a basis. To encourage the development of digital resources in colleges and universities, the "action plan for the revitalization of education in the twenty-first century" in China launched the project "construction of online public resources of Modern Distance Education - Construction of University Digital Museum" at the end of 2001 [9]. It was in 2012 that Baidu Encyclopedia collaborated with eight well-known museums in China to launch the Baidu Encyclopedia digital museum, which introduced cultural relics through text and image as well as interpretation audio, animation, and virtual reality, among other methods, greatly improving the user experience and allowing the cultural relics of the museum to come into the public's vision in a more personable, easy to understand, and convenient manner [6, 7].

Through the use of algorithms, data mining is the process of extracting hidden and market-value information from massive amounts of data. The field of data mining technologies is typically associated with computer science and technological fields. Many other technologies, including

as statistics, online analysis and processing, information retrieval, machine learning, and so on, can be used to fulfill the aforementioned aims. In data depth mining, the most commonly used approaches are clustering, classification and prediction, and association and deviation detection [10], among others. Generally speaking, the functions performed by data mining technologies are primarily divided into two categories: (1) characterizing specific objects and (2) making appropriate projections [11].

CiteSpace software is utilized for bibliometric analysis in this study to mine the significant documents of digital museums released in China from the beginning of the twenty-first century (from 2000 to 2021). The following issues are addressed in this study, which is conducted in the setting of China: there will be a statistical analysis on the number of academic achievements representing China's digital museum; (2) distribution and cooperation of document quantity analysis and research institutions in different periods; (3) exploration of the evolution of the digital museum from 2001 to 2021, as well as identification of research hotspots; and (4) identification of research frontiers and trends.

2. Materials and Methods

The topics covered in this part are primarily four: (1) research tool selection; (2) research data collection; (3) parameter configuration and analysis of the CiteSpace; and (4) an explanation of the primary reference indexes (see Section 1).

2.1. Research Tool Selection. Chen Chaomei, professor of computer and information technology at Drexel University, developed the CiteSpace program in 2004 [12]. CiteSpace is citation visual analysis software that allows users to visually analyze citations. With visual approaches, it is capable of objectively processing a huge quantity of scientific literature data and presenting the knowledge structure, law, and distribution included in text materials, resulting in the creation of a "scientific knowledge map" [13]. A wide range of academic study subjects, including library, information, and archives management, management science and engineering, education, sociology, and other fields [14], have been more interested with and applied to the software. The primary reason for using this tool in this study is that it can create a knowledge map of a specific field through visualization functions such as collaboration between authors and institutions, keyword co-occurrence and clustering, emergence, and so on, allowing researchers to explore the development of a subject knowledge field and its research hotspots in greater detail.

2.2. Research Data Collection. The CNKI database (China's National Knowledge Infrastructure) was chosen for this study since it is one of the most important and reputable Chinese academic databases in the country. The information utilized in this study was acquired from the CNKI database on February 29, 2022, and it included all papers pertaining to

the digital museum from January 2000 to December 2021, which was the time period under consideration. CNKI's database contains the first academic publication on digital museums published in 2000, and this is the year in which the data for this study were originally compiled.

Using the term "Digital Museum" as the subject, this study searched the relevant literature in the CNKI database, yielding a total of 1763 publications. Furthermore, we carefully removed the possibility of contamination from seven unrelated literature and only included peer-reviewed scholarly publications on digital museums in our analysis. Finally, 1756 publications were chosen as the subject of the investigation. During February 2022, CiteSpace, V. 5.8. R1 (<https://Citespace.podia.com/download>, accessed on 29 February 2022) was used to process all of the selected articles related to the digital museum.

2.3. Parameter Setting and Analysis. Before you begin processing the data, you must first configure the settings in CiteSpace. (1) The node type was chosen in accordance with the appropriate analysis; (2) the time slice was set to 2000–2021; (3) the length of each time slice was "2"; (4) the selection criteria were set for the top $N = 50$; and (5) pruning was set as the pathfinder in the selection criteria. The default values have been assigned to the remaining parameters.

Second, CiteSpace is used to evaluate the data from three primary routes once the parameters have been specified in order to address the research difficulties. For the first path, we looked at the quantity of papers published during this time period (as supplied by the CNKI study) and the cooperation network of research institutions, in order to gain a more comprehensive view of the whole research environment. The second approach was to do keyword-based co-occurrence analysis of the data. CiteSpace's co-occurrence analysis generates a map of keyword co-occurrences as well as a map of time-zone co-occurrences. It is possible to obtain the research hotspots of China Digital Museum at different times by doing a thorough study of these two maps. This is the third method, which is known as noun-term burst analysis, and it may be used to demonstrate the quick change of keywords in a short period of time while emphasizing the abrupt shift of keywords [15]. It might also represent the current state of research in this topic, as well as new trends.

2.4. Main Metrics Analysis in CiteSpace. Following the completion of CiteSpace according to the specified specifications, three metrics (betweenness centrality, silhouette, and modularity) are primarily utilized to assess the logic of the map that has been produced. When it comes to structural indexes, betweenness centrality is an essential one to consider. It refers to the ratio of the shortest path via a point (connecting two locations in a network) to the total number of shortest path lines between the two points [14]. Nodes having a mediation centrality greater than 0.1 are referred to as key nodes. The presence of purple circles on the map implies that the betweenness centrality is more than 0.1 [16]. The modularity (Q) value and the mean silhouette (S) value were the two most important variables to assess for the

influence of the clustering map. The modularity index of a network is represented by the term modularity. The greater the value of this parameter, the better the clustering outcome of the network will be. The Q value ranges between 0 and 1, and a value greater than 0.3 indicates that the split clustering structure is statistically significant. The closer the value is to one, the greater the effectiveness of the clustering effect. The silhouette (S) index is used to assess the homogeneity of a network's structure. The closer it is to 1, the more accurately it reflects the homogeneity of the network. When the value is more than 0.5, it is acceptable to conclude that the clustering result is reasonable. Generally speaking, when the S value is 0.7, clustering is both efficient and convincing [12].

3. Descriptive Statistical Analysis of the Research Literature in the Field of Digital Museum in China

This section discusses two issues: (1) the analysis of the number of papers published in different periods aims to gain a comprehensive understanding of the dynamic distribution of research in this field over the past 20 years; and (2) the analysis of the number of papers published in different periods aims to gain a comprehensive understanding of the dynamic distribution of research in this field over the past 20 years. (2) The cooperation study of research institutions seeks to identify prominent academic institutions as well as the nature of the cooperative connection between research institutions, among other things.

3.1. Distribution of the Selected Papers in the Field of Digital Museum. As reported in the yearly analysis report provided in the CNIK database, 1756 papers relating to intangible cultural assets were published between January 2000 and December 2021, representing an average annual publishing rate of 83 publications each year. Figure 1 depicts the distribution track of the number of selected research articles from 2000 to 2021, as represented by the number of selected research papers. There are three periods that may be identified: the steady development period (2000–2010), the rapid growth period (2010–2015, 2017–2019), and the fluctuation decrease period (2000–2010), (2015–2017).

The number of papers published during the time of stable development, particularly between 2000 and 2002, is extremely modest, accounting for only 0.95 percent of the total number of documents issued during this period. Since 2002, the number of papers published has continuously climbed, with an average of 29 articles being published each year. This demonstrates that relevant researchers have begun to pay attention to the growth of the research field of the digital museum, but that the research field is still in its early stages of development, as demonstrated by the findings of the study. Following that, the number of articles published each year has increased significantly since 2010, reaching its first peak (114) in 2015. This can be attributed to an increase in the number of scholars who have invested in the relevant research field of digital museum, which has received an increasing amount of attention from the academic

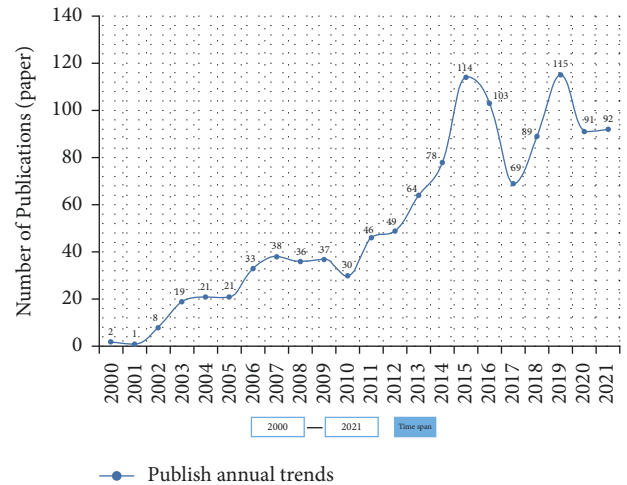


FIGURE 1: Distribution of selected digital museum papers from 2000 to 2021 in China.

community in recent years. During the next four years, there was a large fluctuation in the amount of study material published. From 2015 to 2017, the number of papers published declined precipitously, but the number of articles published again increased to reach its second peak (115) in 2019. Each number's trajectory is shown as a horizontal "s" curve, suggesting that there have been considerable changes when compared to preceding periods. From 2015 to 2021, the number of published papers accounted for 58.0 percent of the total, demonstrating that digital museum research is continuously becoming more mature, which may be attributable to the steady establishment of a research system in this subject.

3.2. Distribution of Core Institutions in the Field of Digital Museum. The institution network map depicts the spatial distribution of research power in this discipline, and it provides additional information. CiteSpace's cooperative network analysis feature allows researchers in the field of digital museum research to get insight into the network relationships of other institutions in the field. The network link might naturally depict the collaboration between institutions and serve as a point of reference for scientific assessment institutions in the academic domain, according to some researchers. Figure 2 depicts the distribution network map of digital museum research institutions that was created after the operation was completed. The size of the nodes represents the number of journal papers published by the research institution, and the strength of the connections between nodes represents the intensity of cooperation between different institutions (see text).

In Figure 2, it can be seen that the research sample consists of 310 nodes connected by 39 connections, and the network density is 0.0008, showing that there are many institutions exploring digital museums, but the level of interaction between research institutions is rather low. The fact that the institutions researching digital museums are dispersed and have not created a strong and wide



FIGURE 2: Institution cooperation network map of digital museum research from 2000 to 2021 in China.

TABLE 1: The ten most productive research institutions in digital museum in China.

Rank	Institution	Year began	Number of papers	Cooperation degree
1	Nanjing University	2002	22	3
2	National Museum of China	2006	13	0
3	Shanghai University	2012	13	1
4	Institute of Information Technology, Beijing Union University	2004	11	0
5	Beijing University of Aeronautics and Astronautics	2003	10	0
6	Communication University of China	2012	10	1
7	Suzhou Higher Vocational and Technical School of Tourism and Finance	2014	7	0
8	China Urban Planning and Design Institute	2019	7	0
9	The Palace Museum	2003	7	0
10	Department of Computer Science, Northwestern University	2004	6	0

cooperative network link can also be observed. Figure 2 depicts the formation of an obvious institutional collaboration network, which is comprised of the Guangdong Museum and the Department of cultural relics and museology of Zhejiang University, as seen in the figure. Finally, it is discovered via literature research that the cooperation network has only talked about the issue of “creating a bridge of connection across museum collections” since the 2014 International Museum Day [17], and that it has not engaged in any significant collaboration since then [18]. A collaborative study project involving Zhejiang University, Nanjing University, and four other universities investigated the digital conservation of sports cultural assets.

A further data mining operation is carried out on Figure 2 in order to thoroughly study the accomplishments and collaborative relationships of research institutions, and the top ten research institutions in terms of the number of documents produced are identified, as shown in Table 1. On the list of starting texts, the top three institutions are Nanjing University (number 22), the China National Museum (number 13), and Shanghai University (13). According to Figure 1, the cooperation network association between the top three universities with research results is not statistically significant, demonstrating that even the institutions with high research results do not necessarily have close collaboration with other research institutions. When looking at

research units from the perspective of cooperation degree, the cooperation degree of main institutions is low, indicating that, at present, the majority of domestic scholars’ research on digital museums is carried out by independent institutions, and there is still a large space for cooperation among institutions, which requires the establishment of a more in-depth cooperation relationship between research institutions from interdisciplinary and cross-regional perspectives.

Table 1 shows the top ten most prolific institutions in the field of China’s digital museum, while Table 1 contains the top ten most productive institutions overall. Six universities, two scientific research institutes, and two big museums are among the institutions on the list. This demonstrates that universities are the driving force for scientific investigation in the field of digital museums. Institutions such as research institutions and huge museums play a key role in academic research, which is being driven by digital museums. To provide an example, the State Vital Laboratory of new computer software technology at Nanjing University is tasked with the critical responsibility of conducting research into key technologies for the development of China’s digital museum. This is a significant undertaking. The research on digital humanistic storage [19] is being carried out in parallel by Nanjing University and the Tibet Institute for Nationalities, with the goal of laying the theoretical groundwork for the “National

digital museums are typically created by museums themselves. Digital websites that provide services for material and intangible cultural heritage are being improved in order to realize the vision of equitable enjoyment of cultural and Expo resources for the entire society and to meet the maximum demand of individuals for cultural and Expo resources [18]. This is in order to realize the vision of equitable enjoyment of cultural and Expo resources for the entire society and to meet the maximum demand of individuals for cultural and Expo resources. Prof. Zheng thinks that the creation of a digital museum is founded on the implementation of museum-related services. In its digital version, it is an information service system for the collection, protection, administration, usage, and transmission of natural and cultural assets in the digital domain. Its presentation and instruction can be carried out in physical museums or across a network [21], depending on the situation.

- (2) *Digitalized Museum*. It was pointed out by Liu and Zhu that the term “digital museum” has two meanings: “the first refers to the digitization of a physical museum”; and “the second refers to the emergence of virtual museums as a result of the information age’s approach, which is not only the result of the digitization of traditional museums, but also the symbol of the modernization of traditional museums [22].” According to some scholars, the term “Digital Museum” is an acronym for “digital museum” in order to separate it from the term “physical museum” [23]. However, some researchers have expressed differing views, claiming that the digital museum is a modern museum following information and digital transformation, which is not the same as a digital museum that exists simply on the Internet [24].
- (3) *Virtual Museum*. According to Li et al., the virtual museum refers to the re-establishment of the conventional museum through the application of digital technology. There are several components to this digital endeavor. Among the major components of the virtual museum are the fundamental operating platform, the browsing platform, and the administration platform [25]. According to Lv et al., the virtual museum makes use of virtual reality technology to generate simulation models or scenes, which are utilized to convey the overall image of the museum, or situations that were actual in history but have since vanished or are in risk of disappearing completely. Its most distinguishing aspect is its high level of involvement. Visitors can attain the goal of autonomous inquiry learning in the roaming system by engaging with the models and buttons depicted in the figure [26].
- (4) *Wisdom Museum*. In the same way that the notion of “smart earth” developed, the concept of “smart museum” did as well. It was initially proposed in

2012, and it has since gained popularity. Earlier this year, IBM announced a partnership with the Louvre in Paris, France to optimize the service, operation, and management of the museum by digital technology methods, therefore designating it as the world’s first smart Museum. A major project by the State Administration of Cultural Relics of China, titled “feasibility research on the construction of China wisdom Museum” [19], was also launched in the same year. The concept of smart museum building was first proposed by the Chinese government in 2013. A widely accepted definition of smart museum currently exists as follows: “by fully utilizing new generation information technologies, such as the Internet of things, cloud computing, big data, and artificial intelligence in the operation of the museum, we are able to perceive, calculate, and analyze information pertaining to people, things, and activities associated with the museum’s operation, realize the intellectualization of museum collection and ensure its protection, display, and dissemination, and conduct research and management of a smart museum.” A comparison was made between the smart museum and conventional museum as well as digital museum, and the “ecological chain under the smart Museum mode” was proposed, which is based on “people,” “things,” and “Museum” [27]. In a systematic manner, the concept of smart Museum was introduced, the similarities and differences between smart Museum and digital museum were discussed, and it was concluded that smart Museum has realized two-way information exchange and effective management of collaborative relationship among “things, numbers, and people” of Digital Museum [28].

4.1.2. Research on Digital Museum in Digital Protection of Intangible Cultural Heritage. Cultural heritage, cultural heritage digitization, cultural relics protection, intangible cultural heritage (protection), traditional villages, and other terms are included in this field.

Human civilization and the crystallization of our predecessors’ knowledge are both documented in our cultural legacy, which serves as an essential testimony to the progress of human civilization. Mankind has a responsibility to safeguard and pass it on to future generations. The application of digital technology in the sphere of cultural heritage, according to Song, has the potential to effectively tackle the problems of correct storage, development, and exhibition at this level. To be more specific, the digital filing of three-dimensional modeling and other technologies can exhibit a large number of cultural relics photographs in 360 degrees, allowing the cultural relics to be kept in a more rigorous environment, reducing man-made harm, and extending their service life [29]. It was proposed by Ding and Guo [30] that digital technology represented by augmented reality plays an increasingly important role in cultural heritage protection, and a detailed analysis was conducted of the

application of domestic digital technology in cultural heritage protection. Zhou and Yang stepped away from the perspective of technical research and proposed an alternative definition of cultural heritage digitization, which refers to the process of creating relationships through digitization, exploring the core value of cultural heritage, connecting it closely to real life, and assuming today's social and cultural roles [31].

In the United Nations Convention on Cultural Heritage, nonmaterial heritage, culturally significant forms of expression and expressions of noncultural arts and crafts, and the instruments of their protection, are defined as "cultural heritage, nonmaterial objects, performances, and their related groups," according to the United Nations Educational, Scientific, and Cultural Organization [18, 32]. In 2005, the general office of the State Council issued opinions on strengthening the protection of China's intangible cultural heritage, which stated unequivocally that "we should make true, systematic, and comprehensive records of intangible cultural heritage, in addition to establishing archives and databases." It was the following year that the "China Intangible Cultural Heritage Network China Intangible Cultural Heritage digital museum" (<http://www.ihchina.cn>), the world's first worldwide gateway website dedicated to the conservation of China's intangible cultural heritage, was launched.

The specific application of digital technology in the protection, development, and utilization of intangible cultural heritage and its cultural and economic value was developed by Run from the perspectives of museum information technology's investigation and research on intangible cultural heritage, data sorting, data protection, and display [33]. Accordingly, researchers at Tsinghua University, such as Ma et al., investigated the reasons why fundamental digital technologies such as text, picture, audio, and video are still the most commonly used technical applications in intangible cultural heritage digitization, and concentrated on the research and practical results of 3D scanning and reconstruction, virtual reality, augmented reality, motion capture, and other comprehensive technical applications. Although the application of emerging technologies in intangible cultural heritage digitization is still at an exploratory and embryonic stage, it is expected to have a significant impact on future digital protection, particularly in the areas of exhibition and communication [34], as previously stated. Zhou and Yang analyzed and studied the concept and characteristics of intangible cultural heritage digital museum and concluded that the intangible cultural heritage database, the intangible cultural heritage virtual exhibition hall, and the protection and education window are three important components of the intangible cultural heritage Digital Museum [31]. Yan et al. analyzed and studied the concept and characteristics of intangible cultural heritage digital museum, and concluded that the intangible cultural heritage database [35–37]. A large number of other scholars have conducted case studies and strategic analyses on the digital protection of intangible cultural heritage from the perspectives of local characteristic culture, dialect, arts, sports, and so on, thereby broadening the research latitude available in this area.

4.1.3. Research on Information Technology of Digital Museum. Among the terms used to describe the field of information technology are digital technology, virtual reality (VR), augmented reality (AR), new media, and other terms. The development of a digital museum is inextricably linked to the assistance provided by information technology.

According to a survey of the literature, virtual reality technology and augmented reality technology are now two of the most active research areas in the field of digital museum exhibition technology. A high-tech innovation that has emerged in recent years is virtual reality (VR), often known as "spiritual environment technology." [38] It has the ability to present the audience with an immersive experience through digital reality and virtual digital picture, and it has a unique immersion experience and interactivity. AR technology is a development of virtual reality technology. In this way, the audience can see the virtual model object in the foreground of the real environment, visualize and visualize the abstract learning content, improve the audience's sense of existence as well as their intuition and concentration, and have a more immersive exhibition viewing experience [39, 40]. These technologies are intended to improve the interaction between museum visitors and the museum itself, as well as the overall visitor experience.

The "virtual palace museum" and the "digital Dunhuang" in China are the forerunners in the application of this technology in the field of museum research, providing a cutting-edge learning experience for other museums [40]. For theoretical study purposes, Wang and a large number of other researchers have compiled a list of the relevant technologies of the digital museum. The most significant of them are 3D technology, 360 panorama technology, augmented reality technology, virtual reality technology, and so on. As well as technical assistance for the creation of Digital Museum [41], the development of hardware or technology such as high-definition printers and 3D printing also contributes to its advancement. In addition to the aforementioned technical characteristics, Wu summarized the current web page form of digital museum into four categories: plane web page form, panoramic web page form, 3D cultural relics web page form, and full 3D interactive [42]. This was based on his research into the information technology of digital museum based on the Web.

New media is a type of media that makes use of digital technology, network technology, and mobile communication technology to connect computers, mobile phones, digital television, and other terminals through network channels such as the Internet and mobile communication networks in order to facilitate the exchange and dissemination of user information [43]. A large part of the domestic study on digital museum new media communication is devoted to the investigation of Internet communication routes and challenges. They used unique museum new media communication means such as social media (Wechat and Microblog), webcast, mobile intelligent terminal navigation application, and others to conduct their research [44]. They also compared the communication benefits of the various methods they used. As an example, Hu and Liu developed a development strategy for promoting the Museum of

causes. First and foremost, as digital museum technology continues to grow and mature, huge museums are striving to be the first to incorporate digital technology into the digital building of museums, in which virtual reality technology plays an important part. In [38], Yang examined the major technologies that are involved in the application of virtual reality technology to digital museums. Using six virtual reality technologies in digital museums, Zhang et al. conducted an analysis of the technical routes of their application and proposed that, on the basis of not affecting the needs of virtual exhibition and access in digital museums, image-based panoramic virtual reality technology is more suitable for the digital construction of general local museums due to its short development cycle and low development cost [48]. University museums have become high-frequency keywords in part because the State Administration of cultural relics and the Ministry of Education issued a joint notice on strengthening the construction and development of university museums in May 2011, with the goal of strengthening the construction and development of university museums as well as fully developing their role in rejuvenating the country through education and science, learning society, and the consulship system. In the same year, Zhejiang University officially opened the IPv6 site of the China University Digital Museum, which is now available to the public. The digital museum, which is based on the collections of 21 colleges, contains hundreds of thousands of digital items [49] and is constantly expanding. Within a short period of time, university museum research has emerged as a topic of interest in the fields of culture and museum, education, and society.

While these key words were being used, the key words for museums such as the Palace Museum and the Capital Museum were also being used. This was primarily due to the fact that the Capital Museum hosted a Beijing Digital Library seminar in 2013 with the theme of “integration, innovation, and development: the Digital Library in the Constructing of a Cultural Power” [50]. A total of 68 researchers from large museums, universities, and many well-known enterprises from both home and abroad participated in the seminar. They demonstrated through presentations and exhibit the new direction, new achievements, and fresh experience in the field of digital museum construction that have emerged in the last two years, as well as jointly discussed museum construction and provided new ideas and method for building a digital museum. The Beijing Association of Science and Technology, the Beijing Municipal Bureau of Cultural Relics, and the Beijing Municipal Commission of Economy and Information Technology came together in 2005 to form the Symposium on Science and Technology. It is held every two years and has played a significant role in the establishment of China’s digital museum, which is now in its third year.

4.2.3. Stage 3 (2014–2019): Research on Digital Museum in New Media Environment. It was during this time period that smart museums, smart services, new media, digital media,

and high-frequency phrases and keywords relating to “design” appeared on the Internet. As can be observed, the wisdom museum has emerged as a new research center throughout this time period. The primary reason for this is that, after more than 10 years of building and expansion, the China Digital Museum has amassed an extensive collection of successful examples and research findings. The State Administration for Cultural Relics of China organized the important research topic “feasibility study on the building of China smart Museum” [19] in 2012, and the concept of smart museum construction was first proposed in 2013. As part of its national wisdom Museum pilot program, the State Administration of cultural treasures designated seven museums as the initial pilot units in 2014. Since then, an increasing number of academics have made investments in the study of this topic. Since 2018, smart museum research has emerged as a prominent topic in the museum sector, with the number of papers published in 2019 more than double the number of articles released the previous year. For the most part, the study is concerned with the concept analysis of smart museums [19, 26, 27], associated technologies [51, 52], case studies at home and abroad [53, 54], and the development trend [55, 56], among other things.

It was in 2015 that Chinese Premier Keqiang Li introduced the China Internet Plus Action Plan [57], which was first included in the government work report. On the one hand, the Chinese government premier introduced the notion of “Internet plus” for the first time in his two-year report, which was released in December. New media, as defined by Liu and others, is a new media form that is produced by the mix of conventional media with developing technologies based on the Internet, i.e., “traditional media Plus Internet.” In Digital Museums, new media technologies such as big data and cloud computing, mobile Internet technology and social network technology, flexible display technology and wearable devices, and holographic technology are prominent examples of new media technologies. According to some, new media technology is not just a method of communication, but also a mode of experiencing the world around us. It increases the number of visitors to the digital museum while also providing a higher-quality user experience. It is not difficult to conclude from a survey of the literature that academics’ focus has steadily changed from the early stages of museum web design and development [58] to the investigation of interactive design based on new technologies. Specifically, the interactive design of virtual reality technology [59, 60] and mobile client [61] is the focus of the current study.

4.2.4. Stage 4 (2020–2021): Research on Digital Museum in the Postepidemic Era. Postepidemic era, educational purpose, folk culture, and living inheritance are some of the keywords that appear often throughout this time period (Figure 4). In this age, a new COVID-19 “hot postepidemic era” was forming, mostly as a result of the advent of the new crown pneumonia outbreak in late 2019, which was the cause of the new epidemic. At several points during the epidemic’s emergency prevention and control in the first half of 2020,

public cultural service locations were temporarily shuttered, and offline cultural experiences were put on hold. Several local public cultural institutions have actively changed their service methods and implemented the “cloud” mode in order to meet the spiritual and cultural needs of the public, which includes the ability to seek knowledge and pleasure online without having to leave the comfort of their own homes [62]. The availability of Internet information services is particularly significant in this situation. Construction of online service platforms is moving at a faster pace, which is the general tendency. Because of this, experts have turned their attention to the study of the existing state of digital services provided by cultural and museum institutions, as well as the debate of the specific path for service upgrading, in the aftermath of the pandemic [63].

5. Research Frontiers and Trends in the Field of Digital Museum in China

Table 2 displays the phrases that have been used most frequently in digital museum research in China over the last 20 years, with a total of 53 keywords. By analyzing the data, it has been discovered that the five keywords of user experience (strength = 5.29), traditional village (strength = 4.98), smart museum (strength = 4.92), big data (strength = 3.59), and digital protection (strength = 3.28) represent the research frontiers of digital museum research in China in the recent years. The following will be contrasted and discussed in conjunction with a sample body of the literature with a high citation rate.

5.1. Research on User Experience Design of Digital Museum.

In recent years, as digital technology has progressed and matured, the research direction of the digital museum has gradually shifted away from simple technical research and toward research on user experience. This shift is intended to provide visitors with a more humanized human-computer interaction while also improving the overall experience of visiting and learning. The use of design methodologies to create the display, web page, and mobile client of a digital museum has been a research hotspot in this topic in recent years, particularly from the standpoint of user experience. Interface design [64], exhibition design [61], interaction design [61], and experience design [38] are some of the design research approaches that have been developed. Immersive interface design, experience design, and user experience assessment are just a few of the topics that have emerged as the current study trends.

In his paper, Wang argues that the notions of user experience and user journey are two fundamental concepts in the concept of experience design. The notion of experience design may be used to guide the building of a digital museum, which can assist the museum in better understanding the needs of users and reducing the psychological gap between the museum and the audience to the greatest extent feasible, says the museum. On the one hand, using the network benefits of a digital museum, a virtuous cycle of “sharing receiving resharing” of knowledge may be

generated among visitors, which can be used to pique the interest of future visitors in browsing by stimulating their browsing excitement. To some extent, digital museums may be considered to be cooperative endeavors involving the merging of technology with art, and the digital collection of information and culture can be considered to be a significant benchmark for the advancement and optimization of the museum sector in general [65]. Noteworthy is the suggestion made by Fan [66] that the future development of digital museums should not only be user-centered but should also take into account the efficacy of cultural information distribution. Consequently, he employs the coding and decoding theory of information transmission, along with the user interview technique, to develop the user experience evaluation index and user experience assessment model for the digital museum. By doing this study, we are able to close a research gap in the assessment process for user experience in the digital museum.

5.2. *Research on the Protection of Traditional Villages by Digital Museum.* Earlier this year, China’s Ministry of Housing and Urban-Rural Development officially began construction on the Digital Museum of Chinese Traditional Villages, with the goal of displaying excellent Chinese traditional villages on a digital platform, thus serving as both a “stage for publicizing Chinese traditional villages to the world” and a “window for the world to understand Chinese agricultural civilization.” The CPC Central Committee and the State Council issued the Strategic Plan for Rural Revitalization (2018–2022) in 2018, which defined traditional villages as “characteristic protected villages” and proposed to effectively protect the material and intangible heritage of traditional villages [67]. The plan was approved by the CPC Central Committee and the State Council in 2018. Throughout the years, study into the digital preservation of traditional villages has steadily gained in importance, and it continues to be a hot topic in the field of digital museum research.

The digital preservation of traditional villages can be accomplished in two ways: first, by archiving, displaying, and publicizing villages in the form of a digital museum; second, by assisting traditional villages in fine management, planning prediction, protection, and development status monitoring through digital means. Also being investigated is the unique application of digital technology in traditional village protection, which is now under development. Recent technologies include man-machine tilt photography, lidar scanning, offline sensors, large data platforms, 3D printing, 3D synthesis technology platforms, artificial intelligence, and others. The following are some examples: practice has shown that Guangzhou University has investigated the combination of “three-dimensional modeling + oblique photography + traditional villages” and has successfully realized the digital presentation and publicity of traditional villages through the new digital process of three-dimensional modeling, online platform construction, cultural and creative industry, and new media promotion. A platform for architectural heritage monitoring developed by Beijing

TABLE 2: The burst terms in digital museum research in China.

Keywords	Year	Strength	Begin	End	2000-2021
Physical museum	2000	4.72	2000	2013	=====
Digital museum	2000	3.73	2000	2013	=====
xml	2000	5.9	2002	2011	=====
Multi-media	2000	4.13	2002	2009	=====
Metadata	2000	4.01	2002	2015	=====
Digital watermarking	2000	3.59	2002	2011	=====
Virtual Reality	2000	4.82	2004	2009	=====
vrml	2000	3.59	2004	2007	=====
Virtual museum	2000	2.6	2004	2015	=====
web service	2000	2.35	2004	2009	=====
Digital archaeology	2000	1.96	2004	2005	=====
Digital library	2000	2.85	2006	2015	=====
Internet	2000	2.25	2006	2009	=====
Archaeological	2000	1.92	2006	2007	=====
Digital museum	2000	3.5	2010	2015	=====
library	2000	2.3	2010	2015	=====
University museum	2000	2.22	2010	2017	=====
Capital museum	2000	2.21	2010	2013	=====
Visitor	2000	1.89	2010	2011	=====
Information	2000	3.71	2012	2015	=====
acquisition	2000	3.68	2012	2015	=====
Information esources	2000	3.1	2012	2015	=====
Cloud computing	2000	2.56	2012	2015	=====
Museum Website	2000	2.21	2012	2015	=====
informationize	2000	2	2012	2017	=====
Archives	2000	2.75	2014	2017	=====
Intangible cultural	2000	2.72	2014	2015	=====
heritage protection	2000	2.43	2014	2019	=====
unity3d	2000	2.27	2014	2019	=====
Internet of things	2000	2.18	2014	2015	=====
Museum education	2000	2.18	2014	2015	=====
Digital media	2000	1.94	2014	2017	=====
Panoramic technology	2000	4.37	2016	2019	=====
Music Museum	2000	3.77	2016	2017	=====
Digital resources	2000	3.38	2016	2019	=====
Digital technique	2000	3.22	2016	2019	=====
New media	2000	2.87	2016	2017	=====
technology	2000	2.53	2016	2021	=====
Interactive	2000	2.43	2016	2019	=====
technology	2000	2	2016	2019	=====
"Internet +"	2000	2	2016	2019	=====
New media	2000	1.94	2016	2021	=====
Chinese traditional	2000	5.29	2018	2021	=====
village	2000	4.98	2018	2021	=====
Information	2000	4.92	2018	2021	=====
technology	2000	3.59	2018	2021	=====
International Museum	2000	3.28	2018	2021	=====
Day	2000	2.69	2018	2019	=====
Internet +	2000	2.67	2018	2021	=====
Public cultural services	2000	2.41	2018	2021	=====
User experience	2000	2.11	2018	2021	=====
Traditional villages	2000	2.05	2018	2019	=====
Wisdom Museum	2000	2.67	2018	2021	=====
Big data	2000	2.67	2018	2021	=====
Digital protection	2000	2.41	2018	2021	=====
the Palace Museum	2000	2.11	2018	2021	=====
Traditional residence	2000	2.05	2018	2019	=====
Artificial intelligence	2000	2.05	2018	2019	=====
Digital media	2000	2.05	2018	2019	=====
technology	2000	2.05	2018	2019	=====
New media	2000	2.05	2018	2019	=====
environment	2000	2.05	2018	2019	=====
Library school	2000	2.05	2018	2019	=====
cooperation	2000	2.05	2018	2019	=====

Architecture University has the capability of not only realizing traditional architectural monitoring, but also carrying out restoration prejudgment and simulated restoration [68]. For example, when it comes to the construction of a traditional village digital museum, Song and Zhao proposed that the next task should be to solve the efficiency of spatial data retrieval and the ability of real-time analysis of spatial data so that users can have a more process user experience [69]. Cao and Wu redesigned the information architecture, information presentation, and information dissemination in

the digitization of cultural heritage at the Traditional Village Museum, in order to address the problem of information understanding in the digitization of traditional villages [70]. They did so in accordance with the principles of information and interaction design, which they developed.

5.3. Development Trend of Future Digital Museum: Smart Museum. However, despite the fact that the concept of a smart museum was first proposed in 2014, the actual building and development of a smart museum in China is still in its early stages. Current intelligent museum architecture depends heavily on a number of developing technologies, including big data, cloud computing, and artificial intelligence [19], in order to actualize the intelligent working mode with the Internet serving as the medium. Since then, the research viewpoint offered by Smart Museums has expanded significantly. In particular, the research from the perspective of public service, as well as the research on the top-level design method of smart museums and the construction of modern smart museum systems, which has emerged as the research frontier in this field, is deserving of special attention and consideration.

In Li's opinion, the smart museum is a new information transmission mechanism developed between people, things, and data that can be used to better understand the world. It also represents a step forward in the preservation of museum collections and public services on all sides of the spectrum. Li conducted a thorough investigation of the process that occurs before, during, and after users browse the smart Museum, and he examined the smart Museum's implementation methods from the standpoint of public service [56]. The operation and maintenance management of the smart museum, according to Zhong and Zhang, is considered a system project. He points out that the smart museum is a subsystem project of the overall system project. He concludes by saying that a research method for top-level design of smart museums was also proposed for the first time by him at the same time. He pointed out that the top-level design of the smart museum is the overall design for the development form and progression path for a digitally enabled museum, which is dependent not only on information technology, but also on a research concept and a research method [19]. In the opinion of Bai S., the wisdom museum may be considered a closed-loop control system that integrates dynamic information collecting, intelligent processing, intelligent control, and analytical feedback. In her remarks, she pointed out that, in order to develop and expand the smart museum system, it is necessary to first establish data specifications and standards, then build the smart service and operating platform of the smart museum, and finally establish standardized operation and maintenance processes [71].

5.4. Application of Big Data Technology in Digital Museum. 2015 saw the release of the State Council's action plan for promoting the development of big data, which stated that we should encourage the development and application of big data, accelerate the opening and sharing of data, promote resource integration, improve management ability, improve the big data security system, and deepen the security pillar.

This demonstrates that the development and deployment of big data technology has advanced to the level of national policy in China [72]. China's Digital Museum has been doing research into big data technologies for some years, with the majority of its efforts concentrating on the preservation of intangible cultural treasures and the development of smart museums.

According to Yang when it comes to intangible cultural heritage protection, we can use big data's unstructured characteristics to develop local intangible cultural heritage management systems, build intangible cultural heritage big data platforms, realize multidimensional communication of intangible cultural heritage, and help the public understand *t*. The omnibus thinking method of big data, in addition, may be utilized to forecast and appraise the communication form and future development trend of intangible cultural assets in a given location [73]. The smart museum, which integrates the Internet of Things, big data, and "cloud computing" technology, can effectively carry out data analysis and processing, allowing the museum's services to be more intelligent and systematic as a result of this combination. At the moment, the use of big data technologies in the management and maintenance of smart museums is being considered. Network information technology and Internet of things technology, as well as real-time monitoring systems and intelligent agent terminal technology [71], are some of the most recent innovations to emerge.

5.5. Research on Digital Protection in Digital Museum.

When Dr. Zhao of Shandong University wrote his doctoral dissertation "Research on the protection and development of historical and cultural resources under digital survival" in 2014, he went into great detail about how important digital technology is to the protection of China's historical and cultural resources. Dr. Zhao is a professor of history and cultural resources at Shandong University. His research has revealed that, in addition to three-dimensional technology and virtual reality technology, digital image processing technology, multimedia technology, digital content management and publishing technology, three-dimensional scanning technology, and network technology are all essential for the protection of cultural resources [74]. Following a review of the literature, it is shown that the preservation of intangible heritage, cultural heritage, and traditional villages, all of which have been described above, have been the study frontiers of digital protection in the last three years.

The author, after examining the current state of China's intangible cultural heritage digital protection, proposed that in order to ensure the long-term development of China's intangible cultural heritage digital protection, it is also necessary to improve database construction, enrich the communication mode of digital protection, and most importantly to strengthen the intellectual property protection of intangible cultural heritage digitization [36]. Sun et al. [75–77] put forward new perspectives, proposed the application of relevant theories of public management to

countermeasures to solve the problem of digitization of cultural heritage, and constructed a New Trinity protection and inheritance model of "government, society, and market" of cultural heritage, which has implications for sustainable development in the context of digital cultural heritage protection. A fresh research viewpoint for the digital safeguarding of cultural assets is also provided by this point of view.

6. Discussion

Following the results of the CNKI's literature review, it can be determined that the number of papers dealing with the issue of digital museum rose fast between 2000 and 2021. Despite the fact that the number of papers fell for a little period of time, it quickly grew again. This demonstrates that academics have maintained a significant interest in this topic over the years. Researchers are mostly found at academic institutions, research institutes, and museums, among other places. Universities are the driving force behind academic research in the field of digital museums; research institutes and big museums play a key role in academic research that is facilitated by digital museums. Unfortunately, there is little collaboration across research institutions, and just a few colleges have engaged in collaborative research efforts to far. Intangible cultural heritage, including cultural heritage and traditional villages, is a hot topic in digital museum research, and digital protection of intangible cultural assets, including cultural heritage and traditional villages, is a research frontier. The debate of local situations, in particular, has long been the primary focus of scholars' deliberations. It can be shown that the digital protection of intangible cultural assets is the top priority of the literary research conducted by China's digital museum and that this is supported by the findings of the research.

According to the findings of this study, the emergence of research hotspots in digital museums is primarily influenced by national policies, digital technology, and the requirements of the public. Chinese society and culture are characterized by a top-down social and cultural framework, and government policy supervision and support are believed to be the most important factors in promoting its development. There is no question that information technology has aided the advancement of Chinese digital museum research, and the debate over the practical use of important technologies has emerged as a popular issue among academics. In addition, the advancement of digital technology is an unavoidable consequence of the efforts made to advance the growth of the digital museum into a smart museum. It is worth emphasizing that the post-epidemic age has catapulted the field of digital museum research to a new level of sophistication. It is both a difficulty and an opportunity for the development of China's digital museum that the virus has had such a devastating impact. During this time period, the investigation of the unique path of service platform development and service upgrading has been a popular topic of discussion.

It is fascinating, in terms of the research frontier and trend, to investigate the digital museum from the viewpoint

of the user experience and to employ the design technique in order to better understand the digital museum, which has drawn the attention of a significant number of scholars. Thus, research on Digital Museum in China has evolved from purely technical to people-oriented and service experience focused research, in line with the overall trend of international research on Digital Museum. It also demonstrates that the perspective and methods of research in this field in China are constantly evolving. Moreover, the researchers who conducted the study discovered that the digital preservation of traditional villages has been gaining more and more attention in recent years, which is sufficient evidence that the Chinese government has elevated Rural Revitalization to a major strategic position. It may be predicted that the research fervor in this sector will continue for a lengthy period of time in the foreseeable future. Furthermore, big data has emerged as a new research frontier for the digital museum, demonstrating that big data technology is a significant driving force for the development of the digital museum, particularly in the digital protection of intangible cultural heritage and the construction of smart museums. Finally, in accordance with the findings of the study hotspots, smart museum is an unavoidable trend in the development of the digital museum in the future.

7. Conclusions

The purpose of this study is to give scholars in the area of Chinese digital museum with a quantitative examination of the literature in this subject, in order to get a more in-depth understanding of the growth and evolution of the field of Chinese digital museum during the past 20 years (2000–2021). The findings of the study are mostly congruent with those of other researchers working in the same field of study. The exception is that this study argues that the establishment of a Chinese Digital Museum comprises a research issue with Chinese characteristics for the conservation of Chinese cultural heritage and intangible cultural heritage, and that this is a research topic with Chinese characteristics. In addition, new media technology is a critical component in the construction of a digital museum's infrastructure. The digital museum benefits from not just the utilization of this significant modern technological tool, but also from the high-quality user experience that it gives. It is particularly important in this postepidemic period to optimize and upgrade user-centered information service platforms since this is the only way to ensure the long-term viability of the digital museum. Finally, this study concludes that the ultimate goal of the research on China's digital museum is to find a way to ensure the long-term development of digital museums that is appropriate for China's national conditions, in order to preserve the diversity of Chinese traditional culture while also meeting the growing cultural needs of the general public.

Data Availability

The datasets 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 Youth Fund for Scientific Research Projects of Nanjing Institute of Engineering in 2019, "Research on educational service design of Nanjing Museum from the perspective of cultural experience" (project no. QKJ201903), General Project of Philosophy and Social Sciences in Colleges and Universities of Jiangsu Province in 2020, "Research on experience design innovation strategy of Nanjing Museum driven by new technology" (project no. 2020SJA0445), and Innovation Fund General Project, the authors' project proposed by Nanjing University of Technology in 2021, "Research on the Design and Application of Intelligent Interaction Platform based on 5G Technology" (project no. CKJB202111), Teaching Reform and Construction Project of Nanjing University of Technology in 2021, "Exploration and practice of "Art and Engineering" in product design under the background of new liberal arts" (project no. JXGG2021032), and Nanjing Lanju Information Technology Co., Ltd., Ministry of Education Industry-Science Cooperation Collaborative Education Project in 2021 "Product design and digital molding teaching and research platform practice and base construction" (project no. 202102355023).

References

- [1] Y. Yan and Y. Ryu, "Exploring Google Street View with deep learning for crop type mapping," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 278–296, 2021.
- [2] L. Scrofani, L. Ruggiero, and L. Ruggiero, "Museum networks in the Mediterranean area: real and virtual opportunities," *Journal of Cultural Heritage*, vol. 14, no. 3, pp. S75–S79, 2013.
- [3] M. Milosz, S. Skulimowski, J. Keşik, and J. Montusiewicz, "Virtual and interactive museum of archaeological artefacts from Afrasiyab – an ancient city on the silk road," *J. Digital Applications in Archaeology and Cultural Heritage*, vol. 18, 2020.
- [4] B. Pauget, J. M. Tobelem, and J. P. Bootz, "The future of French museums in 2030," *Technological Forecasting and Social Change*, vol. 162, Article ID 120384, 2021.
- [5] C. Erolin, M. Jarron, and L. J. Csetenyi, "Zoology 3D: creating a digital collection of specimens from the D'arcy thompson zoology museum," *Digital Applications in Archaeology and Cultural Heritage*, vol. 7, pp. 51–55, 2017.
- [6] S. Dong, X. Wang, S. Xu, G. Wu, and H. Yin, "The development and evaluation of Chinese digital science and technology museum," *Journal of Cultural Heritage*, vol. 12, no. 1, pp. 111–115, 2011.
- [7] M.-C. Tang, "Representational practices in digital museums: a case study of the national digital museum project of taiwan," *The International Information & Library Review*, vol. 37, no. 1, pp. 51–60, 2005.
- [8] E. Ch'ng, C. Shengdan, L. Fui-Theng, and Z. Tong Evelyn, "Adoption and use of emerging cultural technologies in China's museums," *Journal of Cultural Heritage*, vol. 37, pp. 170–180, 2019.

- [9] Y. Q. Hu, "On the current situation and development strategy of University Museums," *J. Technology Information*, vol. 4, no. 17, p. 246+251, 2019.
- [10] P. Q. Chen, "Research on the development of personalized service of digital library based on data mining technology," *J. Technology Information*, vol. 19, no. 36, pp. 149-151, 2021.
- [11] W. Q. Li, "Application of big data mining technology in College Archives Management," *J. Journal of Zhengzhou Railway Vocational and Technical College*, vol. 33, no. 04, pp. 106-107+112, 2021.
- [12] C. C. Chen, *A Practical Guide for Mapping Scientific Literature*, Nova Science Publishers, Hauppauge, NY, USA, 2016.
- [13] C. Chen, "Citespace II: detecting and visualizing emerging trends and transient patterns in scientific literature," *Journal of the American Society for Information Science and Technology*, vol. 57, no. 3, pp. 359-377, 2006.
- [14] R. Rousseau, L. Egghe, and R. Guns, *Becoming Metric-Wise: A Bibliometric Guide for Researchers*, Chandos Publishing, Oxford, UK, 2018.
- [15] M. Kabil, S. Priatmoko, R. Magda, and L. D. Dávid, "Blue Economy and coastal tourism: a comprehensive visualization bibliometric analysis," *Sustainability*, vol. 13, no. 7, p. 3650, 2021.
- [16] U. Brandes, "A faster algorithm for betweenness centrality*," *Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163-177, 2001.
- [17] L. S. An, S. Pan, J. Lv et al., "Building a bridge between museum collections" written by experts," *J. Southeast Culture*, no. 3, pp. 11-25, 2014.
- [18] Y. N. Huang, S. Sun, J. Sun, T. Hu, Q. Tang, and H. Zhao, "Research and application of digital protection of sports cultural heritage," *Journal of Sports Science*, vol. 3, pp. 12-16+67, 2007.
- [19] S. H. Zhao and X. F. Zhu, "Construction and implementation of digital humanistic storage," *J. Information Work*, no. 4, pp. 42-47, 2015.
- [20] G. W. Zhong and J. L. Zhang, "A summary of the research on smart museums in China," *J. Science education and Museum*, vol. 6, pp. 347-354, 2020.
- [21] W. R. Le, "Research on digital museum - a guide to the construction of digital museum," *J. Science Education and Museum*, vol. 2, no. 2, pp. 153-154, 2016.
- [22] J. H. Liu and W. P. Zhu, "Research on visualization of development trend of digital museum in China based on CO word clustering," *J. Science Popularization*, vol. 14, no. 3, pp. 36-48+57+111, 2019.
- [23] J. Y. Zhou, "Research on the digital development of Museums," *J. China's National Conditions and Strength*, vol. 6, pp. 16-21, 2021.
- [24] S. S. Gao, "On the digital construction of museum cultural resources," *J. Chinese Character Culture*, vol. 9, p. 97, 2018.
- [25] J. Li, Y. Dong, and C. L. He, "Research on Key Technologies of virtual museum construction," *J. Computer Programming Skills and Maintenance*, vol. 7, pp. 14-15+45, 2018.
- [26] P. Lv, P. F. Yang, and X. Li, "Interactive design of virtual museum based on VR technology," *J. Packaging Engineering*, vol. 38, no. 24, pp. 137-141, 2017.
- [27] S. J. Li, "Integrated development of physical museum and Virtual Museum: smart Museum," *J. Science and education guide (Part I)*, vol. 34, pp. 183-184+187, 2015.
- [28] D. L. Sun, "The development trend of digital museum from the perspective of mass culture communication -- a comment on the new development trend of Digital Museum," *J. News Lovers*, vol. 9, p. 98, 2017.
- [29] Y. W. Song, "Research on digital filing, protection and inheritance of cultural heritage of Gongyi grotto Temple," *J. File Management*, vol. 4, pp. 71-72, 2020.
- [30] F. Ding and J. T. Guo, "Research methods of art history of technology and the protection and restoration of cultural heritage," *J. Science and Technology Bulletin*, vol. 33, no. 4, pp. 249-253, 2017.
- [31] Y. Zhou and S. Yang, "Digitalization makes cultural heritage 'live,'" *J. Culture Monthly*, vol. 5, pp. 12-13, 2021.
- [32] W. L. Ding and Y. B. Wang, "Responsibility and coordination path of archives department in intangible cultural heritage digital archiving," *J. Zhejiang Archives*, vol. 10, pp. 17-20, 2021.
- [33] H. F. Run, "Risks and Countermeasures of digital protection of intangible cultural heritage," *J. Journal of Luoyang Normal University*, vol. 34, no. 4, pp. 69-72, 2015.
- [34] N. X. L. Tu, and Y. Q. Xu, "Development status of intangible cultural heritage digitization," *J. Chinese Science: The Information of the Science*, vol. 49, no. 2, pp. 121-142, 2019.
- [35] S. Yan, "Research on digital protection of intangible cultural heritage in Shanxi Province," *J. Technology and Innovation*, vol. 6, pp. 61-63, 2016.
- [36] Y. H. Lin, "Research on the protection of intangible cultural heritage from the perspective of digital technology -- Taking the traditional spinning, dyeing, weaving and embroidery technology of Li nationality as an example," *J. Research on National Art*, vol. 24, no. 5, pp. 116-121, 2011.
- [37] T. K. Ren, "Protection path, successful experience and localization enlightenment of British sports intangible cultural heritage," *Journal of Sports Science*, vol. 40, no. 3, pp. 67-73, 2019.
- [38] Z. G. Yang, "The role of VR technology in modern science and Technology Museum," *J. Science and Technology Horizon*, vol. 2, p. 18, 2017.
- [39] S. Cai, P. Wang, Y. Yang, and E. Liu, "An overview of the educational application of Augmented Reality Technology," *J. Journal of Distance Education*, vol. 5, pp. 27-40, 2016.
- [40] M. Wang, "Research on the application of AR and VR technology in digital museum," *J. Electro-acoustic Technology*, vol. 1, pp. 46-47+51, 2021.
- [41] Q. Wang, "Problem analysis and Improvement Strategy Research of domestic digital museum under the background of all media," *J. Journal of Nanjing University of Posts and Telecommunications (Social Science Edition)*, vol. 21, no. 2, pp. 83-90, 2019.
- [42] J. P. Wu, "Research on the presentation form of digital museum based on Web," *J. Journal of Changsha civil affairs vocational and Technical College*, vol. 27, no. 4, pp. 132-134, 2020.
- [43] W. Zhang, X. Zheng, and Y. B. Ye, "New exploration of museum service mode in the era of mobile Internet," *J. Science and Technology Bulletin*, vol. 7, p. 53, 2016.
- [44] Y. Hu and P. Liu, "Research on the communication channels of traditional Chinese Medicine Museum in the new media era," *J. Introduction to Traditional Chinese Medicine*, vol. 8, pp. 63-68, 2020.
- [45] N. Luo, S. Xu, X. Lu, H. Hu, I. Zhao, and R. Zhao, "Museum management in the digital age," *J. Chinese Museum*, vol. 2000, no. 1, pp. 17-19+31, 2000.
- [46] S. C. Dong, S. Xu, X. Lu, H. Hu, I. Zhao, and R. Zhao, "GOMS metadata description system based on XML," *J. Earth Information Science*, vol. 5, no. 1, pp. 90-94, 2003.
- [47] S. Sun, M. Zhang, L. Li, and Z. Pan, "Application of digital watermarking technology in Digital Museum," *J. Journal of System Simulation*, vol. 15, no. 3, pp. 347-349, 2003.

- [48] H. Zhang, L. W. Zheng, and Y. Q. Lu, "Research on the application of virtual reality technology in the development of Digital Museum," *J. Educational Technology (Software Guide)*, vol. 9, pp. 49-50, 2010.
- [49] Z. Y. Xing and X. H. Xu, "Research on the value of university digital museum," *J. Technology Information*, no. 6, pp. 3-4, 2013.
- [50] Beijing digital museum Symposium on integration, and Beijing Science Association, "Beijing digital museum Symposium on integration, innovation and development," *J. Forum of Association for Science and Technology*, vol. 31, no. 7, 2013.
- [51] G. Ke and Q. Jiang, "Application of Internet of Things technology in the construction of wisdom museum," *Concurrency and Computation: Practice and Experience*, vol. 31, no. 10, Article ID e4680, 2019.
- [52] O. Hai, "Research on the application of intelligent perception in smart Museum -- Taking visible light communication technology as an example," *J. Heritage and Conservation Research*, vol. 3, no. 08, pp. 108-110, 2018.
- [53] X. Qi, "Feasibility analysis of Nb IOT application in the forbidden city -- exploration and practice of wisdom museum," in *Proceedings of the 2019 Beijing Digital Museum Symposium*, pp. 106-112, Beijing, China, December 2019.
- [54] Y. Wang, "Smart Museum and intangible cultural heritage protection and inheritance -- Taking the "digital" protection and inheritance of Zhihua Temple Beijing music as an example," in *Proceedings of the 2019 Beijing Digital Museum Symposium*, pp. 253-257, Beijing, China, December 2019.
- [55] Y. Y. Zhang, "How to do a good job of collection management in the era of wisdom, culture and Expo," in *Proceedings of the 2019 Beijing Digital Museum Symposium*, pp. 248-252, Beijing, China, December 2019.
- [56] W. W. Li, "On the construction of smart Museum from the perspective of public service," *J. Digital technology and Application*, vol. 37, no. 7, pp. 219-220+222, 2019.
- [57] Z. Du, "The application of new media technology in Museum publicity and Education -- Based on the Internet+ Chinese civilization action plan for three years," *Journal of Human Mass Media Vocational and Technical College*, vol. 17, no. 6, pp. 17-20, 2017.
- [58] M. Zhou, "Construction of virtual exhibition hall of Suzhou intangible cultural heritage Digital Museum -- Taking Su fan as an example," *J. Science and technology horizon*, vol. 2015, no. 3, p. 49+149, 2015.
- [59] J. M. Wang, "Digital museum design in the new media era," *J. Beauty and the times (Part I)*, no. 5, pp. 110-111, 2014.
- [60] J. Li, "Exploring the construction of digital museum from the perspective of interactive intermediary," in *Proceedings of the 2019 Beijing Digital Museum Symposium*, vol. 2019, pp. 183-189, Beijing, China, December 2019.
- [61] D. Y. Qian and Q. Q. Song, "Research on app interactive design of Digital Museum of stone carving art in Mausoleum of Southern Dynasty," *J. Beauty and the times (Part I)*, vol. 6, pp. 77-79, 2018.
- [62] Y. Jing and J. M. Zheng, "Public digital culture service policy from the perspective of local governance," *J. Library Construction*, vol. 2021, no. 4, pp. 106-112+122, 2021.
- [63] A. Zheng, J. Zheng, and X. Wang, "Realistic challenge and promotion direction of digital communication of sports intangible cultural heritage in the post epidemic Era," *J. Journal of Guangzhou Institute of Physical Education*, vol. 1, pp. 57-61, 2011.
- [64] Z. Zhang, *Research on Web Interface Design of Northwest Silk Road Linear Cultural Heritage Digital Museum*, J. Xi'an University of architecture and technology, Xi'an, China, 2020.
- [65] Q. Wang, "Application and exploration of experience design in Digital Museum -- a case study of virtual exhibition hall of National Museum of China," *J. Decoration*, vol. 2, pp. 134-135, 2020.
- [66] M. Fan, "Research on user experience evaluation of digital museum based on coding and decoding theory," *J. Satellite TV and Broadband Multimedia*, no. 2, pp. 103-105, 2020.
- [67] Ministry of Housing and Urban-Rural Development Source, "Notice on the General Office of the Ministry of housing and urban rural development Construction of Excellent Villages in the Digital Museum of Traditional Chinese Villages," 2017, <http://hongbowang.net/news/yj/2017-03-10/6573.html>.
- [68] C. Gao, Z. N. Wang, and J. Zhou, "On the application strategy of three-dimensional digital technology surveying and mapping involved in the protection of traditional village architectural heritage," *J. Journal of Inner Mongolia University of Technology (Nature Science Edition)*, vol. 39, no. 5, pp. 379-385, 2020.
- [69] G. T. Song and Y. Q. Zhao, "Construction of traditional village digital museum under the background of digital protection of traditional villages," *J. Popular science*, no. 6, pp. 16-19, 2020.
- [70] F. J. Cao and Q. Wu, "Information and interaction design in the digitization of cultural heritage -- Taking the Digital Museum of traditional Chinese villages as an example," *Journal of Design*, vol. 34, no. 22, pp. 107-109, 2021.
- [71] S. Bai, "Discussion on the construction and application of modern wisdom Museum System," *J. Intelligent Building Electrical Technology*, vol. 5, pp. 82-84+88, 2021.
- [72] J. P. Shen, "Research on the design and implementation of Museum big data fusion system," *J. Electronic components and information technology*, vol. 5, no. 10, pp. 21-22+25, 2021.
- [73] Q. Yang, "Research on the protection and inheritance of Huizhou intangible cultural heritage in big data environment," *J. Journal of Anhui Agricultural University (Social Science Edition)*, vol. 5, pp. 104-108, 2018.
- [74] D. Zhao, *Research on the protection and Development of Historical and Cultural Resources under Digital Survival*, PhD dissertation, Shandong University, Jinan, China, 2014.
- [75] Q. Sun and P. F. Liu, "Research on the value of cultural heritage digitization in public management," *J. Business News*, vol. 35, pp. 142-143, 2019.
- [76] Q. Dang, Z. Luo, C. Ouyang, L. Wang, and M. Xie, "Intangible cultural heritage in China: a visual analysis of research hot-spots," *Frontiers, and Trends Using CiteSpace, Sustainability*, vol. 13, no. 17, p. 9865, 2021.
- [77] M. Atiquzzaman, N. Yen, and Z. Xu, "Big Data Analytics for Cyber-Physical System in Smart City, BDCPS," in *Proceedings International Conference on Big Data Analytics for Cyber-Physical-Systems*, Springer, Shenyang, China, December 2019.

Research Article

Design of Nuclear Radiation Monitoring System in Floor Exploration Based on Deep Learning

Bochen Zong 

School of Nuclear Science and Engineering, North China Electric Power University, Beijing 102206, China

Correspondence should be addressed to Bochen Zong; 120201110522@ncepu.edu.cn

Received 28 March 2022; Revised 8 April 2022; Accepted 5 May 2022; Published 21 May 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Bochen Zong. 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.

Nuclear radiation environmental monitoring has become an important issue in floor surveys. From the perspective of regional environmental nuclear radiation monitoring, it is of great practical significance to establish a scientific and reliable wireless sensor monitoring network for timely and accurately grasping nuclear radiation status and ensuring nuclear safety. In this article, we design a regional environmental nuclear radiation monitoring system based on Zigbee wireless sensor network by using Zigbee wireless technology. First, the network topology suitable for nuclear radiation environment monitoring is designed. Second, the JN5121 module is selected as the core of the Zigbee wireless sensor to build the network hardware platform. Finally, the article focuses on the receiving mechanism, data transmission, coordination between network nodes, network structure, and dynamic network management, and carries on the software development to the Zigbee wireless environment monitoring network. The system can measure the data in real time, display the dose rate of γ radiation effectively, and realize the functions of remote control, field control, security alarm, and environmental monitoring. It has good promotion and application value.

1. Introduction

In recent years, the frequent occurrence of the urban environment and land gas events makes the study of the release law of urban environment and land gas become the focus. For densely populated cities, artificial changes in the particularity of the surface structure will affect and change the natural release of Earth and air materials in the soil. It is easy to release in weak coverage and often causes anomalies. Zhou and Tao of Peking University put forward the concept of macro and micro effects of geog as based on the release characteristics of geog and its impact on the ecological environment and studied the phenomenon of long-distance transport of matter under the action of geog as and its significance in environmental research [1]. Du believes that upward emission of underground gases can cause at least 13 effects, revealing the environmental hazards of underground gases [2].

The floor survey and monitoring system needs to collect a lot of data, and it has a very high demand for wireless

sensor networks [3]. Wireless sensor networks should not only track and monitor environmental changes but also not destroy the environment. At present, deep learning has become an innovative technology to solve the problems of floor exploration. In particular, Zigbee wireless sensor network is a research hotspot of scholars at home and abroad. Zigbee wireless sensor environmental monitoring system has a broad application prospect [4]. The research emphases include node device miniaturization technology and real-time operating system design based on the embedded device, protocol, and architecture construction of sensor network [5, 6].

Based on the above research, this topic from the floor to the exploration of the radiation monitoring perspective, using Zigbee wireless technology to design based on Zigbee wireless sensor network floor exploration of the radiation monitoring system, real-time measuring data, shows that the radiation dose rate effectively realizes the remote control, scene control, security alarm and environmental monitoring, and other functions. The system is of great significance

to timely and accurately grasp the situation of nuclear radiation, nuclear accident emergency protection, ensure nuclear safety, and reduce the incidence of nuclear accidents.

2. System Requirement Analysis

To meet the needs of wireless ad hoc network floor survey and radiation environmental monitoring requirements, the first task is to determine the depth study adopts the wireless communication technology, according to the functional requirements of the whole system mainly from the wireless ad hoc network, collection and processing of various sensors, positioning and alarm functions, and combining with the characteristics of the regional environment, and determine various modules for the system. A wireless ad hoc network monitoring system is designed for floor investigation of nuclear radiation.

2.1. Demand Analysis of Key Technologies. At present, the network communication technologies used in wireless sensor networks mainly include wireless local area networks (Wi-Fi), ultra-wideband communication (UWB), Bluetooth, and Zigbee.

Wi-Fi has a long transmission distance and high speed, but it costs a lot and consumes a lot of energy. Wi-Fi technology relies on wireless network cards or access point devices applied to campus and corporate LANs and is suitable for indoor monitoring, not for outdoor monitoring needs [7].

Bluetooth has the advantages of low cost, low system power consumption, and small chip size. However, the biggest disadvantages of Bluetooth technology are high engineering cost, short communication distance (10 m), and easy eavesdropping of data communication, which puts Bluetooth technology at a disadvantage in the detection of the nuclear radiation environment.

UWB is characterized by fast transmission speed, strong anti-interference ability, very small transmitting power, and low electromagnetic radiation to the human body. That is because of these advantages that some people believe that it will become the mainstream technology of short-distance wireless communication in the future [8]. But the current working group on UWB standards has failed to come up with a final standard, and the technology still needs to be refined because there is no consensus.

Table 1 shows the existing mainstream wireless sensor network technologies.

As can be seen from the comparison in Table 1, Zigbee is selected mainly from the perspective of low power consumption (compared with Bluetooth and WiFi), communication distance, and connected devices. WiFi consumes a lot of power, and Bluetooth supports up to 8 connected devices in theory. The ability to connect devices is poor. The key is that the device does not have the ability of ad hoc networking, so it must rely on the external network. Zigbee is a bidirectional wireless communication technology based on IEEE 802.15.4 standard (2.4 GHz), low power LAN protocol, and ad hoc low-cost wireless network. It supports up to 65000 devices and has a valuable feature of high-

security performance and adopts the AES-128 encryption algorithm [9]. The uses of Zigbee are not as widely promoted as WiFi and Bluetooth, but Zigbee has advantages of low cost, low power consumption (but only for terminal nodes), low complexity, flexible networking (advantages when there are more devices in the network), strong self-recovery ability, small delay, high reliability, and high-security performance.

2.2. Zigbee Protocol

2.2.1. Zigbee Networking Mode. The Zigbee protocol specifies three device types for nodes based on their roles in the network: coordinator, router, and terminal. Zigbee technology has strong networking capability and can form three typical self-organizing wireless network types, namely star structure, network structure, and tree structure [10]. As shown in Figure 1:

- (1) Star network consists of a coordinator (as the central node) and several other terminal nodes, without a routing algorithm. Each node must be within the communication range of the coordinator. The structure has fewer nodes, a simple structure, and low cost. A large amount of data flooding into the coordinator, resulting in network congestion and poor stability. Usually suitable for the network capacity is not high requirements of the environment, such as home automation and other small areas.
- (2) Tree network is an extension of a star network, adding router nodes. In the tree structure, each router can also have its child nodes. Long-distance transmission can be realized through multilevel hop of router nodes. Terminals cannot connect to their child nodes. Once a routing node in the transmission channel fails, the other nodes of the communication link are disconnected from the network. According to the tree network of the parent-child relationship, the data transmission mode is simple and the transmission path is increased. However, as the scale of the network increases, the transmission delay will become larger and the stability of the network will become worse, which is suitable for the occasion where the timing requirement is not high.
- (3) Any two router nodes of the mesh network can communicate with each other, and the optimal path selection of the network is completed by the routing table mechanism [11]. It provides multiple routing paths for data packets. If a routing path is damaged (e.g., a module is powered off), the network can choose a new routing path to repair the routing path freely, avoiding data flow collision and blockage, and improving data transmission reliability. Mesh network is suitable for a large number of nodes, so the relative overhead and power consumption is large, suitable for occasions requiring high timeliness and reliability, and is also the most Zigbee network application topology.

TABLE 1: Comparison of mainstream wireless sensor networks.

Designation	Zigbee technology	Bluetooth technology	Wi-Fi technology	UWB technology
Cost of the chip	\$4/group	\$5/group	\$20/group	\$25/group
Battery life	6–24 months	Several days	Several hours	Several hours
Transmission distance	10–100 m	<10 m	100–300 m	<10 m
Transmission rate	20–250 kbps	1–3 Mbps	11–54 Mbps	1 Gps
Frequency band	868 MHz–2.4 GHz	2.4 GHz	2.4 GHz	3.1–10.6 GHz
Maximum number of network nodes	65535	8	30	100

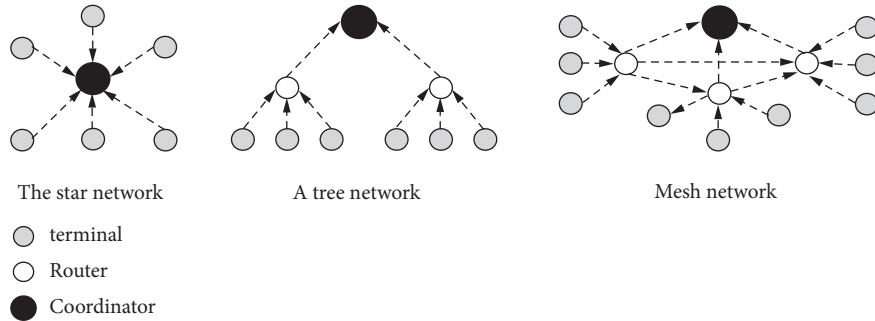


FIGURE 1: Three network topologies.

Each network topology has its advantages and disadvantages. Considering that the monitoring system has many nodes and large scale, to improve the reliability of data transmission, the network topology is preferred in this system. The mesh network structure does not need to consider the shape of the network when establishing the network and has strong adaptability to the external environment.

Compared with Zigbee 2007, Zigbee 2007 Pro provides greater network support and expands the number of networks. Expanded from 31101 to 65540 [12], it is more suitable for commercial applications and has many enhanced functions, such as alternate addressing, many-to-one routing, and higher security performance [13]. Zigbee 2007 Pro adopts a random address assignment scheme to assign random addresses to newly added devices [14].

Zigbee protocol is different from other network architecture systems, which are divided into 7 layers. That is mainly based on IEEE802.15.4 protocol specification and Zigbee protocol specification. Data transmission specification of NWK layer, APL layer, and APS layer adopts Zigbee protocol specification. The data transmission specification of the physical layer (PHY layer) and Media Access Control layer (MAC layer) adopts IEEE802.15.4 protocol specification, which is Zigbee wireless network [15].

2.2.2. Operation Mechanism of Zstack. To complete data communication between wireless nodes, the whole protocol stack needs to run. The Zstack adds a layer of real-time manipulation called OSAL [16], which operates on top of the OSAL, and the OSAL, a priority-based rotating system, allows for task switching and memory management [17]. The whole system starts from the main function in the ZMain. ZMain.c file in the directory.

The main function is used to initialize the system and perform the multitask polling and query function of the operating system.

(1) *Initialization of the system.* The `osal_init_system()` function is called to complete the initialization of each module, mainly divided into the initialization of the system clock, initialization stack, initialization of each hardware module, such as I/O, LED, etc., initialization of Flash memory, determine the IEEE64-bit address, initialization of nonvolatile variables, initialization of the operating system, etc. Until `osal_start_system()`, the function call is the actual operation of the stack.

(2) *The execution of the operating system.* After a series of preparations are made for the operation of the operating system after system initialization, the operating system is started. The `osal_start_system()` function is an infinite loop [18], which continuously detects the occurrence of events and performs corresponding function operations once detected. The operating system needs to do two things: one is to create a task event table. The operating system sets the `taskEvents[]` array to hold all the data and adds the event to the task event table. Each item in the array is a pointer to the event handler of each task event table. After processing, it continues to access the event table to see if any events have occurred. `TasksArr []` is set to store the address of each task event handler. When an event occurs, you can find the corresponding task event handler in the function table. Call `events = tasksEvents[idx]` to retrieve the events to be processed from the task event list. This array contains the sequence number from 0 to `tasksCnt`. `Idx` is the sequence number of the priority task (the smaller the value, the higher the priority task). `TasksArr [idx]` can then be called to perform specific task event handlers depending on the `idx` in the `events = (tasksArr[idx]) (idx, events)` statement. Figure 2 below is a diagram of the task event table and the event handler function table.


```

int main(void)
{.....
osal_init_system(); //Initialize the operating system.
osal_start_system(); //After system initialization, the system starts to execute the operating system
.....}

```

ALGORITHM 1: ZMain. C file code in directory.

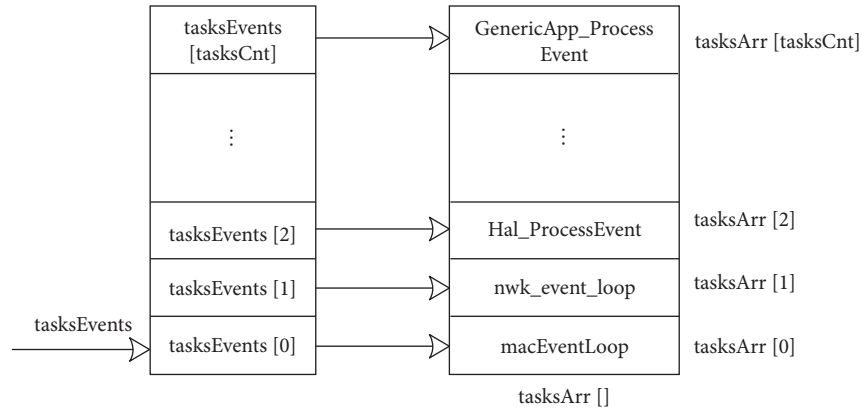


FIGURE 2: The relationship between the task event table and the task event handler table.

2.3. Feasibility Requirement Analysis of the System

2.3.1. Analysis of Overall Functional Requirements of the System. Based on the sensor node, router, coordinator, and monitoring center, the ground radiation environment monitoring system is constructed into Zigbee wireless sensor network. Due to the complex geographical location of the environmental area, there are many obstacles and the location of nodes will change, so this design chooses a mesh network structure with multiple data transmission paths to bypass the obstacles by relying on routing relay and migration back to ensure the reliability of transmission and is conducive to centralized supervision distributed in multiple areas. According to each requirement of the monitoring environment of the system, a corresponding sensor node is set. Figure 3 shows the overall network architecture diagram of the system.

Through the analysis of the overall system workflow, the functional requirements of each part are analyzed as follows:

- (1) Sensor node function analysis: sensor node is also a terminal node, the node can be configured with one or more sensors, mainly responsible for the acquisition of γ radiation dose rate, temperature, humidity, illumination, CO, rainfall, and other environmental parameters, latitude, and longitude coordinates and real-time time information.
- (2) Router function analysis: mainly responsible for forwarding sensor data collected by nodes in different areas to the coordinator. Relying on the relay function of router uploading and sending, the multi-

hop route migration mode is selected to store the collected data and upload it to the coordinator, which makes the whole Zigbee network more extendable and increases the coverage of the network.

- (3) Function analysis of coordinator: mainly responsible for networking, receiving data, alarm processing, uploading data to an upper computer, and forwarding upper computer commands. Sensor nodes and routers aggregate sensor data to the coordinator. If the data received by the coordinator exceeds the set threshold concentration, the alarm module will be triggered to alarm, and it will be transmitted to the upper computer software of the monitoring center for processing through serial port or Internet, mobile communication network, or system private network.
- (4) Functional analysis of monitoring center: mainly responsible for displaying environmental data in intuitive decimal numbers and graphs through upper computer software. By reading the encapsulated data directly from the serial port of the coordinator, the environment data are analyzed, processed, and stored according to the communication protocol, and various control operations are carried out according to the analysis results.

2.3.2. System Module Requirement Analysis. Because the monitoring environment is complex and there are multiple interference sources, the following modules are required based on system requirements:

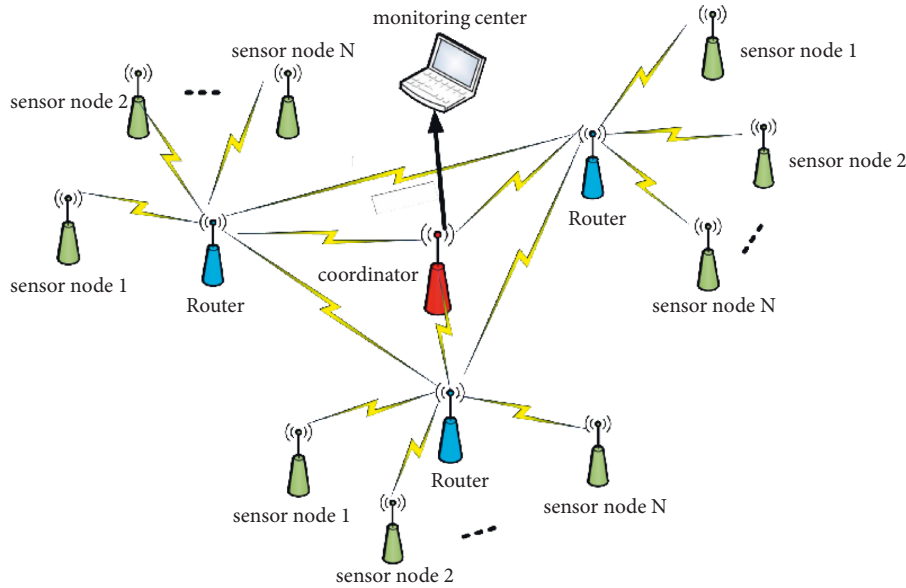


FIGURE 3: Overall network topology of the system.

- (1) Dust sensor: used to detect dust concentration in the environment.
- (2) Alarm module: used for alarm greater than the danger threshold value; there are two ways to alarm: one is the buzzer and LED sound and light alarm, and the other is GSM SMS remote alarm.
- (3) GPS positioning module: used to obtain the location and monitoring time of nodes in the environment area. If the monitored data lack geographical location information, the monitoring system will be meaningless. Nodes must have the ability of real-time and geographical location.

3. Design of Nuclear Radiation Monitoring System in Floor Survey

Using sensors (photoelectric sensors information collection devices, design a set of the regional environmental radiation monitoring system, collect the regional environmental radiation doses of radiation, after data encoding, using wireless transmission technology to the control center, control center software after receiving information analysis processing, draw the corresponding curve, As a prerequisite for the controller's next move.

Zigbee network and monitoring host together constitute the regional environmental nuclear radiation monitoring system. The whole system is a hierarchical network structure, with Zigbee sensor nodes at the bottom, Zigbee coordinator and Zigbee router at the middle, and monitoring host at the top [19].

As the high-end equipment of the whole monitoring system, the monitoring host not only needs to monitor the radiation dose data of environmental radiation but also needs to master the health and working status of the sensor node itself [20]. The host not only displays the collected data, but also stores the address of the data source to judge the

good or bad situation of the node itself, and timely adjusts the tasks assigned to the node according to the change of the data, as well as prejudices the life cycle of the sensor node. The sensor node itself includes the working conditions of communication components, sensors, residual energy, etc. The remaining energy information of the wireless sensor node determines the working voltage of the node. When the voltage value is low, the reliability of the data transmitted by the sensor node decreases, and the voltage alarm of the monitoring center is triggered. At this time, the battery of the node should be replaced.

3.1. Design of Sensor Node. The sensor network node structure of the regional environmental nuclear radiation monitoring system is shown in Figure 4.

The structure is composed of an energy unit, Zigbee terminal equipment, Zigbee processing unit, positioning system, and Zigbee communication unit[21]. The Zigbee processing unit is an embedded system with memory, CPU, and operating system. The communication unit includes a serial port direct communication interface and wireless communication module.

3.2. System Network Model. When the monitoring environment is not convenient for the host to use on-site for a long time, a base station can be set up close to the monitoring environment. The base station acts as a gateway between Zigbee wireless sensor network and the wired network. Zigbee sensor network first wireless signal to the base station, and the base station through the wired connection will be transmitted to the monitoring host information management module. At the same time, the information management module transmits query and monitoring commands to the base station, which transmits them to the nodes in the Zigbee sensor network.

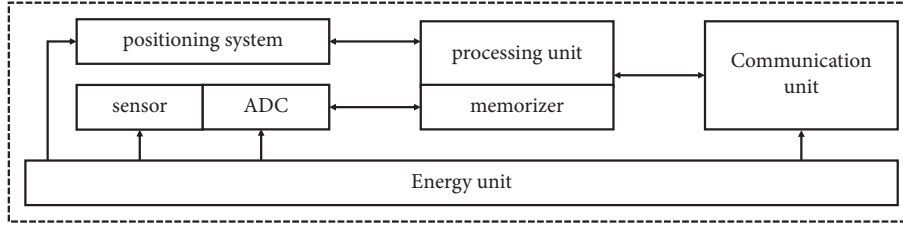


FIGURE 4: The structure of network node in the system.

3.3. *Distribution of Sensor Nodes.* Nodes of the regional environmental nuclear radiation monitoring system based on the Zigbee wireless sensor network adopt a hierarchical structure. The whole network is composed of multiple star networks, which we call clusters. Each star network consists of a central node and multiple terminal nodes. The terminal node is responsible for data collection, and the central node is responsible for data forwarding between nodes. Therefore, the central node of the network (Zigbee router) needs to control the information. The central node can be automatically generated by a clustering algorithm or specified by the monitoring system.

3.4. *Selection of Evaluation Indexes for Radionuclide Dose.* Radiation in the contaminated area formed after regional environmental nuclear radiation is mainly composed of X-rays and rays[22]. The contamination zone itself acts as a radioactive source. Radionuclides (X-rays, rays, etc.) are evaluated by the following criteria:

3.4.1. *Activity.* Radioactivity is the number of spontaneous decays of radionuclides per unit time, expressed in [Bq]. The definition is as follows:

$$A = \frac{dN}{dt}, \quad (1)$$

3.4.2. *Absorbed Dose.* The amount of radiation absorbed by a unit mass of tissue or organ is the absorbed dose, which is measured in Gy. One unit of absorbed dose is equivalent to 1 joule of energy absorbed by a unit mass of tissue or organ. The definition is as follows:

$$D = \frac{d\bar{\epsilon}}{dm}, \quad (2)$$

where D represents absorbed dose, expressed in grays, $1 \text{ Gy} = 1 \text{ J/kg}$; $\bar{\epsilon}$ is the average amount of material absorbed by a tissue or organ per unit mass and is an expected value; and m is the mass.

3.4.3. *Absorbed Dose Rate.* The absorbed dose rate is the amount of radiation absorbed per unit mass of tissue or organ over unit time. The definition is as follows:

$$\dot{D} = \frac{dD}{dt}, \quad (3)$$

where \dot{D} is the absorbed dose rate in Gy/s.

It is difficult to calculate the absorbed dose and absorbed dose rate of radionuclide in the nuclear radiation areas. The absorbed dose of radio nuclear elements in dry air is usually not calculated directly.

3.4.4. *Dose Equivalent.* Because the total amount of radionuclide absorbed by different people or different organisms is different, the same absorbed dose may not produce the same biological effect. Biological effects are therefore influenced by a variety of factors, such as exposure conditions, types and energies of radiation, individual differences, and dose and dose rate. Considering the above factors, the evaluation index of dose equivalent is introduced. The definition is as follows:

$$H = D \times Q \times N, \quad (4)$$

where H is dose equivalent in Sv, $1 \text{ Sv} = 1 \text{ J/kg}$; D is absorbed dose; and Q is the quality factor. The quality coefficient of X or ν rays is usually a constant 1, that is, $Q = 1$. N represents the product of all other correction factors, usually $N = 1$. So, for X or γ ray, the focus of this study, the dose equivalent is equivalent to the absorbed dose.

3.4.5. *Dose Equivalent Rate.* The dose equivalent of a radionuclide element per unit time is the dose equivalent rate. The definition is as follows:

$$\dot{H} = \frac{dH}{dt}, \quad (5)$$

where H is the dose equivalent rate, expressed in Sv/H.

It can be seen that dose equivalent and dose equivalent rate more accurately reflect the damage degree of various rays to tissues or organs, while absorbed dose and absorbed dose rate is difficult to be calculated directly and accurately.

3.5. *Sensors Used by the System.* The sensor data acquisition circuit ensures the working performance of the whole system and is mainly responsible for the data acquisition of the entire nuclear radiation environment. Different environmental data require different sensors, and different sensors have different requirements for communication interfaces, so the form of sensor data acquisition circuit is also different [23]. In the selection of sensors, it is necessary to consider both the needs of the nuclear radiation monitoring environment and the interface between sensors and microcontrollers.

Currently, sensors commonly used in the field of environmental monitoring include biosensors, gas sensors, liquid level sensors, thermal sensors, photoelectric sensors, noise sensors, temperature and humidity sensors, etc. According to the different detection indexes, the selected sensor is also different. Considering that the indicators in the field of water quality monitoring include ammonia content, dissolved oxygen, pH value, biochemical oxygen demand, REDOX potential, electricity, etc., optical fiber oxygen sensor and conductivity sensor are usually selected. Given the large number of parameters to be monitored in the field of atmospheric monitoring, there are many types of sensors applicable, including dust particle sensors, sulfur oxide sensors, airborne-burning sensors, etc.

The sensitivity of photocell can be improved by using g photomultiplier tubes because of the high requirement of light particle sensors. The basic working process of the photomultiplier tube is as follows: the photoelectron escapes from the cathode and passes through the cluster electrode for secondary amplification, so the amount of electron escapes from the photocell increases greatly. The Zigbee wireless sensor terminal node designed in this topic selects the abovementioned photoelectric sensor, model FVDK 10N5101, which is used to collect radiation dose in the contaminated area.

3.6. Wireless Transmission Technology Route and Analysis. With the development of integration technology, microelectronics technology, and wireless communication technology; the continuous popularization of cheap wireless modules; and the continuous reduction of the installation cost of wireless networking technology, wireless communication module design technology is becoming more and more mature. Mature wireless communication technologies include Bluetooth wireless fidelity, 802.11 b, IrDA, Zigbee, HomeRF, etc.

3.6.1. Track of Bluetooth. Bluetooth technology supports long communication distances and can form a small wireless LAN. A Bluetooth device can establish up to seven simultaneous connections, constantly announce its presence to surrounding devices, and is password protected. The 2.45 GHz channel in the Industrial Scientific Medical (ISM) band is a working channel of the Bluetooth protocol. The transmission distance is 10 m and the transmission rate is 1 Mb/s.

Bluetooth technology can be used in the following areas: as a wireless connection between peripheral devices; realization of data or voice message, information, n , and wan connection; and set up personal wireless LAN to realize personal network sharing. However, the chip used in Bluetooth technology is too expensive and has some problems such as information security, short transmission distance, and weak anti-interference ability.

3.6.2. IrDA. The Infrared Data Association (IrDA) is a point-to-point communication technology developed by the

Infrared Data Standards Association. The initial transmission rate of 4 Mb/s has reached 16 Mb/s, the initial reception angle was 30°, and now it has reached 120°.

3.6.3. Wireless Fidelity Technology. Wireless Fidel (Wi-Fi) is also a short-range wireless communication technology commonly used in homes and offices. The 2.4 GHz ISM band is also the working band of Wi-Fi. The 802.11 b physical layer defines another data transmission mode in the 2.4 GHz ISM band, and the transmission rate can reach 11 Mbit/s. Although the Wi-Fi technology has the advantages of low threshold, convenient network layout cost, and wide coverage of radio waves, 802.11a and 802.11 b are incompatible with each other, which brings adverse effects on the market promotion.

3.6.4. Zigbee. Zigbee technology is an emerging wireless network communication technology with low power consumption, short distance, and flow rate, the transmission distance can reach 100 meters, and the transmission rate range is 20 Kb/s-250 KB/s. It realizes low speed, short distance, and low cost and is suitable for the regional environmental nuclear radiation monitoring systems.

In the regional environmental nuclear radiation monitoring system, the area contaminated by nuclear radiation is generally large, so the requirements for wireless sensor network are as follows: First, considering that the monitoring of environmental nuclear pollution takes a long time, nodes must save power to ensure low energy consumption. Second, considering that the system does not require high data transmission rate, the cost should be reduced as much as possible after ensuring the required transmission rate. Third, considering the large monitoring range, the number of sensor nodes required is large. It can be seen from the above requirements that the design of a regional environmental nuclear radiation monitoring system based on Zigbee technology will be the best choice.

4. System Performance Test and Data Analysis

This chapter mainly uses hardware and software platforms to test the overall function of the wireless network floor survey nuclear radiation environment monitoring system and verify the operation effect of the system design scheme. The main work of the test includes testing the communication of various sensor modules, GPS positioning module, GSM SMS module, and radio-frequency module, and finally realizing data acquisition and transmission.

4.1. Node Networking Test. First, five monitoring nodes and one coordinator are prepared to build a wireless ad hoc network, and the topology of the entire Zigbee network ad hoc network is displayed through the Z-Sensor Monitor host computer software of TI Company, as shown in Figure 5:

As can be seen from Figure 5, a network with a mesh topology is successfully formed between nodes. In the figure, two routers (blue) and a sensor node (yellow) are attached to

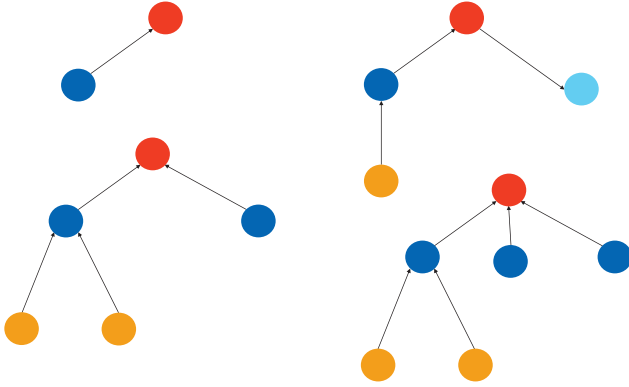


FIGURE 5: Network topology.

TABLE 2: The DC parameters of Zigbee-based wireless sensor network system.

DC character	
Deep sleep current	$<1 \mu\text{A}$
Sleeping current	$<7 \mu\text{A}$
Wireless transmission current	50 mA
Wireless receiving current	60 mA

a coordinator (red), among which two sensor nodes are attached to a router. Sensor nodes and routers will calculate the best path to join the network by themselves, so the network structure is flexible and connected everywhere.

4.2. Implementation of the Nuclear Radiation Monitoring System

4.2.1. Monitoring System Parameter Setting. Parameters of the nuclear radiation monitoring system based on the Zigbee wireless sensor designed in this article are shown in Tables 2 and 3, whose parameters are set by the JN5121 chip.

4.2.2. Working Mode of Wireless Sensor Network. Zigbee wireless sensor network has two working modes as follows: one supports beacon mode, and the second is the beacon-free mode. In mode 1, after the sensor node is powered on and works normally, the beacon of the sensor network coordinator is first monitored and registered. The coordinator can also set the device's wake mode or sleep mode. Once the data are sent and received, the Zigbee coordinator goes into sleep mode, reducing the energy consumption of the entire sensor network. In mode 2, the Zigbee sensor node is equivalent to an autonomous device that can initiate conversations autonomously. The node only wakes up from sleep mode when an event is triggered. The Zigbee coordinator does not enter sleep mode and needs to continuously listen to external information, so energy consumption is higher than in mode 2.

Considering the convenience of system implementation, the system developed in this article adopts a beacon-free mode to realize data transmission and reception in interrupt mode.

TABLE 3: The RF parameters of Zigbee-based wireless sensor network system.

RF character	
Receive sensitivity	-90 dBm
Sending power	0 dBm
Coverage area	80--100 m
Maximum input signal	-10 dBm

4.2.3. Network Configuration for Wireless Sensors. The Zigbee Coordinator sets the wireless sensor network configuration parameters in the configuration file. The Zigbee Alliance defines standards for configuration files, including how nodes in a network are described and how interfaces to specifications are defined for specific applications.

Considering the convenience of system operation, the system developed in this study configures the network directly in the application program and does not use Zigbee standard configuration file to configure the network parameters. The `JZA--boAPPStart ()` function is used to configure network parameters directly.

4.2.4. Establishment of Wireless Sensor Network. The Zigbee Coordinator is responsible for establishing the entire Zigbee wireless sensor network. The network implementation process is as follows:

First, at the beginning of power-on, the Zigbee Coordinator searches the network for an appropriate transmission channel, which should be the one that is least used or has not been used. Second, a sensor network is established by using the channels selected by the Zigbee coordinator. To facilitate communication identification, a PAN ID is defined for the network. Moreover, once the wireless sensor network is established, Zigbee routers (central nodes) and terminal nodes can be added to the network. After these two types of nodes are added, the Zigbee coordinator assigns appropriate network addresses to them. After the above steps, the nuclear radiation monitoring system based on Zigbee wireless sensor network can be realized.

4.3. Commissioning of the Nuclear Radiation Monitoring System. In the process of product development, software and hardware debugging is a key link. In the process of software-hardware debugging, it is possible to simulate various situations that may occur in the field operation, to find and solve problems, and finally improve the performance of the system and achieve the functions that the monitoring system should be able to achieve.

4.3.1. Debug the Hardware Circuit of the System. During debugging, it was found that the program could run on the computer, but not on JN5121. It is considered that there may be a fault in the program download interface circuit. Through continuous checking and debugging, it is finally determined that the program download interface circuit on JN5121 has a pin short connection problem, resulting in the program download failure. Once the short connection problem is resolved, the program can run normally.

The communication between Zigbee wireless sensor network nodes and monitoring hosts was debugged and found that the communication could not be realized. After troubleshooting, the crux of the problem that is rS-485 communication interface circuit failure is found. Since the normal working voltage of the RS-485 communication chip used in this system is +5V, the working voltage provided by the power supply is +3.3V in the debugging process, resulting in the normal operation of the RS-485 communication interface circuit. After the problem is solved, the RS-485 communication interface circuit runs normally, and the data receiving and receiving between the sensor node and the monitoring host are normal.

Finally, in the operation process of the nuclear radiation monitoring system, the output signal of the sensor node circuit is not ideal, which is greatly affected by external interference, and the output waveform fluctuates greatly and is unstable. After extensive reference, the following conclusions are drawn: the analog circuit is highly sensitive to external disturbance. Therefore, in the combination of the two circuits of hardware, the use of digital and analog ground should not be confused, once confused, will cause high-frequency interference. Usually, there are several ways to separate digital ground and analog ground: for example, to use ohm resistance, inductance, or magnetic beads for separation. This system uses the former separation method, selects 0Ω resistance to separate digital ground and analog ground, and adopts inductance to separate digital power and analog power, so that the reliability and anti-interference characteristics of the whole hardware circuit are enhanced, and the output signal waveform becomes stable.

4.3.2. System Software Debugging. When debugging the software program of a nuclear radiation monitoring system based on Zigbee wireless sensor, there is no problem with the basic framework, but there are some programming language problems, which can be solved quickly. For example, in the process of programming the monitoring host data display program, it is found that the display menu has the phenomenon of operation menu dislocation, and the above problem is solved by unifying the signs at the same level.

4.4. Performance Test of the Nuclear Radiation Monitoring System. For the built nuclear radiation monitoring system, the network implementation test, data transceiver test, Zigbee coordinator and Zigbee router function test, and stability test were carried out. This study realizes the interface of serial communication in a visual basic environment by writing a serial communication program. One by one, the data receiving situation of the above three nodes are tested. By sending a data ton to check the receiving situation of the corresponding nodes, it is found that the receiving situation of the above three nodes is normal without data loss or garbled characters.

Stability tests are performed in a laboratory environment. The laboratory environment is as follows: the lowest temperature is about 10 °C, the highest temperature is about

TABLE 4: Background ambient dose rate measurements.

Background surrounding environment	Maximum value	Minimum value	Average	Relative error (%)
1	68.80	67.06	67.52	1.41
2	69.21	68.00	68.53	1.22
3	74.10	71.15	72.23	1.65
4	72.25	68.88	70.06	0.95
5	83.91	80.54	82.12	2.45
6	76.50	74.41	75.72	7.37
7	71.86	68.60	69.78	5.86
8	81.12	78.50	80.00	4.12
9	79.30	75.19	76.55	1.96

TABLE 5: ¹³⁷Cs source environmental dose rate measurements.

Background surrounding environment	Maximum value	Minimum value	Average	Relative error (%)
1	980.13	972.10	978.15	0.32
2	975.28	970.55	972.62	0.76
3	977.61	980.80	973.78	1.32
4	979.56	973.41	975.88	0.87
5	984.91	981.10	983.76	0.21
6	978.52	974.33	976.32	1.43
7	988.78	983.30	985.67	0.64
8	985.21	977.78	979.72	0.82
9	986.57	980.00	982.78	0.75

25 °C, and the air humidity is about 50%. The nuclear radiation environment monitoring system is placed in the above environment, allowed to continuously for 12 hours, and the stability of the system operation is observed for an emergency. The experimental results show that the regional environmental nuclear radiation monitoring system designed in this study works well and has high stability.

The instrument was placed in the radiation measurement point for field real-time monitoring, and the ambient background and the surrounding environment with ¹³⁷Cs activity were measured respectively. The measurement data are shown in Tables 4, and 5.

By comparing Tables 4 and 5, we can see that: (1) the relative error of environmental dose rate measurement near ¹³⁷Cs source is relatively low; (2) the relative error of dose rate measurement in background environment is high; and (3) through multiple measurements, it is found that the measured value (maximum, minimum, and average values) and relative error are within the fault tolerance range, and the detection accuracy is high, which can be put into practical application.

5. Conclusion

Nuclear radiation environmental monitoring has become an important problem of floor exploration. Zigbee wireless sensor network based on deep learning is a powerful tool to realize nuclear radiation environmental monitoring systems. This study adopts Zigbee wireless sensor with JN5121 module as the core to establish a regional environmental

nuclear radiation monitoring system and draws the following conclusions:

- (1) Based on network structure analysis, the sensor terminal node and the central node are arranged manually or in other ways to establish Zigbee wireless sensor network.
- (2) Zigbee wireless sensor network adopts the JN5121-Z01-M01 chip. The transmission rate of sensor nodes can reach more than 100 Mb/s, and the transmission distance can reach more than 300 m. The data transmission mode is the multi-jump mode. The data collected by the terminal node are transmitted to the central node through this mode, and the central node will transmit the data to the monitoring center after fusion and optimization.
- (3) CRC Check mechanism and AES-128-bit encryption algorithm are adopted. The safety and reliability of data transmission in the regional environmental nuclear radiation monitoring system is ensured.
- (4) The regional environmental nuclear radiation monitoring system can not only realize real-time data display but also realize database storage, which can provide historical data for further data analysis.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] D. N. Wategaonkar, S. V. Nagaraj, and T. R. Reshmi, "Multi-hop Energy-Efficient reliable Cluster-based sectoring scheme using Markov chain model to improve QOS parameters in a WSN," *Wireless Personal Communications*, vol. 10, no. 12, pp. 1–29, 2021.
- [2] W. Sun and L. M. Wei, "Design of intelligent home control system based on ZigBee wireless network," *North Architecture*, vol. 4, no. 4, pp. 35–38, 2019.
- [3] D. Y. Tian, "Experimental teaching exploration of Z-STACK operation mechanism," *Electronic Production*, vol. 20, pp. 44–45+53, 2020.
- [4] X. Guo, Y. He, X. Zheng, L. Yu, and O. Gnawali, "ZigFi: Harnessing channel state information for cross-technology communication," *IEEE/ACM Transactions on Networking*, vol. 28, no. 1, pp. 301–311, 2020.
- [5] M. Collotta, G. Pau, T. Talty, and O. K. Tonguz, "Bluetooth 5: A concrete step forward toward the IOT," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 125–131, 2018.
- [6] A. Hernandez-Solana, D. Perez-Diaz-De-Cerio, M. Garcia-Lozano, A. V. Bardaji, and J.-L. Valenzuela, "Bluetooth mesh analysis, issues, and challenges," *IEEE Access*, vol. 8, pp. 53784–53800, 2020.
- [7] D. Gao, S. Zhang, F. Zhang, T. He, and J. Zhang, "RowBee: A routing protocol based on cross-technology communication for energy-harvesting wireless sensor networks," *IEEE Access*, vol. 7, pp. 40663–40673, 2019.
- [8] T. T. Nguyen and H. Oh, "SCSMA: A smart CSMA/CA using blind learning for wireless sensor networks," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 12, pp. 10981–10988, 2020.
- [9] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, 2018.
- [10] T. H. Yang and F. Diao, "Design of subway environmental safety early warning system based on wireless sensor network," *Instrument Technology and sensor*, vol. 8, pp. 105–107+112, 2018.
- [11] S. X. Yao and Y. N. Xie, "Design of timing terminal for WiFi interface chip ESP8266," *Microcontroller and embedded system application*, vol. 19, no. 10, pp. 63–66, 2019.
- [12] J. Zhang, S. Rajendran, Z. Sun, R. Woods, and L. Hanzo, "Physical layer security for the Internet of things: Authentication and key Generation," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 92–98, 2019.
- [13] S. Wang, Z. Yin, Z. Li, Y. Chen, S. M. Kim, and T. He, "Networking support for bidirectional cross-technology communication," *IEEE Transactions on Mobile Computing*, vol. 20, no. 1, pp. 204–216, 2021.
- [14] H. Wang, "Design of illuminance acquisition and Control program for mobile terminal based on ZigBee Technology and ESP8266WIFI Platform," *Software Engineering*, vol. 22, no. 9, pp. 28–30+56, 2019.
- [15] M. Gamal, N. Sadek, M. R. Rizk, and M. A. E. Ahmed, "Optimization and modeling of modified unslotted CSMA/CA for wireless sensor networks," *Alexandria Engineering Journal*, vol. 59, no. 2, pp. 681–691, 2020.
- [16] M. Yang, F. Gu, J. Liu, and L. Wang, "An anti-interference synchronization for OFDM systems based on scrambling sequence," *IEEE Access*, vol. 7, pp. 51121–51128, 2019.
- [17] M. Xiao, "Application research of ZigBee technology in intelligent agricultural greenhouse Monitoring system," *Hubei agricultural mechanization*, vol. 16, pp. 48–49, 2019.
- [18] W. Y. Wang, G. L. Li, C. Zhou, and C. Chen, "Design of remote monitoring system for greenhouse based on ESP8266," *Science and Technology innovation and Application*, vol. 6, pp. 53–56, 2021.
- [19] X. Li, T. Xu, and Q. T. Le, "Exploration of underground mechanical environment monitoring system according to Zigbee wireless communication," *International Journal of Mechatronics and Applied Mechanics*, vol. 2, no. 8, pp. 136–143, 2020.
- [20] R. Paul, Y. J. Choi, J. Jang, and Y. S. Kim, "Channel hopping using $\{p\}$ -ary m-sequence for rendezvous in cognitive radio networks," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1516–1519, 2019.
- [21] J. H. Li, Z. T. Chi, and S. C. Cui, "Design of intelligent classroom control system based on ZigBee," *Journal of Qingdao University (Engineering and Technology Edition)*, vol. 349, no. 3, pp. 84–88, 2019.
- [22] Y. L. Xia, Z. M. Zhou, J. Wang, and W. Li, "Design of wireless sensor network for structural strain monitoring based on ZigBee," *Electronics Test*, vol. 17, pp. 61–63, 2019.
- [23] K. X. Mo, Y. F. Li, Y. J. Zhuang, and C. Wang, "Data acquisition and monitoring system of Photovoltaic power station Environmental monitoring Instrument based on ZigBee," *Engineering research*, vol. 4, no. 18, pp. 1–3, 2019.

Research Article

Machine Learning Application of Transcranial Motor-Evoked Potential to Predict Positive Functional Outcomes of Patients

Mohd Redzuan Jamaludin ¹, Khin Wee Lai ¹, Joon Huang Chuah ²,
Muhammad Afiq Zaki ³, Khairunnisa Hasikin ¹, Nasrul Anuar Abd Razak ¹,
Samiappan Dhanalakshmi ⁴, Lim Beng Saw ⁵, and Xiang Wu ⁶

¹Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia

²Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur 50603, Malaysia

³Center of Environmental Health and Safety, Faculty of Health Sciences, Puncak Alam Campus, Universiti Teknologi Mara Selangor, Bandar Puncak Alam 42300, Selangor Darul Ehsan, Malaysia

⁴Department of Electronics and Communication Engineering, College of Engineering and Technology, Faculty of Engineering and Technology, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Chengalpattu, Chennai, Tamil Nadu, India

⁵Department of Orthopaedic Surgery, Sunway Medical Centre, Selangor, Malaysia

⁶School of Medical Information & Engineering, Xuzhou Medical University, Xuzhou 221000, China

Correspondence should be addressed to Khin Wee Lai; lai.khinwee@um.edu.my and Xiang Wu; wuxiang@xzhmu.edu.cn

Received 3 March 2022; Revised 14 April 2022; Accepted 21 April 2022; Published 20 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Mohd Redzuan Jamaludin 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.

Intraoperative neuromonitoring (IONM) has been used to help monitor the integrity of the nervous system during spine surgery. Transcranial motor-evoked potential (TcMEP) has been used lately for lower lumbar surgery to prevent nerve root injuries and also to predict positive functional outcomes of patients. There were a number of studies that proved that the TcMEP signal's improvement is significant towards positive functional outcomes of patients. In this paper, we explored the possibilities of using a machine learning approach to TcMEP signal to predict positive functional outcomes of patients. With 55 patients who underwent various types of lumbar surgeries, the data were divided into 70:30 and 80:20 ratios for training and testing of the machine learning models. The highest sensitivity and specificity were achieved by Fine KNN of 80:20 ratio with 87.5% and 33.33%, respectively. In the meantime, we also tested the existing improvement criteria presented in the literature, and 50% of TcMEP improvement criteria achieved 83.33% sensitivity and 75% specificity. But the rigidity of this threshold method proved unreliable in this study when different datasets were used as the sensitivity and specificity dropped. The proposed method by using machine learning has more room to advance with a larger dataset and various signals' features to choose from.

1. Introduction

Disc herniation and prolapsed disc that compresses the nerve roots in the lumbar region can cause sensory and motor disturbances, which contribute to low back pain, leg pain, and weakened leg's motor strength [1]. Decompression surgery or discectomy is a treatment surgery of removing the bulging disc from compressing the nerve roots.

The use of intraoperative neuromonitoring (IONM) in the lumbar discectomy procedure helps to monitor the

integrity of the nervous system from further injury. IONM modalities such as somatosensory-evoked potential (SSEP), motor-evoked potential (MEP), and electromyogram (EMG) are commonly used in lumbar surgery. However, SSEP has limited function since it only monitors specific nerve roots that are innervated from the S1 level when the posterior tibial nerve is stimulated [2]. It is also less sensitive to the changes of nerve root function because the signal is the result of a summation of neural signal from multiple segments before it enters the spinal cord [3]. EMG can show

continuous nerve root events since it is free running, but it has a high false positive rate and low specificity in determining events that are significant for the surgeon to alert on [4]. MEP or transcranial motor-evoked potential (TcMEP) has high sensitivity and specificity towards nerve root injury, but it relies on the alarm criteria used and which myotome is monitored [5].

Besides showing potential nerve injury during surgery, several studies which were reviewed by [6] presented in Table 1 had shown that improvement to the TcMEP signal is significant towards positive functional outcomes of patients. Most of the researches utilized amplitude increment as the improvement indicator. However, they have no common agreement on what is the increment percentage that is considered significant to show actual improvement in the postsurgery patient outcome. This paper aims to use machine learning algorithm to classify TcMEP signals into no improvement and improvement.

2. Related Works

2.1. Literature Reviews on Automated IONM, Objective Interpretation on IONM, and Machine Learning Applications on IONM. Currently, automated feature in commercial IONM machines only exists for triggered EMG (trigEMG) modality. It is done by attaching/clipping a stimulating device with a conductive surgical instrument. While the conductive surgical instrument is advanced deeper inside the spine body, increasing continuous stimulus is applied automatically to detect any nearby nerve structure so that nerve injury can be avoided [16]. The technique is straightforward, such that if the nerve is nearby the instrument, less stimulus is required to trigger the nerve. If a higher stimulus is required to trigger the nerve, it means that the nerve is further away from the instrument and safe to operate. It uses a basic principle by setting a threshold level and does not require an algorithm to interpret the signal. However, this technique in Malaysia is often neglected because of the high additional cost incurred by the IONM service. Moreover, this modality does not serve as a prognostic tool to determine the neurological condition of the patient.

There were several researches that used IONM to predict patients' functionality outcome by using an automatic algorithm. The first research was done by [17] to apply a deep learning algorithm on visual evoked potential (VEP) modality in order to detect changes to the VEP signal during sellar region tumours surgery. Another research was done for automatic SSEP interpretation by [18] to be used in cardiac surgery for peripheral nerve injury prevention. The application of deep learning and automatic interpretation of VEP signal and SSEP signal is supported by the fact that there is a universal acceptable criterion of normal and abnormal VEP and SSEP signal patterns to make the prediction possible. As far as our knowledge is concerned, no TcMEP signal has been applied for automatic patients' functionality outcomes categorization.

The interpretation of TcMEP is more difficult even by the IONM team because of high variability trial-to-trial, anaesthetic effect, and high sensitivity [19]. Interpretation of

TcMEP signal drop or usually known as alarm criteria is based on the complexity of the surgical procedure [20]. Furthermore, if changes are observed on TcMEP signals by the monitoring personnel, he or she has to go through a checklist of troubleshooting before the final interpretation is made to minimize the possibility of false positive or false negative events [21, 22]. Among the checklist that needs to be clarified before the surgical reversal is initiated are technical aspects (electrodes and machine connections and stimulation parameters) and anaesthesia/systemic (patient's mean arterial pressure (MAP), blood pressure, body temperature, and anaesthesia used).

2.2. Literature Review on the Applicability of TcMEP as Prognostic Tool to Justify Positive Functional Outcome of Surgery. In creating a meaningful prognostic application to justify the outcome of the surgery, we first need to identify if TcMEP has been used for that particular purpose in the literature. Considerable efforts have been exerted to address the ability of TcMEP to predict the risk of injuries to the nervous system, but only a few have focused on proving the ability of TcMEP to predict positive functional outcome of the patients intraoperatively. Studies that showed TcMEP as a prognostic tool to correlate the TcMEP improvement with the improvement after surgery were reviewed by [6] and are presented in Table 1.

A previous study by Barley et al. [7] utilized TcMEP on a 15-year-old boy who presented with upper and lower limbs motor weakness during a tethered spinal cord release procedure. Only the right abductor pollicis brevis (APB) of the upper limb was obtained, and no other responses were observed. Postdetethering revealed that left APB response appeared and increment of right APB amplitude, but only left upper extremity had notified improvement after surgery. Another research had proven that patients with improved TcMEP signal had better American Spinal Injury Association Impairment Scale (AIS) [13]. He et al. [15] presented a case report of a patient that had percutaneous endoscopic lumbar discectomy with TcMEP monitored and discovered that the TcMEP amplitude increment after the decompression was associated with low back and leg pain relief immediately after the patient was awake. Another study used IONM on 12375 patients who had spinal surgeries over 25 years with 386 patients exhibiting IONM signals improvement [10]. However, there were several IONM modalities (including TcMEP) that they monitored without specifying the improvement indicator. One patient had a permanent neurological deficit despite having IONM signal improvement, but this is statistically significant with the 14 true negative cases. Meanwhile, Wang et al. [12] had found that improvement of TcMEP signals (specifically amplitude rather than latency) in 59 patients who went through cervical laminoplasty or laminectomy was highly correlated with a modified Japanese Orthopaedic Association scale or mJOA improvement rate. Even though these studies had proven that the improvement on TcMEP had a high correlation with functional patient outcome, these five studies did not mention any specific improvement criteria from the IONM

TABLE 1: The summary of studies that indicated that TcMEP can be used as a prognostic tool [6].

Number	Reference	Number of samples	IONM modalities used	Stimulation parameters	Muscles used to monitor MEP	Improvement criteria	Results
1	Barley et al. [7]	One (15-month-old boy)	TcMEP and SSEP	C1-C2 scalp electrode positioning, current stimulation (145 mA to 187 mA for the left extremities and 175 mA to 200 mA for the right extremities) C1-C2 with multipulse current stimulation, 0 mA to 200 mA, stimulus duration 0.2 ms to 0.5 ms	Bilateral quadriceps femoris, tibialis anterior, gastrocnemius, sphincter, abductor pollicis brevis, and abductor hallucis	Not mentioned	TcMEP response of the left APB had an increment in amplitude. The patient had observable left upper extremity improvement
2	Voulgaris et al. [8]	25 (2 had no IONM results)	TcMEP and EMG	C3-C4 stimulation	Not mentioned	>50% MEP amplitude improvement	17 patients with >50% improvement had better VAS score improvement
3	Rodrigues et al. [9]	One (case report)	SSEP, MEP, and free running EMG	C3-C4 stimulation	Not mentioned muscles' names specifically but monitoring covered L3-S2 myotomes Upper extremity TcMEP was recorded from deltoid, flexor/extensor carpi radialis, and/or abductor digiti minimi/abductor pollicis brevis. Lower extremity TcMEP was recorded from anterior tibialis, medial gastrocnemius, and/or extensor hallucis longus	Not mentioned	MEP improved as much as 30%, and patient had returned to sports The results did not mention specifically TcMEP improvement, but out of the modalities used, 88.7% of patients had IOM signals improvement, but one patient out of this percentage had permanent neurological deficit
4	Raynor et al. [10]	386 patients had IOM signals improvement out of 12375 patients who had spinal surgeries over 25 years	DNEP, TcMEP, spontaneous EMG, triggered EMG, and dermatomal SSEP	C3-C4 TcMEP scalp electrode stimulation montage	Not mentioned	Not mentioned	

TABLE 1: Continued.

Number	Reference	Number of samples	IONM modalities used	Stimulation parameters	Muscles used to monitor MEP	Improvement criteria	Results
5	Visser et al. [11]	74 patients	TcMEP	Cz-Fz with monophasic stimulation and C3-C4 with biphasic stimulation	For the lower limbs, the quadriceps muscle (L2-L4), the tibialis anterior muscle (L4-L5), the hamstrings (L5-S1), or the gastrocnemius muscle (S1-S2). For cervical, the bilateral trapezoid muscle (C2-C4), the biceps (C5-C6), and triceps muscle (C7-C8) of the arm; the extensor muscles of the forearm (C6-C7); or the abductor digitus V muscle (C6-C8)	>200% of amplitude increment	There is a correlation between the duration of symptoms onset and the MEP improvement. MEP improvement can be accurate if the symptoms' onset duration is less than half a year
6	Wang et al. [12]	59 patients who had cervical myelopathy who underwent laminoplasty or laminectomy	MEP and SSEP	Not mentioned	Not mentioned	Not mentioned	Patients who had MEP signals improvement had a significant mJOA improvement rate. MEP amplitude was found to be a more accurate parameter compared to MEP latency in predicting surgery outcome
7	Dhall et al. [13]	32	EMG, MEP, and SSEP (not used for the study)	100 V-1000 V constant voltage stimulation, C1-C2 anodal stimulation, double train with a total of 9 pulses, 50 ms pulse width, 1.7 ms interstimulus, and 13.1 ms ISI	Not mentioned	Comparison with AIS grade and BASIC score of MRI images	MEP outcome (present) highly correlated with better AIS grade and BASIC grade
8	Piasecki et al. [2]	18	MEP and SSEP (not used for the study)	50 V-150 V C1-C2 biphasic stimulation, 5 to 7 train pulses, 500 Hz, and 1 ms interstimulus pulse	One upper limb muscle (control), bilateral tibialis anterior/bilateral abductor hallucis	>20% of AUC MEP; > 50% of ZCQ score	The MEP improvement was related to the early follow-up functional outcome

TABLE 1: Continued.

Number	Reference	Number of samples	IONM modalities used	Stimulation parameters	Muscles used to monitor MEP	Improvement criteria	Results
9	Wi et al. (2019) [14]	29 patients who had improvement in IONM signals out of 317 cases	MEP and SSEP	Not mentioned	Upper extremity TcMEP was recorded from deltoid, triceps, and thenar muscles. Lower extremity TcMEP was recorded from anterior tibialis and abductor hallucis Bilateral iliopsoas, rectus femoris, tibialis anterior, and medial gastrocnemius	Comparison with MISS, SF-36, JOA, NDI, and Oswestry Disability Index	The patients with MEP improvement had a better MISS improvement rate, while the patients with SSEP improvement only had a better SF-36 improvement rate MEP improvement aligned with the patient's relieved symptoms
10	He et al. [15]	One (case report)	MEP and free running EMG	Not mentioned		Not mentioned	

signal that they used to indicate significance towards postsurgery improvement.

Prospective research was done by Piasecki et al. [2] on patients who went through lumbar decompression surgery with TcMEP monitored. By using 20% of TcMEP area under the curve (AUC) increment as improvement criteria, they compared the findings with the Zurich Claudication Questionnaire (ZCQ) assessment of patients. It was found that the patients with improved TcMEP's AUC had higher ZCQ scores. A study conducted by Rodrigues et al. [9] had TcMEP monitored during a decompression surgery of a 22-year-old male athlete who was having lumbar pain and weakness on the right foot. The patient had a 30% of TcMEP amplitude increased after the discectomy procedure and was able to return to competitive athletic activities a month after surgery. Visser et al. [11] had used 200% of TcMEP amplitude increment criteria as significant to show actual patients' neurological improvement. This threshold was suggested by them to rule out any influence possibilities on the TcMEP generations. They found out that the MEP improvement should also be associated with the symptom's duration of less than 6 months for it to be significant. Another research that was done by Voulgaris et al. [8] compared the TcMEP outcomes with pain visual analogue scale (VAS) in patients who underwent lumbar decompressive laminectomy. The TcMEP improvement criteria were set on a 50% increment mark, and it was shown that the patients who had more than 50% TcMEP-increased amplitude had better VAS score at 12-month follow-up compared to the others who had lower amplitude increment. Wi et al. [14] found out that patients with more than 100% amplitude increment had better Motor Index Scoring System (MISS). It was then concluded by Wi et al. [14] that improvement of IONM signals could indicate the success of decompression. However, among these five studies, they have no common amplitude increment percentage that they used in their studies to indicate significance (>50% in [8] and >200% in [11]).

The proposed approach of this paper requires features that are the signal characteristics or parameters that we need

to feed to the machine learning models. Hence, we will make use of the TcMEP signal's parameters that were already established as presented in Table 1, which are the peak-to-peak amplitude and the AUC values. We also added the onset latency of the signal as one of the features selected for the machine learning models. The onset latency has not been used to predict the functional positive outcome. But it was used in [23] as the alarm criteria to indicate significant postoperative motor deficit. It was proven in the study that the onset latency had high sensitivity and specificity (100% and 84%, resp.) towards the detection of motor deficit compared to the amplitude threshold criteria (using more than 70% drop of amplitude as a significant indicator) at 100% sensitivity and 72% specificity. But when they combined both criteria (amplitude and onset latency), the sensitivity remained at 100%, and the specificity increased to 93%. Hence, in our study, we decided to include the onset latency parameter to be experimented as one of the features to run with the machine learning models.

In this paper, we are proposing a machine learning approach to be applied to the TcMEP signal that could identify patients that would have positive functional outcome and patients that has no changes from presurgery to postsurgery. We will also compare the efficacy of our results with two of the presented approaches in Table 1 from [8, 11] that utilized the TcMEP amplitudes threshold (>50% of amplitude increment and >200% of amplitude increment, resp.) as their improvement criteria since we are also using the peak-to-peak amplitude as one of the features that we used for the machine learning.

3. Methods

3.1. Data Source. The TcMEP data are the selected 55 patients who underwent lumbar disc decompression surgeries and patients who underwent instrumentation and correction surgeries from August 2021 until January 2022 at Sunway Medical Centre, Malaysia. Among the 55 patients, 13 patients had presurgery motor weakness and developed

positive motor improvement after surgery (named as group MI for motor improvement), 34 patients had presurgery symptoms such as numbness, back pain, and leg pain which had improved after surgery (named as group PNR for pain and numbness relief), and eight patients had no symptoms before surgery and after surgery (named as group NC for no changes). The actual outcomes of the postsurgeries were recorded based on the attending surgeons' evaluations. The eight patients who had no symptoms before surgery and after surgery were patients who only had instrumentation and correction surgeries.

All of the patients went through the surgery with the aid of IONM consisting of SSEP, MEP, and EMG modalities as requested by the surgeons. The myotomes involved in all of the surgeries were different depending on the spine levels that were being operated on. Since all of the operations involved spine lumbar L2 and below, the monitoring included vastus lateralis (VL) muscle (innervated from L2 to L4 nerve roots), tibialis anterior (TA) that innervated from the L5 nerve root, and abductor hallucis (AH) that innervated from S1 to S2 nerve root. Reference electrode was placed on hand muscle abductor digiti minimi (ADM) or abductor pollicis brevis (APB), whichever is easily accessible. Only TcMEP data were chosen for this study. Since the parameters used on the IONM and the approaches of the IONM setup on the patients were not experimental, they were all applied accordingly to the necessities of the surgeries and the surgeons' requests, and this study is not categorized as a prospective study. This study was approved by Sunway Medical Research Council, and since it was only a retrospective study, no informed consent was obtained from the patients.

3.2. Intraoperative TcMEP Monitoring. TcMEP monitoring was applied using the NIM Eclipse E4 system (Medtronic, Minneapolis, MN). The TcMEP was stimulated by using corkscrew electrodes placed at the C3 and C4 over the motor cortex on the scalp according to the International 10–20 System scalp electrode placements. The monitoring electrodes were placed at different muscles bilaterally by using dual subdermal needle electrodes.

The stimulus intensity varied from 250 V to 600 V. Train pulse stimulus of three to five pulses was applied to most of the patients to overcome the response variability or inconsistency. Sometimes, double train stimulation of five pulses and three pulses was used if it was difficult to elicit TcMEP response. Interstimulus interval was set to be either at 5 ms or 10 ms, which was based on which produced the better MEP response. Overall, the ideal TcMEP stimulation would be to elicit a response of more than 20 μ V of peak-to-peak amplitude for each channel with minimal patient movement.

Short-acting muscle relaxant was used in all of the patients during intubation, and the anaesthetic protocol was maintained with total intravenous or TIVA for the rest of the surgery. However, some of the anaesthetists applied inhalational agents such as desflurane and/or sevoflurane during intubation which made the baseline reading establishment

difficult and affected the interpretation of IONM. One anaesthetist had used midazolam on the patient.

Ideally, the baseline reading was obtained after the patient was intubated and before any incision was made, but most of the time, the muscle relaxant used during intubation lasted longer and required some time to completely wear off. So, the baseline readings were at least obtained before the spine area was fully exposed and operated on. The TcMEP was then stimulated from time to time to compare with the baseline reading so that any deterioration could be detected and reversed if needed.

3.3. Feature Selection and Machine Learning. The features that are used for the proposed approach in this paper were the onset latency which is the start point of the MEP signal in ms [23], the peak-to-peak amplitude in μ V, and AUC. Figure 1 shows the frame of an MEP signal from the start of the response to the end of the response. Figure 2 shows the definition of onset latency, peak latency, peak amplitude, and peak-to-peak amplitude of an MEP signal. These values were obtained straight from the NIM Eclipse E4 system. The AUC was obtained through NIM Eclipse E4 by selecting the section of TcMEP response which is within the start point of the signal until the end point of the signal as presented in Figure 1.

We went through the recorded patients' history in the IONM machine to identify which of the patients had presurgery neurological symptoms and which of the patients had no presurgery neurological symptoms. These data were obtained from preclinical evaluation made by the surgeons on their patients and were recorded in NIM Eclipse E4 during the surgery for future references. Then, we further investigated along the timeline of each patient's comment history to find at what time was the best TcMEP baseline achieved and the final reading for analysis. Then, we recorded three TcMEP features (peak-to-peak amplitude, onset latency, and AUC) from these two times. The readings were obtained from one target muscle and three reference muscles. The target muscle is the muscle that indicates weakness or pain. The three reference muscles were supposed to represent the asymptomatic myotome of the patient and not involved in the surgical site such as the hand muscles or the side of the limb that was not symptomatic.

The relevance behind this was because it is the similar approach used by the IONM technicians intraoperatively, which is that any change (either drop or increase of amplitude or latency delay) of a certain myotome, especially the target myotome, is compared with reference readings from reference myotome. This is based on the idea that there should be no significant changes in the reference myotome during the surgery, and any changes to the target myotome can be interpreted as significant and require further attention.

The final features that were used for the machine learning were obtained from the percentage difference between features from the baseline reading against features from the final reading. We also added the averaged values of peak-to-peak amplitude, onset latency, and AUC from all of the four muscles as another feature.

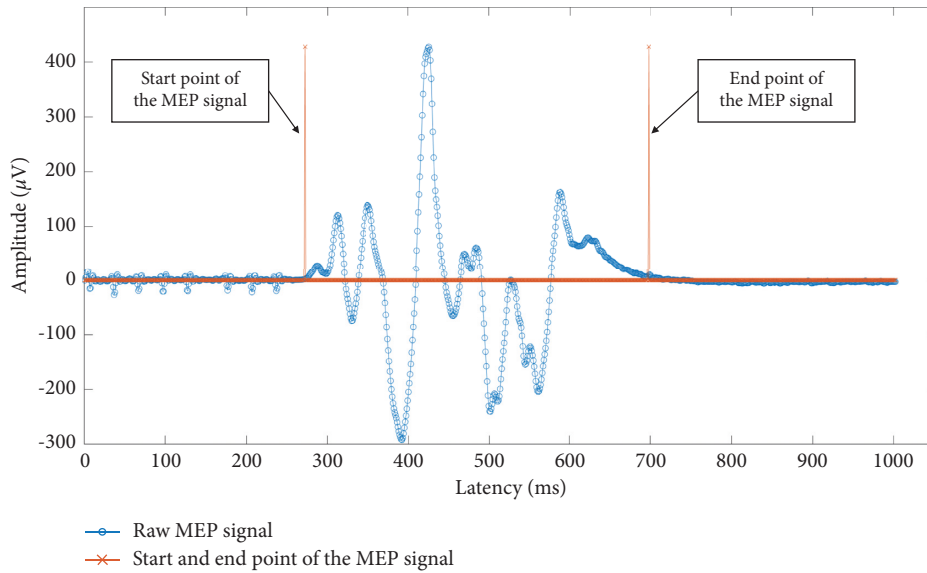


FIGURE 1: Start and end point marks at the raw MEP signal.

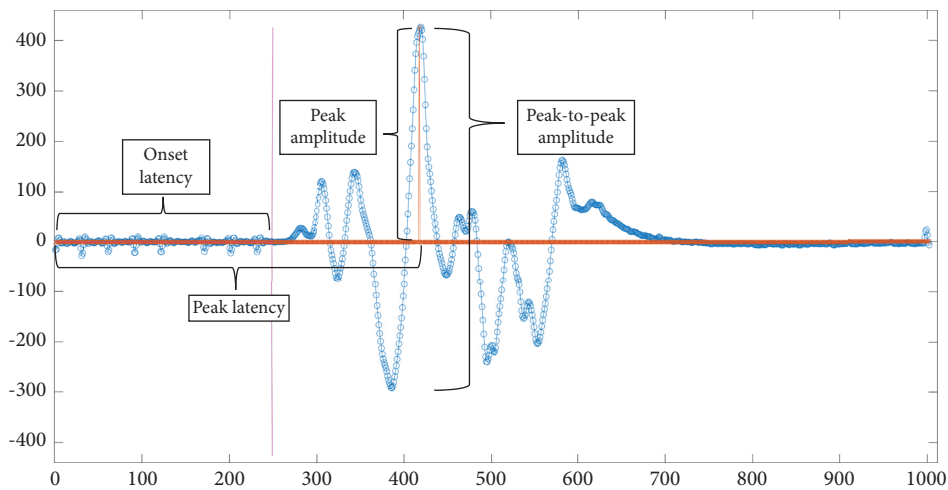


FIGURE 2: Onset latency, peak-to-peak amplitude, peak amplitude, and peak latency labels on raw MEP signal.

Consequently, the 55 patients were split into training and testing at two different ratios, which were 70:30 (39 patients for training and 16 patients for testing) and 80:20 (44 patients for training and 11 patients for testing). Since the data was quite small, we carefully selected the patients from MI, PNR, and NC groups for them to have equal numbers in both training and testing. The baseline readings of three MI patients and one NC patient were highly influenced by anaesthesia. The anaesthetist for these four patients used midazolam during the intubation, which highly suppressed the baseline reading. We included these samples because firstly we need the NC sample because the sample number of NC is limited and small. Secondly, if the proposed approach is able to classify the different groups despite the baseline readings being poor because of the anaesthetic factor, this study could be beneficial for current and future works since the chance of getting a false positive result is high with these types of samples. The two ratios (70:30 and 80:20) were chosen as it is commonly accepted

as the training and testing ratio in machine learning studies. Moreover, through our experimental procedures, the results obtained from the ratios 70:30 and 80:20 are worth presented and significant in this study compared to the results obtained from the other ratio sizes.

Test run was done with different sets of features:

- (i) Set A: all features included.
- (ii) Set B: only target and three reference muscles peak-to-peak amplitudes and onset latencies.
- (iii) Set C: target and three reference muscles peak-to-peak amplitudes, onset latencies, and AUCs.

We utilized the classification learner application in MATLAB to run the samples for training. After the training run was done with all of the models available in the classification learner application, we identified which of the models had 100% accuracy. We then chose these models with 100% of accuracy to run testing with the remaining

samples. Specific models that were used and the results achieved will be presented in Section 4.

We also utilized the improvement criteria presented by [8, 11], which were more than 50% of peak-to-peak amplitude increment and more than 200% of peak-to-peak amplitude increment, respectively, to compare the accuracy of their proposed threshold against our proposed model. The way we applied these criteria was by setting the rule as follows:

- (i) We used the target muscle for each patient. Target muscle is the muscle that is claimed to be having symptoms before surgery.
- (ii) We obtained the percentage difference between the amplitudes from the baseline reading and the final reading of the target muscle's TcMEP response.
- (iii) For 50% rule, if the difference in amplitude is more than 50%, then the result will be a true improvement. Otherwise, the result will be no changes.
- (iv) For 200% rule, if the difference in amplitude is more than 200%, then the result will be a true improvement. Otherwise, the result will be no changes.

4. Results

The results are shown in Table 2 for 70 : 30 ratio of training and testing samples and Table 3 for 80 : 20 ratio of training and testing samples. The first two columns of both tables showed the results of using the 50% amplitude increment criteria and 200% amplitude increment criteria on the same testing samples for comparison with the proposed method. The second column of each table represents Set A, the third column of each table represents Set B, and the fourth column of each table represents Set C as described in Section 3.3. Only the machine learning models that achieved 100% accuracy during the training session were selected for the testing session, and these are presented in both tables.

The machine learning models' ability to detect the positive outcome showed high sensitivity percentage. In Table 2, Fine KNN and Weighted KNN achieved 100% sensitivity in Sets A, B, and C. But they had no specificity, which means they were unable to identify patients who have no positive outcome. However, 50% amplitude increment criteria showed high specificity percentage on the same test samples used, which was 75% and had a high sensitivity of 83.33%.

Meanwhile, in Table 3, Fine KNN in Set B and Set C achieved 100% and 87.5% sensitivities, respectively, with 33.33% specificity in both sets. This indicates that the additional feature, which was the AUC in the Fine KNN model, dropped the sensitivity percentage. But when the 50% amplitude increment criterion was applied to the dataset in Table 3, the sensitivity and specificity dropped to 75% and 67%, respectively, compared to the result achieved by using the dataset in Table 2.

On the other hand, the other machine learning models that achieved 100% accuracy during the training session (Weighted KNN, Ensemble Subspace KNN, and Ensemble

Bagged Trees) in Sets A, B, and C produced 0% specificity in both Tables 2 and 3. Ensemble Subspace KNN in Table 2 had 8.33% false negative and 12.5% false negative in Table 3 when AUC was added as one of the features compared to when only peak-to-peak amplitude and onset latency were used. In fact, in Table 3, when peak-to-peak amplitude, onset latency, and AUC were used as features, Fine KNN, Ensemble Bagged Trees, and Ensemble Subspace KNN exhibited a 12.5% false negative rate. There was no false negative rate at both tables when Set B was used as the features.

The 200% amplitude increment criteria had lower sensitivity in both tables, which is 25% but had relatively higher specificity (75% in 70 : 30 ratio and 67% in 80 : 20 ratio) than the machine learning approach.

Overall, the proposed machine learning approach had higher sensitivity and lower specificity compared to the improvement criteria proposed by [8, 11]. We will look in-depth on what are the possibilities that lead to these findings in Section 5.

5. Discussion

In this paper, we proposed the utilization of machine learning to analyse the characteristics of TcMEP signals in order to group them objectively into TcMEP signals that significantly show positive functional outcome or TcMEP signals that do not indicate any positive functional outcome. We also compared the performance of the proposed approach with the amplitude criteria presented in the literature, and we found that the proposed method could potentially assist in the interpretation of TcMEP monitoring during lower lumbar decompression surgeries.

Among the tested machine learning models, Fine KNN with peak-to-peak amplitude and onset latency as the features achieved high sensitivity (100%) on TcMEP response to predict whether there will be a positive functional outcome to the patient or not. Several limitations to the proposed method are acknowledged. The proposed method needs input from the IONM technical personnel to select the best baseline signal for comparison with the final signal. Some of the surgeries finished in under two hours of time, for which the baseline readings were still influenced by muscle relaxant and inhalational agents induced during intubation, causing some false positive events to the analysis. After several encounters of difficulty to achieve the best baseline in some of the patients who had the same anaesthetist, the anaesthetist was enquired about his/her technique, and he/she admitted that midazolam was used in all of his/her patients. It is understood from the previous study made by [24] that midazolam could also highly suppress the TcMEP response. These patients were still included in this study for the sake of testing whether the prediction can still be made even though anaesthetic influence was involved.

With the small number of patients from the NC group, the machine learning was unable to distinguish clearly between patients who were expected to have improvement and patients who were expected to maintain their strength and functionality. This caused the low specificity in most of the models even though Fine KNN in 80 : 20 ratio exhibited

TABLE 3: Prediction performance of the models with 80 : 20 training samples and test samples ratio as the ability to identify the positive outcome.

	>200% method	>50% method	%All features										%Target and reference p2p amplitude and onset latency				%Target and reference p2p amplitude, onset latency, and AUC			
			Fine KNN	Fine KNN	Weighted KNN	Ensemble Bagged Trees	Ensemble Subspace KNN	SVM Fine Gaussian	Fine KNN	Weighted KNN	Ensemble Bagged Trees	Ensemble Subspace KNN	Fine KNN	Weighted KNN	Ensemble Bagged Trees	Ensemble Subspace KNN	Fine KNN	Weighted KNN	Ensemble Bagged Trees	Ensemble Subspace KNN
True positive	25%	75%	75.00%	100.00%	100.00%	100.00%	87.50%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	87.50%	100.00%	87.50%	87.50%	
True negative	67%	67%	33.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	33.33%	0.00%	0.00%	0.00%	
False positive	33%	33%	66.67%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	66.67%	100.00%	100.00%	100.00%	
False negative	75%	25%	25.00%	0.00%	0.00%	0.00%	12.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.50%	0.00%	12.50%	12.50%	

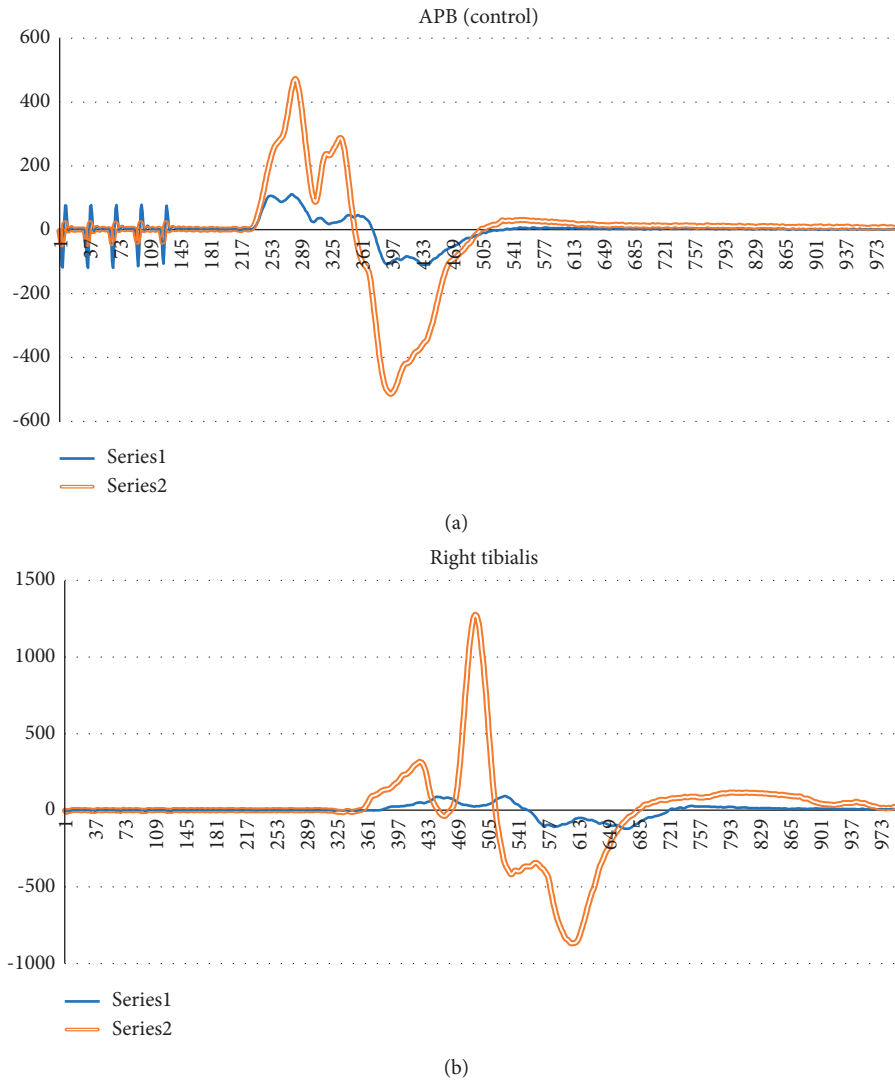


FIGURE 3: TcMEP responses of an NC sample. (a) Control muscle APB. (b) Target muscle right tibialis. Series 1 in both figures are baseline readings, and series 2 in both figures are final readings.

33.33% specificity. Additional numbers of data in the NC group would help in increasing the prediction capability of the proposed approach. Moreover, it can be observed that the specificity of Fine KNN improved when more samples were added to the training dataset (70:30 ratio compared with the 80:20 ratio). But we did not go for 90:10 ratio as 10% of samples for testing are too little and do not produce a significant finding.

Another finding was when AUC was added as the feature (in Set A and Set C), the false negative rate increased for Fine KNN, Ensemble Bagged Trees, and Ensemble Subspace KNN. Even though the study made by [2] presented that they used 20% AUC increment as the criteria of improvement, AUC had no significance in increasing the accuracy of the machine learning models in our study when it was added to the models' feature.

The NC group introduced by this paper should not be assumed as patients who develop postsurgery deficit. Although the introduction of a group of patients who develop

postsurgery deficit would be another valuable information that could lead to better patient management intra-operatively, we have not come across enough data for training and testing purposes (only two cases appeared during the period of the data collection).

Surprisingly, the 50% amplitude increment criteria achieved relatively high sensitivity and high specificity. It proves that this criterion is still relevant in most cases if it is added with additional input and interpretation from the neuromonitoring technician or specialist. However, it was observed with different sample sets (test samples in Table 2) that the sensitivity and the specificity by using this criterion dropped. Moreover, this threshold method could only be applied onto one target muscle. If we look further into one sample that this method predicted wrongly on one NC sample, the target muscle was set as right tibialis while the control muscle was set as the APB. This sample is shown in Figure 3. The behaviour of this sample shows that the baseline readings were influenced by muscle relaxants from

the intubation period since the final response of the APB muscle (nonsurgical site) increased drastically compared to the baseline response. By using this technique, the sample can be mistakenly identified as improved.

Meanwhile, the 200% criterion had a low true positive rate because the threshold was too high, and most of the MI and PNR samples were unable to achieve this criterion. Hence, we suggest that the threshold method is inflexible in such a way that if the threshold is too high, the rate of true positive will be lower, and if the threshold is too low, most samples will be predicted as having improvement, thus increasing the false positive rate.

We propose for future works that any prediction system by using the TcMEP signal can incorporate with the input from the comment history that records the events during the surgery. The system should be able to screen through the comment history to look for certain keywords such as “baseline,” “after decompression,” or “final TcMEP” so that the system can locate the best TcMEP signal for the interpretation process automatically.

6. Conclusion

IONM has been used widely to prevent or reduce the risk of nerve injury in spine surgeries. Alarm criteria to predict significant nerve injuries have been widely studied and established. However, the study of the use of IONM to predict improvement in patient’s symptoms can be improved further. Predicting the positive outcome of the decompressive surgery on the lower lumbar level intraoperative via IONM will be valuable information for the surgeon to make a decision on the depth of decompression needed. If the decompression is not done enough, patient might not get the intended result, but if the decompression is done too deep and too long, the risk of nerve injury increases, and it will prolong the duration of the surgery.

In this paper, we utilized machine learning models to predict the outcome of the intraoperative TcMEP. The best machine learning model that achieved a high percentage of sensitivity and specificity percentages was Fine KNN with only peak-to-peak amplitude of the target muscle and three reference muscles used as the feature parameters. The Fine KNN model performance was able to predict a positive functional outcome from the TcMEP response with high sensitivity but with low specificity. With the limitations on the whole TcMEP modality itself, any drop or change to the signal requires human technical involvement before the actual interpretation is passed. Thus, it is advised that the proposed system can only act as an assistant to the monitoring personnel. On the other hand, we managed to show that the 50% amplitude increment criterion is relevant as a predictor of positive functional outcome.

Data Availability

All the data are available from the list of references. Data presented in Figure 1, Figure 2, Figure 3, Table 2 and Table 3 were collected from Sunway Medical Centre, Malaysia upon medical ethics approval.

Ethical Approval

The data collection was approved by the medical ethics committee of SunMed Clinical Research Centre: SMRR no. SMRR/IIR/SMC/21/002 and SRB Approval no. SRB/NF/21/002.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Acknowledgments

This work was supported in part by the Impact-Oriented Interdisciplinary Research Grant, Universiti Malaya, under Grants nos. IIRG001B-2021IISS and ACU UK IF063-2021.

References

- [1] H. Matsui, M. Kanamori, Y. Kawaguchi, H. Kitagawa, H. Nakamura, and H. Tsuji, “Clinical and electrophysiologic characteristics of compressed lumbar nerve roots,” *Spine*, vol. 22, no. 18, pp. 2100–2105, 1997.
- [2] K. Piasecki, G. Kulik, K. Pierzchala, E. Pralong, P. J. Rao, and C. Schizas, “Do intra-operative neurophysiological changes predict functional outcome following decompressive surgery for lumbar spinal stenosis? A prospective study,” *Journal of Spine Surgery*, vol. 4, no. 1, pp. 86–92, 2018.
- [3] A. A. Gonzalez, D. Jeyanandarajan, C. Hansen, G. Zada, and P. C. Hsieh, “Intraoperative neurophysiological monitoring during spine surgery: a review,” *Neurosurgical Focus*, vol. 27, no. 4, p. E6, 2009.
- [4] R. N. Rattenni, T. Cheriyan, A. Lee, J. A. Bendo, T. J. Errico, and J. E. Goldstein, “Intraoperative spinal cord and nerve root monitoring: a hospital survey and review,” *Bulletin of the Hospital for Joint Diseases*, vol. 73, pp. 25–36, 2015.
- [5] W. B. Wilent, E. A. Tesdahl, J. S. Harrop et al., “Utility of motor evoked potentials to diagnose and reduce lower extremity motor nerve root injuries during 4,386 extradural posterior lumbosacral spine procedures,” *The Spine Journal*, vol. 20, no. 2, pp. 191–198, 2020.
- [6] M. R. Jamaludin, K. W. Lai, J. H. Chuah et al., “Transcranial electrical motor evoked potential in predicting positive functional outcome of patients after decompressive spine surgery: review on challenges and recommendations towards objective interpretation,” *Behavioural Neurology*, vol. 2021, Article ID 2684855, 16 pages, 2021.
- [7] J. L. Barley, J. F. Mooney, S. S. Glazier et al., “Sudden appearance of new upper extremity motor function while performing neurophysiologic intraoperative monitoring during tethered cord release: a case report,” *Journal of Pediatric Orthopaedics*, vol. 30, no. 6, pp. 624–628, 2010.
- [8] S. Voulgaris, D. Karagiorgiadis, G. A. Alexiou et al., “Continuous intraoperative electromyographic and transcranial motor evoked potential recordings in spinal stenosis surgery,” *Journal of Clinical Neuroscience*, vol. 17, no. 2, pp. 274–276, 2010.
- [9] L. M. R. Rodrigues, F. W. F. d Rosa, R. J. R. Ferreira, F. Ueno, and C. Milani, “Herniated lumbar disc surgery in triathlon athletes with intraoperative neurophysiologic monitoring,” *Einstein (Sao Paulo)*, vol. 9, no. 4, pp. 530–533, 2011.

- [10] B. L. Raynor, J. D. Bright, L. G. Lenke et al., "Significant change or loss of intraoperative monitoring data: a 25-year experience in 12,375 spinal surgeries," *Spine*, vol. 38, no. 2, pp. E101–E108, 2013.
- [11] J. Visser, W. C. Verra, J. M. Kuijlen, P. P. Horsting, and H. L. Journee, "Recovery of TES-MEPs during surgical decompression of the spine: a case series of eight patients," *Journal of Clinical Neurophysiology*, vol. 31, no. 6, pp. 568–574, 2014.
- [12] S. Wang, Y. Tian, C. Wang et al., "Prognostic value of intraoperative MEP signal improvement during surgical treatment of cervical compressive myelopathy," *European Spine Journal*, vol. 25, no. 6, pp. 1875–1880, 2016.
- [13] S. S. Dhall, J. Haefeli, J. F. Talbott et al., "Motor evoked potentials correlate with magnetic resonance imaging and early recovery after acute spinal cord injury," *Neurosurgery*, vol. 82, no. 6, pp. 870–876, 2018.
- [14] S. M. Wi, H.-J. Lee, T. Kang et al., "Clinical significance of improved intraoperative neurophysiological monitoring signal during spine surgery: a retrospective study of a single-institution prospective cohort," *Asian spine journal*, vol. 14, no. 1, pp. 79–87, 2020.
- [15] S. He, Z. Ren, X. Zhang, and J. Li, "Neurophysiologic monitoring for treatment of upper lumbar disc herniation with percutaneous endoscopic lumbar discectomy: a case report on the significance of an increase in the amplitude of motor evoked potential responses after decompression and literature review," *International Journal of Surgery Case Reports*, vol. 67, pp. 271–276, 2020.
- [16] B. M. Ozgur, H. E. Aryan, L. Pimenta, and W. R. Taylor, "Extreme Lateral Interbody Fusion (XLIF): a novel surgical technique for anterior lumbar interbody fusion," *The Spine Journal*, vol. 6, no. 4, pp. 435–443, 2006.
- [17] N. Qiao, M. Song, Z. Ye et al., "Deep learning for automatically visual evoked potential classification during surgical decompression of sellar region tumors," *Translational Vision Science & Technology*, vol. 8, no. 6, p. 21, 2019.
- [18] J. Chui, J. M. Murkin, T. Turkstra, N. McKenzie, L. Guo, and M. Quantz, "A novel automated somatosensory evoked potential (SSEP) monitoring device for detection of intraoperative peripheral nerve injury in cardiac surgery: a clinical feasibility study," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 31, no. 4, pp. 1174–1182, 2017.
- [19] D. B. MacDonald, S. Skinner, J. Shils, and C. Yingling, "Intraoperative motor evoked potential monitoring – a position statement by the American Society of Neurophysiological Monitoring," *Clinical Neurophysiology*, vol. 124, no. 12, pp. 2291–2316, 2013.
- [20] M. R. Weinzierl, P. Reinacher, J. M. Gilsbach, and V. Rohde, "Combined motor and somatosensory evoked potentials for intraoperative monitoring: intra- and postoperative data in a series of 69 operations," *Neurosurgical Review*, vol. 30, no. 2, pp. 109–116, 2007.
- [21] J. E. Ziewacz, S. H. Berven, V. P. Mummaneni et al., "The design, development, and implementation of a checklist for intraoperative neuromonitoring changes," *Neurosurgical Focus*, vol. 33, no. 5, p. E11, 2012.
- [22] M. G. Vitale, D. L. Skaggs, G. I. Pace et al., "Best practices in intraoperative neuromonitoring in spine deformity surgery: development of an intraoperative checklist to optimize response," *Spine Deformity*, vol. 2, no. 5, pp. 333–339, 2014.
- [23] K. Kobayashi, K. Ando, R. Shinjo et al., "A new criterion for the alarm point using a combination of waveform amplitude and onset latency in Br(E)-MsEP monitoring in spine surgery," *Journal of Neurosurgery: Spine*, vol. 29, no. 4, pp. 435–441, 2018.
- [24] C. Kalkman, J. Drummond, A. Ribberink, M. Patel Piyush, T. Sano, and R. Bickford, "Effects of propofol, etomidate, midazolam, and fentanyl on motor evoked responses to transcranial electrical or magnetic stimulation in humans," *Anesthesiology*, vol. 76, no. 4, pp. 502–509, 1992.

Research Article

Greedy Strategies with Multiobjective Optimization for Investment Portfolio Problem Modeling

Xinchen Zhang ¹, Linghao Zhang ¹, Qincheng Zhou,² and Xu Jin ²

¹School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

²School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

Correspondence should be addressed to Xinchen Zhang; b19011721@njupt.edu.cn and Xu Jin; jinxu@njupt.edu.cn

Received 8 April 2022; Revised 29 April 2022; Accepted 3 May 2022; Published 19 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Xinchen 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 ultimate purpose of portfolio investment is to reduce investment risk and improve total return on the premise of ensuring reasonable allocation of capital. In this paper, we build a quantitative model to advise on trading based on the price movement of Bitcoin and gold between 2016 and 2021; our goal is to maximize profit while minimizing risk. We mainly use greedy strategies with multiobjective optimization models. For the purpose of obtaining the correct price trend, some popular trend indicator strategies are referred to predict the future price trend in the medium and long term. In addition, we also consider people with different trading preferences and divided them into aggressive, advanced, balanced, and cautious and provided trading strategies for each of these four groups. This gives our model scalability. Finally, we analyze the sensitivity of the model and discuss the impact of trading commission costs on the model results. The model can be applicable to various investment situations.

1. Introduction

Investment portfolio refers to a series of assets, such as bonds, stocks, and currency derivatives. Usually, a high-quality investment portfolio should be high liquidity, high stability, high return, low investment risk, etc. Portfolio problem is a kind of very complex decision-making problem. The research on the problem of portfolio is helpful to improve the risk consciousness of decision-makers and can improve the rationality and effectiveness of investment decisions, avoid the blindness of investment, reduce its cost, and improve its return. Under normal circumstances, when investors make decisions, the most concerned issue for investors is to allocate the limited capital effectively and to arrange the order of asset investment reasonably, so as to maximize the overall return and minimize the risk of the portfolio. Investors have different preferences. Risk-averse people want to make investments with as little risk as possible and then they want to make returns; risk-loving people pursue the return on investment, but do not care

much about the risks of investment. They believe that high returns can be achieved only by taking high risks. However, even for investors with different attitudes, the ultimate purpose of portfolio investment is to reduce investment risk and improve total return on the premise of ensuring reasonable allocation of capital, so they should still be considered as the most important objective.

As a result of the advanced computer technology, quantitative trading based on modern financial investment theory has become a new hot spot in the development of capital investment market at home and abroad. Rapid response, product diversification, and risk control are the main advantages of quantitative trading. In addition to traditional hedge fund and futures trading, quantitative trading is increasingly being used to trade digital currencies (e.g., Bitcoin). Benefiting from the high volatility of digital currencies, many quantitative strategies have realized excess returns. Generally speaking, we believe that a quantitative strategy is a comprehensive study of market characteristics, market sentiment, and other multifaceted information in

order to obtain maximum returns with relatively tolerable risk.

In recent years, the application of machine learning in market forecasting has increased and has basically surpassed the prediction accuracy of traditional time series forecasting models. At the same time, we believe that Bitcoin and gold represent high-risk (up to more than 100 times in five years) and low-risk assets, respectively, and how to consider the dynamic allocation of assets to balance risk and return in trading is a central issue in our article.

In our article, we are required to accomplish the following tasks: (1) Bitcoin and gold spot are selected as a portfolio and we are required to use a quantitative strategy to determine the maximum return at the end of the period for a ternary portfolio (C, B, G) of cash, Bitcoin, and gold. (2) Proving the optimality of our trading strategy. (3) Analyzing the sensitivity of the model and discussing the impact of trading commission costs on the model results.

2. Related Works

2.1. Portfolio Optimization Theory. Portfolio optimization theory has been developed for nearly 60 years. Markowitz proposed the return as a random variable and created the mean-variance theory of portfolio selection in 1952, which marks the beginning of quantitative analysis method into the field of financial research [1]. It gradually became the focus of investment research. Konno used a mean-absolute deviation portfolio optimization model in the applications to Tokyo Stock Market [2]. Dellino et al. established a portfolio optimization model using the dynamic target aggregation method [3]. Mercurio et al. used a generalized entropic portfolio optimization (GEPO) [4] and a return-entropy portfolio optimization (REPO) [5] for option portfolio selection. Zhou proposes a multiobjective optimization model for investment portfolio problem and solves it by MOEAs [6]. Gong integrated dynamic risk-tolerance and expected-return levels to develop two coherent fuzzy multiperiod portfolio selection models [7]. Yin's study showed that basing on high-frequency prices data, the realized covariance matrix model can improve the accuracy of the portfolio market risk forecast significantly [8]. Jelena researched the effects of portfolio optimization and especially the Bitcoin investment on portfolio optimization [9]. Taking the risk and return variables into account, we can use different optimization techniques to create an optimal portfolio [10–13], which are aimed to create a balance between risk and return in the financial market.

2.2. Greedy Algorithm. Greedy algorithm is characterized for its simpler and faster method to solve some optimization problems, often based on the current situation as the basis of an optimization measure to make the optimal choice, without considering all possible overall situation, saving many time to find the optimal solution to exhaust all possible. It generally follow 4 steps: (1) setting up a mathematical model to describe the problem; (2) dividing the problem into several subproblems; (3) solving each

subproblem to obtain the local optimal solution; (4) synthesizing the local optimal solution of the subproblem to get a solution of the original problem. Because greedy algorithms always solve problems from the parts, there is no guarantee that the solution is optimal for the whole. Greedy algorithms have many important applications in n traffic offloading, medical service, flexible manufacturing, investment planning, etc. [14–19]. By introducing Chritos's greedy algorithm and the simulated algorithm into PSO, Tang proposed a hybrid particle swarm optimization algorithm with adaptive mulisections to make up for the deficiency of PSO [20]. Based on the greedy strategy, Khuller et al. proposed an approximate algorithm with an approximate ratio of $(1 - 1/e)$, although they used the local enumeration method which optimizes the approximate ratio of the algorithm, but the algorithm still has a high time complexity $O(n^{(k+1)} \log n)$, in solving large-scale BMCP (budgeted maximum coverage problem) cases, but also poor in solving results [21]. Zhang et al. proposed a novel $(1 - 1/e)$ approximation method for BMCP with $O(3n)$ [4] time complexity to improve the speed of this algorithm, but the approximate ratio is not yet improved [22]. Ashwin proposed an approximate algorithm A-SUKP based on greedy strategy to solve SUKP, in which the approximate ratio of A-SUKP is $1/(1 - e^{1/d})$, where d ($d \geq 2$) is the upper bound on the occurrence of all elements. Obviously, the approximate solution of A-SUKP is unsatisfactory and inefficient as d gets larger [23]. Li studied group orthogonal greedy algorithm for change-point estimation of multivariate time series [24].

3. Model Preparation

3.1. Data. The data used in our model for the empirical application consist only of historical price series for Bitcoin and gold, which were collected between 9/11/2016 and 9/10/2021. Gold daily prices (in US dollars per troy ounce) are sourced from the London Bullion Market Association, while Bitcoin daily prices (in US dollars per Bitcoin) are sourced from the Nasdaq Stock Market, take a look at Figures 1 and 2. Bitcoin can be traded every day, so the data are continuous. However, gold has a difference between trading days and nontrading days, and the data are not continuous. We first do a smoothed interpolation of the historical price of gold to facilitate forecasting model (nontrading day scenarios are taken into account in the trading strategy).

3.2. Our Research Framework. To simplify the problem, our model does not overly consider low probability events such as global financial crisis, financial systemic risk, natural disasters, and exchange closures, in other words, most of the normal risk is already reflected in the price changes. We consider Bitcoin to be a high risk asset and gold to be low. If we want to get higher returns, we will hold more Bitcoins and give up gold in bitcoin uptrends (under controlled risk). Trend and timing are important propellants to obtain excess returns. We only consider trading on the right side, making trades at the moment of uptrend or trend turn and will not ambush in downtrends.

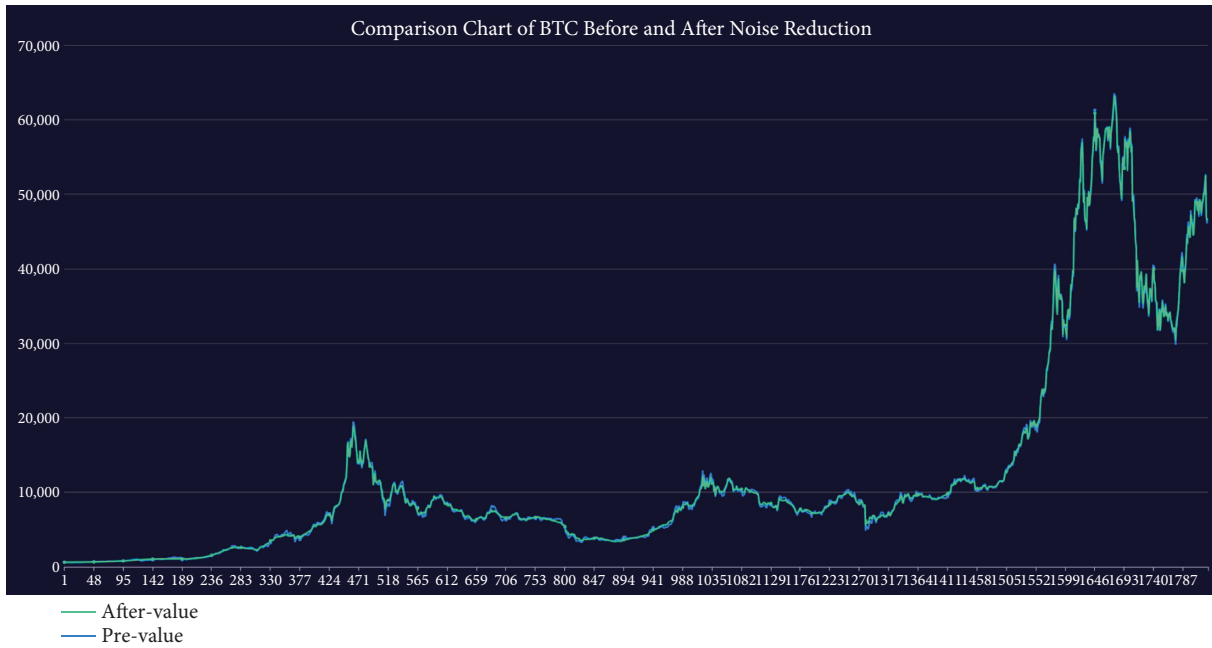


FIGURE 1: Gold daily price.

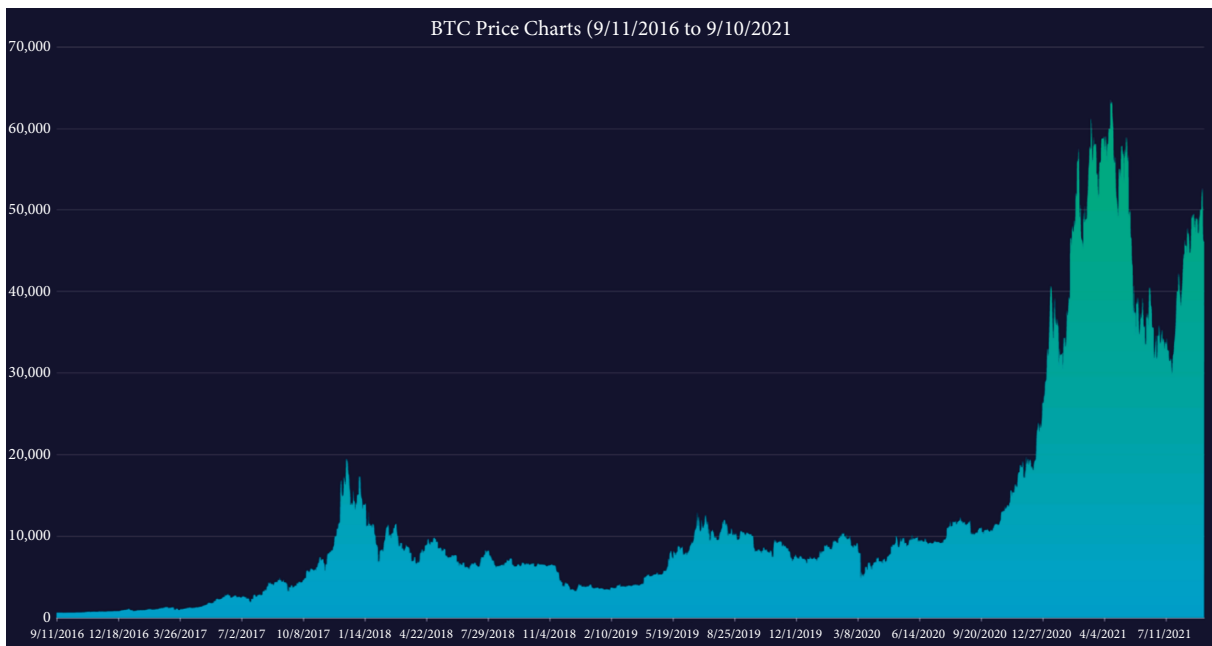


FIGURE 2: BTC daily price.

In this article, considering the background information and restricted conditions identified above, we will use multiobjective optimization to solve this problem. The paper is divided into three main parts for trend trading strategy, multiobjective optimization, and model sensitivity analysis, respectively.

3.2.1. Trend Trading Strategy. We believe that trends are very important. We need to consider the proportion of our portfolio

at every moment. A common quantitative strategy is commodity trading advisor strategy, which is called the CTA strategy, a strategy based on price trends. In the trading strategy, we use a LSTM-P neural network model for predicting the values of Bitcoin and gold for the next five days (in this article, we will not discuss LSTM-P neural network model). Then, based on our predicted prices, we combine some trending indicators to identify trading signals. We get the price trend with high probability through trend indicator analysis and use it to classify the trading behaviors into 5 categories.

3.2.2. Portfolio Optimization. We need to consider a return-risk biobjective optimization model. The objective functions are optimized through the expected wealth, the variance, and the conditional value at risk, including transaction costs, investor risk appetite, and investment limits for each asset. Common risk models include mean-variance model and multiobjective evolutionary algorithm. We finally refer to multiobjective model predictive control [25] and Pareto optimization principle to simplify and derive our return-risk model. We obtain the optimal portfolio allocation ratio as the local optimal solution by some constraints obtained by the greedy algorithm.

3.2.3. Model Sensitivity Analysis. After validity and stationarity tests, our article effectively demonstrates the validity and applicability.

4. Multiobjective Optimization Control (MOC)

By using the LSTM-P neural network model, we have got the price forecast for the next 5 days from the price data so far, and we have to determine the trading and portfolio ratios together based on the forecast and some other indicators. We believe that both risk and return are required to be considered, so we are going to make a multiobjective optimal control model (MOC) for risk and return.

Before building the MOC model, we introduce some trend technical indicators commonly used in financial quantitative analysis to get the forecast of future price trend. This will be part of the constraints of the multiobjective optimization later.

4.1. Trend Prediction Model. We believe that it is not enough to know the forecast data for the next few days. In order to reduce high-frequency trading and lower the error rate, we need to know the change of a price trend in the medium and long-term future (for example, one month).

We refer to 28 technical indicators such as MACD, BOLL, and KDJ. A correlation test was performed (the correlation heatmap is shown in Figure 3).

We finally choose two indicators, MACD and BOLL, to calculate the trend. The former can be briefly explained as the acceleration change of price movement, and the latter can be briefly described as the channel of price movement. The detailed explanation of the indicators is not described here.

Figure 4 shows the MACD and BOLL charts that we plotted for the Bitcoin price. We have only shown a portion of the time to make it look clear.

It is obvious that these two indicators give a good picture of the trend direction of Bitcoin in the medium to long term. We have divided the trend into an uptrend, a downtrend, and an oscillating trend. We obtain the future medium to long-term trend by combining these two indicators and the forecast price for the next 5 days and highlight the resulting uptrend in green (Figure 5).

We assume that it is more cost-effective to buy in uptrends, while buying in downtrends should be eliminated (or

positions reduced). This greed-like idea is quantified as constraints to perform multiobjective optimization.

That is, we give the future trend a score $Tr \in (-1, 1)$, indicating the probability of being an uptrend. As a greedy idea, we consider that the species cannot be bought when $Tr < 0$. Here, Tr denotes the index of future trend. We have also divided the buy and sell signals corresponding to the gold and Bitcoin trend combinations into 5 categories, which is a constraint and the value of S in the risk function CvaR.

Figure 6 shows the buy and sell signals corresponding to the 5 categories of our trend portfolio (green is BTC and yellow is gold).

4.2. Multiobjective Optimal Control Model. In order to obtain an optimal portfolio, we need to perform multiobjective dynamic optimization control. In this section, we derive the mathematical principles of the model in detail.

4.2.1. Description of the Model. Multiobjective optimization control (MOC) is used to study the optimization of more than one objective function and satisfies the contradiction between the objectives of multiobjective planning. The model of multiobjective planning consists of two parts: more than two objective functions and several constraints.

In the case described in our problem, our objectives are the maximization of benefits and the minimization of risks. In the following, we will describe the quantitative functions of return and risk, respectively, as well as the constraints on the final set of planning equations.

4.2.2. Returns Calculation

(1) Basic Variables. Suppose $\eta(k) \in \mathbb{R}^2$ is a time series of returns on two risky assets, gold and BTC. $\mathbb{F} = (F_k)_{k \geq 1}$ is a filter on the σ domain generated by $\eta(k)$ [26]; $a_1(k)$ and $a_2(k)$ represent the amount of gold and BTC at time k , respectively; $\eta_1(k)$, $\eta_2(k)$ represent the returns on gold and BTC at time k , respectively. Since we only invest the principal at the beginning, we do not add or withdraw wealth from the portfolio, and the variables $a_i(k)$, w_0 are all deterministic. Apparently, the value of the assets held at time k is

$$W(k) = a_1(k) + a_2(k). \quad (1)$$

In the next instant $k + 1$, the value of the assets held is

$$\begin{aligned} W(k+1) &= [1 + R_1(k+1)]a_1(k) + [1 + R_2(k+1)]a_2(k) \\ &= [a_1(k) + a_2(k)] + [R_1(k+1)a_1(k) + R_2(k+1)a_2(k)]. \end{aligned} \quad (2)$$

Then, let the return vector $R(k+1) = (\eta_1(k+1), \eta_2(k+1))^T$, $A(k) = (a_1(k), a_2(k))^T$; thus, (10) can be written as

$$W(k+1) = W(k) + R^T(k+1)A(k). \quad (3)$$

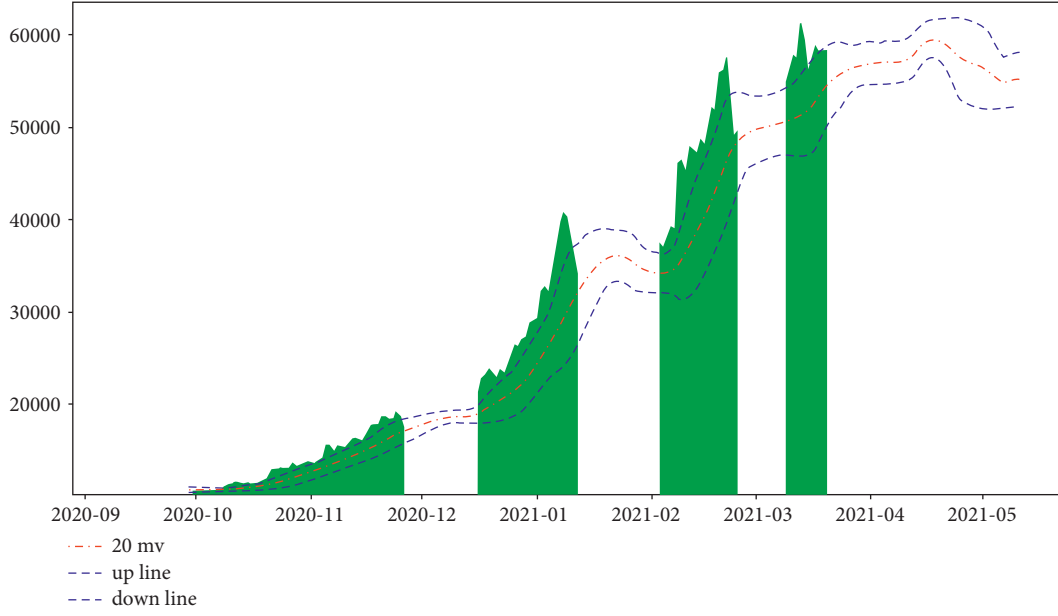


FIGURE 5: Charts of Bitcoin price uptrend period (2020-10 to 2021-05).

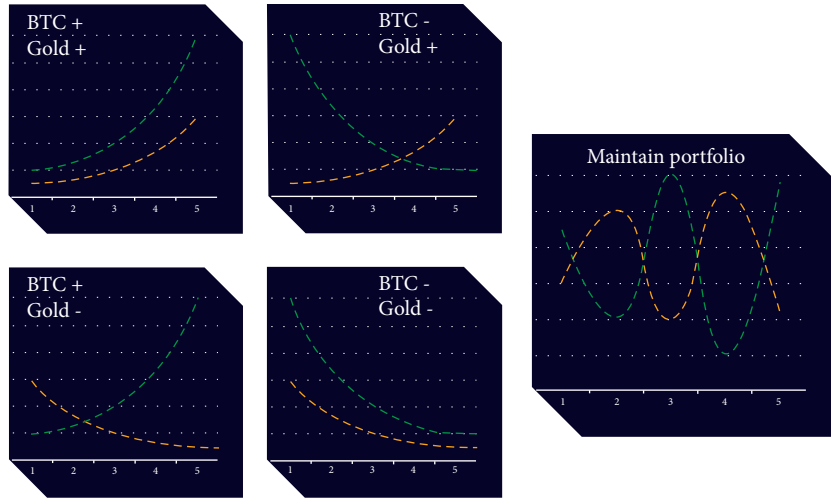


FIGURE 6: Trading operation chart under 5 trend combinations.

$$\begin{aligned}
 \bar{\eta}(k+m) &= E[\eta(k+m)|F_k]e\Sigma \\
 &= E[(\eta(k+i) - \bar{\eta}(k+i))(\eta(k+j) - \bar{\eta}(k+j))^T | F_k] \\
 &= \mathbf{A}_1^m \eta(k) + \sum_{j=0}^{m-1} c\mathbf{A}_1^j,
 \end{aligned} \tag{5}$$

where

$$\Sigma = \sum_{l=0}^{i-1} \mathbf{A}_1^{2i-2(l+1)} \Sigma w = \sum_{l=0}^{i-1} \mathbf{A}_1^{2i-2(l+1)} E[w(k+l+1)w^T(k+l+1)]$$

is a 2-order square matrix.

Iterating over (12), one can also calculate the asset value after time m .

$$\eta(k+m) = \mathbf{A}_1^m \eta(k) + \sum_{j=0}^{m-1} [\mathbf{A}_1^{m-j-1} w(k+j+1) + \mathbf{A}_1^j c]. \tag{6}$$

(3) *Conditional Covariance*. To calculate conditional covariance definition,

$$\hat{\eta}(k+m) = \eta(k+m) - \bar{\eta}(k+m) = \sum_{j=0}^{m-1} \mathbf{A}_1^{m-j-1} w(k+j+1). \tag{7}$$

Then, based on the above definition of the return vector $R(k+j)$, the conditional expectation vector and the conditional covariance vector are defined again here:

$$\bar{R}(k+j) = (\bar{\eta}_1(k+j), \bar{\eta}_2(k+j))^T \bar{R}(k+j) = (\bar{\eta}_1(k+j), \bar{\eta}_2(k+j))^T. \tag{8}$$

$$\text{Thus, } \bar{\Sigma} = E[\bar{R}(k+i)\bar{R}^T(k+j)|F_k] = \begin{bmatrix} \Sigma_{2 \times 2} & \mathbf{0}_{2 \times 1} \\ \mathbf{0}_{1 \times 2} & \mathbf{0} \end{bmatrix}.$$

According to the above corollary, the expression for the conditional value mean at the moment $k+m$ becomes obvious:

$$E[W(k+m)|F_k] = W(k) + \sum_{j=0}^{m-1} E[R^T(k+j+1)|F_k]A(k+j). \quad (9)$$

4.2.3. Risk Calculation. Assume $X_p = (x_1, x_2)^T$ represents the portfolio of gold and BTC bought or sold now, and set vectors $\xi = (\xi_1, \xi_2)$ and $R_n = (\bar{r}_1, \bar{r}_2)$ to represent the vector of returns and the vector of expected returns for the portfolio of gold and BTC, respectively. Here, $\bar{r}_i = E(\xi_i)$. Then, the expected value and return of the corresponding portfolio are

$$\begin{aligned} r_p &= \sum_{i=1}^n x_i \xi_i = X_p \xi, \\ \bar{r}_p &= \sum_{i=1}^n x_i \bar{r}_i = X_p R_n. \end{aligned} \quad (10)$$

We use CVaR to calculate risk. CVaR is risk evaluation (conditional covariance), also known as the average value of excess losses and the conditional value at risk. It represents the conditional mean value of the maximum loss of an investor's asset portfolio over VaR when investing at a certain confidence level.

Assume $f(A, R)$ is the loss function of the investor's assets, where $a \in A \subset \mathbb{R}^n$ and $\eta \in \mathbb{R}^n$. Definition domain U denotes the feasible region of the gold and BTC portfolio.

The random vector R in the n -dimensional vector space denotes the uncertainty variables that have an impact on the loss which the investor cannot capture.

By referring to the relevant literature [26], the CVaR value of an investor's asset portfolio at confidence level β and its approximate representation can be expressed as

$$CVaR_\beta(A) = \min_{\alpha \in \mathbb{R}} F_\beta(A, \alpha), \quad (11)$$

where $F_\beta(A, \alpha)$ can be approximated as $\tilde{F}_\beta(A, \alpha)$:

$$\begin{aligned} E[W(k+m)|F_k] &= W(k) + \sum_{j=0}^{m-1} E[R^T(k+j+1)|F_k]A(k+j) \\ &\quad - \sum_{j=0}^{m-1} \alpha \left| A(k+j) - (1 + E[R^T(k+j)|F_k])A(k+j-1) \right|, \\ E[W(k+m)|F_k] &= S^{-1} \sum_{s=1}^S E[W_S(k+j+1)|F_k] \\ &= S^{-1} \sum_{s=1}^S \left[W(k) + \sum_{j=0}^{m-1} E[R^T(k+j+1)|F_k]A(k+j) \right. \\ &\quad \left. - \sum_{j=0}^{m-1} \alpha \left| A(k+j) - (1 + E[R^T(k+j)|F_k])A(k+j-1) \right| \right]. \end{aligned} \quad (15)$$

4.2.5. Constraints. Investing in gold or BTC has a maximum money limit, which means "you cannot put your eggs in one

$$\tilde{F}_\beta(A, \gamma) = \gamma + (1 - \beta)^{-1} \sum_{s=1}^S p_s [f(A, R_s) - \gamma]^+. \quad (12)$$

It can be considered that $S = 5$, i.e., there are 5 trend combinations of gold and BTC (refer to Section 4.1). The above is the numerical solution of CVaR value.

Based on the above inversion, we can derive the value equation for multiperiod CVaR, which is used to describe the future risky asset trajectory of multiple investment products.

According to the expression for $\eta(k+m)$ after m instants, one of the trend combinations can be obtained. A new predicted value is obtained by adding a random component to $w(k+j+1)$. For each $j > 0$, at each moment, sample from a normal distribution $\Sigma^w = E[w(k+l+1)w^T(k+l+1)]$ with mean and covariance zero. Then, considering the emergence of S trend combinations, wealth value which is expected becomes a random process:

$$E[W(k+m)|F_k] = S^{-1} \sum_{s=1}^S E[W_S(k+j+1)|F_k]. \quad (13)$$

In the same trend combination, the solution of the CVaR value can also be solved:

$$E[\tilde{F}_\beta(A, \gamma)|F_k] = \gamma + [(1 - \beta)S]^{-1} \sum_{s=1}^S [-E[W_S(k+m)|F_k] - \gamma]^+. \quad (14)$$

4.2.4. Trading Commission Costs. Trading commissions are proportional costs, which are expressed as a percentage of the transaction amount. Therefore, we should deduct the commission from the expected value, and equations (12) and (14) should be modified as follows (α (%) is the commission percentage):



FIGURE 7: Total account investment income chart.

basket.” Using the greed algorithm, we set different caps for different investment styles. For example, an aggressive investor can allocate more Bitcoins (100% of his or her assets).

Considering the upper limit of the portfolio, i.e., $a_i^{\max}(k) = \delta W(k)$, $i = 1, 2$, we can express it as the following constraint:

$$\overline{M}(k)A(k) \leq A_{\max}(k). \quad (16)$$

Here, $A_{\max}(k) = (a_{\max}^T(k), \dots, a_{\max}^T(k+m-1))^T$, $A(k) = [a^T(k|k), \dots, a^T(k+m-1|k)]^T$, and $a_{\max}(k) = (a_1^{\max}(k), a_2^{\max}(k))$, with $M = \mathbf{I}_2$.

Also, the profit obtained from each trading operation must exceed the commission.

$$E[W(k+m)|F_k] - E[W(k)|F_k] > \sum_{j=0}^{m-1} \alpha \left[A(k+j) - (1 + E[R^T(k+j)|F_k])A(k+j-1) \right]. \quad (17)$$

4.2.6. Multiobjective Optimization. Taking the above inferences into account, the dual objective of “maximum return and minimum risk” is followed in the investment process. Then, the following multiobjective dynamic optimization can be listed based on the constraints mentioned above [26].

$$\begin{aligned} & \max_U E[W(k+m)|F_k], \\ & \min_U E[\tilde{F}_\beta(A, \gamma)|F_k] \text{ s.t. } \begin{cases} \overline{M}(k)A(k) \leq A_{\max}(k) \\ E[W(k+m)|F_k] - E[W(k)|F_k] > \Delta. \end{cases} \end{aligned} \quad (18)$$

4.3. Results of Our Model. We divide the traders into 4 categories: A1 (cautious), A3 (balanced), A4 (advanced), and A5 (aggressive). They have different risk-tolerance (CVaR).

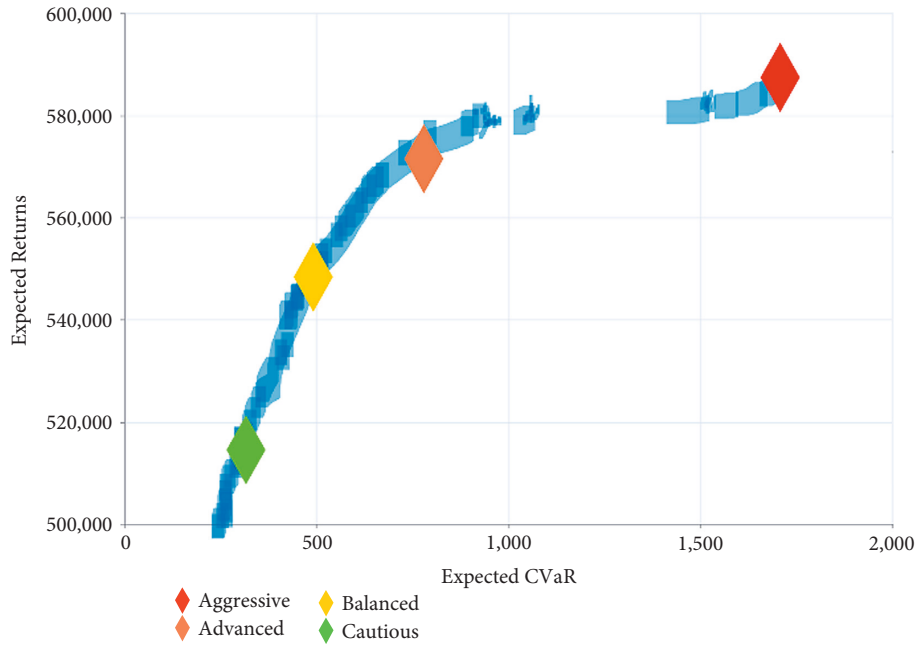


FIGURE 8: Pareto front returns, CVaR.

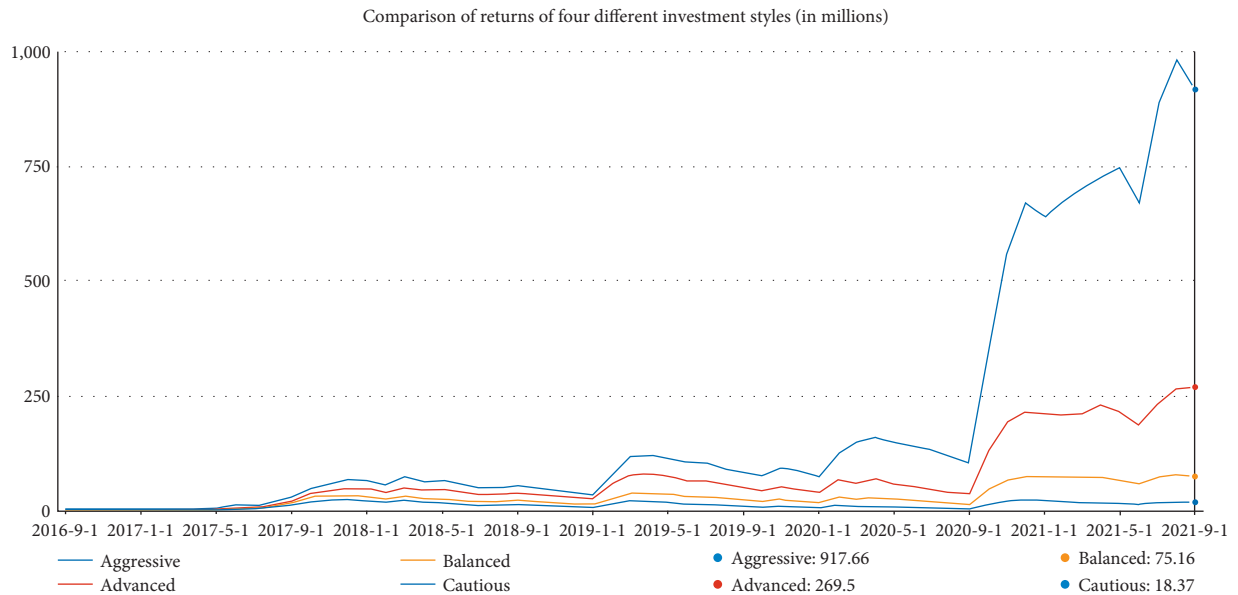


FIGURE 9: Return curves of four different investment styles.

4.3.1. *Maximum Return.* We first give the results for the aggressive investment style, an investor who places profitability above all factors.

As a result, a maximum return of \$ 917,658.45 is obtained with trading commission $\alpha_{gold} = 1\%$ and $\alpha_{BTC} = 2\%$. Here, α_{gold} represents gold commission rate, and α_{BTC} represents Bitcoin commission rate. The tracking process reveals that this strategy invested all the money (100%) in Bitcoin when it is in an uptrend and gained high returns. The following charts (Figure 7) show the return on total accounts and total assets.

Comparing the chart of Bitcoin, you can see that the return chart of this strategy is highly similar to the trend of

Bitcoin’s chart. This also shows that the account’s main return comes from the volatility of Bitcoin and is a high-risk, high-reward strategy. Conversely, the account has had huge retracements. The data show that the strategy has a maximum retracement of more than 52% in 2019.

4.3.2. *Comparison of Different Investment Styles-Pareto Frontier.* To determine the optimal control strategy, we use the multiobjective optimization method. In fact, the solution of the multiobjective model is a sequence of control actions called the nondominant or Pareto optimal set.

TABLE 1: Statistical characteristics of the data before and after processing.

Comparison parameters	Our model	Other model
Number of trades	133	235
Back-test time (same platform)	126.5s	62.7s
Percentage of profitable operations	79.43%	55.56%
Total yield	91765.85%	1940.15%
Max. continuous retracement	15.78%	29.7%

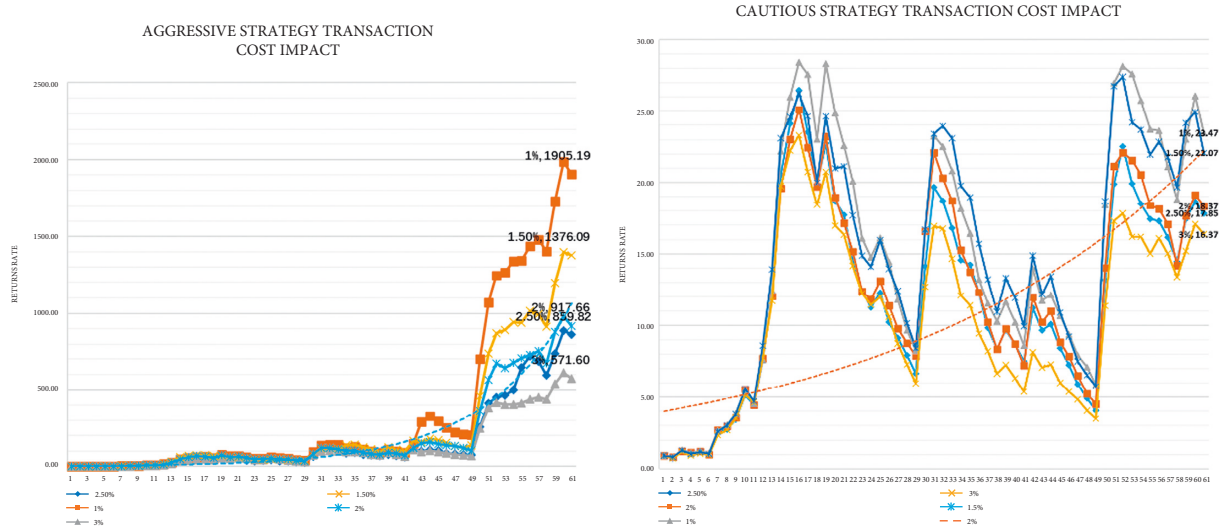


FIGURE 10: Impact of different transaction costs.

Mathematically, $J(A(k)) = [J_1(A(k)), J_2(A(k))]^T$ is a vector-valued function with 2-objective minimization and if no other feasible sequence of control actions $A(k)$ exists, the sequence $A^q(k) = (a^q(k), \dots, a^q(k+m-1))$ is called Pareto optimal. A Pareto front is constructed for the problem's solution. The Pareto solutions are given by $P = \{A^1(k), \dots, A^S(k)\}$, ($S = 5$). Thus, we can select the best Pareto solution for the investor. For instance, the investor can tolerate high risk in order to aim for higher returns, or otherwise, he is just conservative and can only accept low risk. Figure 8 shows the Pareto front charts obtained after the first step of the multiobjective optimization instant.

Figure 9 shows the return curves for the 4 different investment styles with the same trading commissions. Apparently, returns and risks are always proportional.

5. Sensitivity Analysis of Our Model

5.1. Analysis of the Superiority of Our Model. In fact, our multiobjective (return-risk) model has a higher return under an aggressive investment style if we do not apply predictive data but use real data. Nevertheless, this is already very good, because, after all, we cannot have a god's perspective in real trading.

We also back-test our model against a traditional trend indicator strategy. We have also back-tested the model against a traditional trend indicator strategy. We use a portfolio strategy model based on price forecasting and trend indicators, which is relatively popular in the JoinQuant

Quantitative Community, for comparison. The results are shown in Table 1.

In summary, our quantitative strategy has better results than the common strategies.

5.2. Analysis of the Impact of Transaction Costs. Based on common sense, we believe that transaction costs are also an important factor in the returns of a strategy. After our model validation, we find that the lower the transaction cost, the higher the strategy return. Figure 10 shows the impact of different transaction costs on the return curve for aggressive and cautious investment styles.

The data in Figure 10 are obtained when we adjust the commission percentage for Bitcoin and leave the percentage for gold unchanged. This is based on the idea of greed, where getting a bigger return requires investing more Bitcoins, and therefore, the impact of Bitcoin transaction costs on total returns is greater. We control for variables to see the results of running the model, and this is in line with our common sense.

6. Conclusions

In this paper, we build a quantitative model to advise on trading based on the price movement of Bitcoin and gold between 2016 and 2021. We mainly use greedy strategies with multiobjective optimization to balance risk and return in trading. Such model end up with a maximum ending return of 91765.85% (aggressive) for this portfolio after 5

years of quantitative trading, with an annualized return of 291.32% and ending assets of \$ 917,658.45 (opening assets are \$1,000).

The strengths of our work are as follows: (1) to obtain the price trends in the medium and long term, we integrate some common trend indicators with the predicted prices to further improve the correctness of the obtained price trends. (2) Combining greedy strategies with multiobjective optimization to minimize risk and maximize return. (3) Pareto frontier is provided for investors with different preferences. (4) The model can be applicable to various investment situations (*Note*. The model is universal because the portfolios in multiobjective optimization in Section 4 can be extended to N , which only requires adding dimensionality to some vectors and matrices.)

The weaknesses of our work are as follows: (1) the model is not well trained, as only 5 years of price data are used. (2) We filter out the high-frequency price fluctuations in the short term and only considered the medium and long-term price trends, thus reducing the returns to some extent. (3) Some of the models simply use greedy algorithms, which only consider the possible local optimal solutions, but not necessarily the overall optimal solutions.

In short, although our model has a considerable degree of applicability and extremely promising returns, the investment industry is ever-changing and no model can predict all situations and suggest the perfect trading scenario. One needs to be very careful when entering the investment industry and must be well prepared. As Warren Buffett said, "Obviously, every investor will make mistakes. A reasonably intelligent, informed, and diligent person can judge investment risks with a useful degree of accuracy." If you cannot control your emotions and are easily tempted by price fluctuations, then leave the trading to the computer! The future of financial markets will be a battleground between quantitative trading strategies based on objective data analysis and forecasting.

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 declare that they have no conflicts of interest regarding this work.

References

- [1] H. M. Markowitz, *Portfolio Selection Efficient Diversification of Investment*, Oxford UK Blackwell, Cambridge, MA, USA, 1990.
- [2] H. Konno, "Yamazaki mean-absolute deviation portfolio optimization model and its application to Tokyo stock market," *Management Science*, vol. 37, no. 5, pp. 519–531, 1991.
- [3] G. Delino and M. Fedele, "C meloni. DOAM for evolutionary portfolio optimization: a computational study," *Applied New Economics Paper*, pp. 253–266, 2008.
- [4] P. J. Mercurio, Y. Wu, and H. Xie, "Option portfolio selection with generalized entropic portfolio optimization," *Entropy*, vol. 22, no. 8, p. 805, 2020.
- [5] P. J. Mercurio, Y. Wu, and H. Xie, "An entropy-based approach to portfolio optimization," *Entropy*, vol. 22, no. 3, p. 332, 2020.
- [6] Y. Zhou, *Research on Portfolio Problem Modeling and Multi-Objective Evolutionary Algorithm*, Guangdong University of Technology, Guangzhou, China, 2013.
- [7] X. Gong, L. Min, and C. Yu, "Multi-period portfolio selection under the coherent fuzzy environment with dynamic risk-tolerance and expected-return levels," *Applied Soft Computing*, vol. 114, Article ID 108104, 2022.
- [8] L. Yin, "Research on the influence of high-frequency price data on dynamic risk measurement of portfolio -- Based on the analysis of realization covariance matrix model," *Financial Development Research*, no. 7, pp. 3–8, 2016.
- [9] J. Poljašević and M. Grujić, "Portfolio optimization with investment in cryptocurrencies," *Lecture Notes in Networks and Systems*, vol. 315, pp. 35–47, 2022.
- [10] B. A. Mercangöz, "Portfolio optimization," *International Series in Operations Research and Management Science*, vol. 306, pp. 15–27, 2021.
- [11] M. I. Boloş, I. A. Bradea, and C. Delcea, "Optimization of financial asset neurosophic portfolios," *Mathematics*, vol. 9, no. 11, p. 1162, 2021.
- [12] L. H. Pedersen, A. Babu, and A. Levine, "Enhanced portfolio optimization," *Financial Analysts Journal*, vol. 77, no. 2, pp. 124–151, 2021.
- [13] L. Martínez-Nieto, F. Fernández-Navarro, and M. Carbonero-Ruz, "An experimental study on diversification in portfolio optimization," *Expert Systems with Applications*, vol. 181, pp. 0957–4174, 2021.
- [14] W. J. Klerk, W. Kanning, M. Kok, and R. Wolfert, "Optimal planning of flood defence system reinforcements using a greedy search algorithm," *Reliability Engineering & System Safety*, vol. 207, Article ID 107344, 2021.
- [15] R. Mendoza-Gómez, R. Z. Ríos-Mercado, and K. B. Valenzuela-Ocaña, "An iterated greedy algorithm with variable neighborhood descent for the planning of specialized diagnostic services in a segmented healthcare system," *Journal of Industrial and Management Optimization*, vol. 16, no. 2, pp. 857–885, 2020.
- [16] S. Balakrishnan and S. Doraiswamy, "Efficient virtual data center request embedding based on row-epitaxial and batched greedy algorithms," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 27, no. 2, pp. 780–794, 2019.
- [17] D. Xu, Y. Li, T. Xia, J. Li, and P. Hui, "Portfolio optimization in traffic offloading: concept, model, and algorithms," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 691–706, 2021.
- [18] X. Xiong, P. Zhou, Y. Yin, T. C. E. Cheng, and D. Li, "An exact branch-and-price algorithm for multitasking scheduling on unrelated parallel machines," *Naval Research Logistics*, vol. 66, no. 6, pp. 502–516, 2019.
- [19] K. P. Anagnostopoulos, P. D. Chatzoglou, and S. Katsavounis, "A reactive greedy randomized adaptive search procedure for a mixed integer portfolio optimization problem," *Managerial Finance*, vol. 36, no. 12, pp. 1057–1065, 2010.
- [20] C. Tang, *Improvement of PSO Algorithm and its Application in Constrained Multi-Objective*, South China University of Technology, Guangzhou, China, 2008.

- [21] S. Khuller, A. Moss, and J. Naor, "The budgeted maximum coverage problem," *Information Processing Letters*, vol. 70, no. 1, pp. 39–45, 1999.
- [22] S. Zhang and S. He, "Approximate algorithm for maximum coverage problem of budget type," *Journal of Hebei University (Natural Science Edition)*, vol. 28, no. 1, pp. 7–9, 2008.
- [23] A. Arulsevan, "A note on the set union knapsack problem," *Discrete Applied Mathematics*, vol. 169, no. 41, pp. 214–218, 2014.
- [24] Y. Li, N. H. Chan, C. Y. Yau, and R. Zhang, "Group orthogonal greedy algorithm for change-point estimation of multivariate time series," *Journal of Statistical Planning and Inference*, vol. 212, pp. 14–33, 2021.
- [25] M. K. De Melo, R. T. N. Cardoso, and T. A. Jesus, "Multi-objective dynamic optimization of investment portfolio based on model predictive control," *SIAM Journal on Control and Optimization*, vol. 60, no. 1, pp. 104–123, 2022.
- [26] V. Dombrovskii and T. Obedko, "Feedback predictive control strategies for investment in the financial market with serially correlated returns subject to constraints and trading costs," *Optimal Control Applications and Methods*, vol. 38, no. 6, pp. 908–921, 2017.

Research Article

Design and Implementation of Chinese Common Braille Translation System Integrating Braille Word Segmentation and Concatenation Rules

Ju-Xiao Zhang ¹, Hai-Feng Chen,² Bing Chen,³ Bei-Qin Chen ³, Jing-Hua Zhong,⁴
and Xiao-Qin Zeng⁵

¹College of Information & Mathematics Science, Nanjing Normal University of Special Education, Nanjing 210038, Jiangsu, China

²Nanjing Dian-Ming Software Technology Co, Ltd, Nanjing 210038, Jiangsu, China

³China Braille & Sign Language Research & Application Center, Nanjing Normal University of Special Education, Nanjing 210038, Jiangsu, China

⁴College of Special Education, Beijing Union University, Beijing 100075, China

⁵College of Computer and Information, Hohai University, Nanjing 210038, Jiangsu, China

Correspondence should be addressed to Bei-Qin Chen; cbq@njts.edu.cn

Received 30 March 2022; Accepted 27 April 2022; Published 17 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Ju-Xiao 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.

An important sign of the accessibility of Braille information is the realization of the mutual translation between Chinese and the Braille. Due to the irregularity and uncertainty of the Prevailing Mandarin Braille, coupled with the lack of a large-scale Braille corpus, the quality of Chinese-Braille translation seems to be poor. In July 2018, the National Language Commission released the “Chinese Common Braille Scheme” and advocated replacing the “Prevailing Mandarin Braille.” Aimed at improving translation accuracy, this research, which is based on the self-built Chinese Common Braille corpus and combined with the HanLP (Han Language Processing) dictionary and the Chinese-Braille word corpus (a Braille word segmentation and concatenation dictionary for generating a unigram language model), uses the n-gram language model to design and implement a Chinese-Braille intertranslation system that integrates Chinese and Braille Word Segmentation and Concatenation Rules. More importantly, this research proposes an experimental plan for improving the Braille Word Segmentation and Concatenation Rules using a Chinese-Braille word corpus. Experiments show that in the field of educational literature, the accuracy rate of translation from Chinese to Chinese Common Braille has reached 95.01%, and the accuracy of Chinese Common Braille to Chinese translation has reached 90.15%.

1. Introduction

An important sign of the accessibility of Braille information is the mutual translation between Chinese and Braille so that no significant differences can be found between the original and translated Chinese characters on smart devices. Braille is a special script with the properties of the host. Braille generally does not exist independently (there is no Braille used in a country that can be separated from a certain language, and there is no Braille that is used internationally across languages), and there are both associations and differences with the host language. The appearance of Braille is

the same all over the world, but the difference in the host language makes the Braille of the corresponding language completely different from others. The informatization of English Braille is easy to complete, and the level of informatization is also high, so that blind people who use English can be well educated, which helps to promote social equity and to achieve great social significance.

At present, the Braille that bonds with the Chinese is collectively referred to as “Chinese-Braille,” and there are three main types, Prevailing Mandarin Braille, the double spelling Braille, and the Chinese Common Braille Scheme in 2018. The use of double spelling Braille is less often, and now

the Prevailing Mandarin Braille is mainly used, and the use of the Chinese Common Braille is gradually promoted [1]. The Prevailing Mandarin Braille at most uses the three-cells Braille (initial, final, and tone, respectively) to represent a Chinese character and suffers from the following problems:

- (1) The general principle of tone is “generally not to mark the tone, but only marked when necessary” [2], which makes Braille expression rely on expert experience. Besides, there are principles but no norms. In particular, the understanding of the homophones in Chinese itself, no matter with marked tones or not, has to rely on “guessing” and therefore the ambiguity is increased.
- (2) Word segmentation and concatenation rules are not yet perfect. Unlike “characters” and “characters” that are not separated in Chinese, Braille draws on English word segmentation rules (Braille is similar to Pinyin in essence) by adding “blank cells” (or spaces) between words to reduce ambiguity. Braille word segmentation is not only based on semantics but also considers the tactile problem of “touching and reading” for blind people (reducing blank cells and improving reading speed). Therefore, it is necessary to concatenate words that are originally semantically separated, which is called Braille Word Segmentation and Concatenation Rules. For example, “引/无数/英雄/竞/折腰.” There are about 100 rules in the Braille Word Segmentation and Concatenation Rules, which are still not perfect, and they are often done manually by Braille experts.

The above-mentioned irregularities and uncertainties hinder the translation of Chinese to Braille and make it difficult to improve the Braille informatization at a certain level. Researchers have been looking for breakthroughs for many years, but the results are not obvious. This is also the original intention of the nation to promulgate and promote the Chinese Common Braille.

The Chinese Common Braille inherits the Prevailing Mandarin Braille as a whole and “upgrades” the Prevailing Mandarin Braille to some extent. Mainly, all characters are marked with tones, and the tones are written according to the initials. This not only reduces the randomness of Pinyin tone but also reduces the number of cells in Braille [1]. For example, the Pinyin of “更加” is “gèngjiā”

while the Prevailing Mandarin Braille is

without marking the tones for both characters. However, the characters with marked tones should be

●● ○● ○○ ●● ●● ●○
 ●● ○● ●○ ●● ●○ ○○. The Chinese Common Braille
 ○○ ●● ●○ ○○ ○● ○○
 rule is that when the initial consonant is “g,” the tone is omitted so that the Common Braille should be

●● ○● ●● ●● ●○
 ●● ○● ●● ●○ ○○.
 ○○ ●● ○○ ○● ○○

1.1. Research Background. Some scholars have carried out research on the basis of Prevailing Mandarin Braille, such as

Zhou et al. [3] and Zhu et al. [4–6], starting from the rules, segmenting words according to Chinese semantics, and then converting Chinese words into Braille words. The accuracy from Chinese to Braille is high, but the accuracy from Braille to Chinese is not satisfactory. To solve the problem of word segmentation and concatenation, some scholars built a small-scale Braille word segmentation database and use the Trie tree to process word segmentation and concatenation [7]. Some are first-word frequency and word grading weighted word segmentation and then the combination of rules and statistics [8]. Some followed the Chinese-Pinyin-Braille conversion manner [9]. Some scholars used the Markov model to identify Chinese characters, and then the reverse maximum matching word segmentation method is used to segment Chinese words; scholars such as Wang Xiangdong combined the Chinese-Braille word segmentation, Braille Word Segmentation, and Concatenation Rules and tone information for higher translation accuracy [10]. Furthermore, there are also methods based on statistical machine learning to realize Chinese-Braille machine translation [11]. Another work proposes an algorithm that integrates Chinese word segmentation and Braille word segmentation [12, 13] to improve the accuracy. A deep learning-based technique that trains a bidirectional LSTM model achieved a word segmentation accuracy rate of 94.42% [14]. However, the Prevailing Mandarin Braille corpus is self-made, and no in-depth research has been carried out on the Chinese Common Braille. The China Disabled Persons’ Federation project “Development of Common Braille Automatic Translation Software” hosted by Professor Xiao Hang developed a Common Braille automatic translation software by adopting a language model combining N-grams and maximum entropy, and good translation results have been achieved [15].

1.2. Outlines and Contributions. The remainder of this paper is arranged as follows. In Section 2, we will introduce the detailed information of the Chinese-Braille word corpus provided by the team of Professor Jing-Hua Zhong of Beijing Union University as we carried out our research based on this dataset. We will present the detailed procedures of our developed system in Section 3, where we will firstly introduce the details of translation from Chinese to Braille and vice versa. We then proposed our method for Braille word segmentation and concatenation. Specifically, we firstly deploy the n-gram model to segment Chinese words and then train a Braille word segmentation concatenation dictionary to generate a unigram language model for Braille word segmentation adjustment. By doing so, we aimed at improving the translation accuracy from Chinese to Braille. Finally, we improved the Braille Word Segmentation and Concatenation Rules by experiment. The details of the experiment are presented in Section 5. In terms of translation from Chinese Common Braille in the field of educational literature, the accuracy has reached 95.01% while the translation accuracy reached 90.15% when translating Chinese Common Braille to Chinese.

TABLE 1: Field of Chinese-Braille word corpus (parts).

Corpus field	Explanation	Example
BSCR	Braille Word Segmentation and Concatenation Rules	CD
u2v	Chinese Pinyin, <i>u</i> to <i>v</i>	geng4/jia1
Pinyin	The Pinyin's form of Chinese Common Braille. The abbreviation rules of tone	geng/jia1
Initials and finals	Initials and finals of Braille	geng/jia1
ASCII's of Braille	The ASCII's form of Braille	G#G\$A
Common Braille	The Unicode's form of Chinese Common Braille	⠠⠠⠠⠠

TABLE 2: A list of word segmentation and concatenation marks (parts).

Index	Classification	Symbol	Word frequency	Sample
1	Noun	N	6897	人、品德、蓝色、国家、思路、心胸、机器、阿胶、思想家、阿司匹林
	Noun of location	Nf	277	上、下、左、右、东、西、里面、南、北、之中
	Place NOUN	Nd	2587	日本国、韩国、北京、北京市、夏威夷、唐古拉山、三峡、加州、华北
	Personal noun	Nr	1840	唐太宗、马克思、豆豆、鲁迅、黑旋风、卡尔·马龙、小布什
	Other proper nouns	Nz	1639	北京大学、英语、抗日战争、肯德基、左传、诺贝尔奖、和谐号、长城饭店、海豹突击队、开国大典
2	Verb	V	15196	走、爱、调查、同意、喜爱、包含、跨
	Directional verb	V		来、回、上去、下来、出去、进来、出来
3	Adjective	A	3863	美丽、丑陋、雪白、公共、皑皑、金灿灿
4	Numeral	M	7041	一、一百五十一、第一、一百零八、百分之十
5	Classifier	Q	82	个、沓、千克、架次、册、吨、朵、光年、赫兹
6	Adverb	D	4669	很、必、已经、处处、单独、倍加、必定、不妨
	Special adverb		-	不 _{Dbu}
7	Pronoun	R	1859	我、我们、她、这个、什么、怎样、这么、谁、哪里

2. Material

2.1. Chinese-Braille Word Corpus. The team of Professor Jing-Hua Zhong of Beijing Union University has undertaken the National Social Science Fund major project "Research on the Construction of Chinese-Braille Corpus." With the authorization of Professor Jing-Hua Zhong, this study uses the Chinese-Braille word corpus provided by him. We automatically extract words from literature, science, and other books to make a Chinese-Braille word corpus the word corpus and then manually reviewed and revised them. The corpus contains the information shown in Table 1.

This corpus is a Chinese-Braille word dictionary. The corpus contains the words extracted from the corpus text, and a mapping relationship is established.

2.2. Braille Word Segmentation and Concatenation Rules. The word segmentation and concatenation rules can be seen in Table 1 which lists the refinement and annotation of existing word segmentation and concatenation rules. The annotations are slightly different from the Chinese corpus, as shown in Table 2, which was made by Jing-Hua Zhong's team.

2.3. Construction of Braille Word Segmentation and Concatenation Dictionary. According to the Chinese-Braille word corpus, a Braille word segmentation and concatenation dictionary is established. In particular, the existing corpus is only a word corpus, and only a Braille unigram language

TABLE 3: Statistical information of Braille word segmentation and concatenation corpus.

The sum of word frequencies in the corpus	229551
Number of words	31708
Number of Common Braille with Pinyin	1477
The sum of Common Braille word frequencies in the same Pinyin	52762
The sum of the highest word frequencies of Common Braille words in the same Pinyin	46235
The number of Chinese characters in Common Braille with the same Pinyin	3491

model dictionary is constructed here. In this dictionary, the core is to count the frequency of words.

The number of word frequencies in Table 2 equals the sum of the number of frequencies of this type of word segmentation. "-" means unsubdivided statistics. More specifically, the numbers and punctuation marks in the Chinese-Braille word corpus are removed, only Chinese words are retained, the word frequency of each Braille word is counted, and it is stored as a Braille unigram language model dictionary, as shown in Table 3.

There are a total of 229,551 words in the corpus, 31,708 of which are the Chinese Common Braille Pinyin, 1,477 of which correspond to two or more different Chinese words, forming a total of 3,491 Chinese words, 1,477 of which have the same Pinyin and a total of 46,235 Chinese Common Braille words. The sum of the highest frequency of each Pinyin word is 46,235.

Table 4 shows the fragment content of the unigram Braille word segmentation and concatenation dictionary.

TABLE 4: Fragments of Braille word segmentation and concatenation dictionary.

Pinyin	Chinese words	Braille Word Segmentation and Concatenation Rules	ASCII code of Braille	Word frequency
bu4/shi	不是	DbuV/JC	BU:	276
bu4/shi	不事	DbuV	BU:	1
bu4/shi	不适	CD	BU:	4
bu4/shi	不释	I2	BU:	2
bu4/shi2	不时	CD	BU:1	23
bu4/shi2	不识	DbuV	BU:1	2
bu4/shi2/pu	不识谱	DbuVN	BU:1PU'	1

The first column is the Chinese Common Braille Pinyin, the second column is the Chinese words connected according to the Braille word segmentation, the third column is the word segmentation type, the fourth column is the Braille ASCII Pinyin, and the last column is the word frequency.

3. Design of Chinese-Braille Translation System

The release of the Chinese Common Braille has fundamentally solved the problem of Braille marking. Therefore, in the translation process from Chinese to Braille, the core difficulty is the Braille Word Segmentation and Concatenation Rules [12].

The main methods of translation from Chinese to Braille are as follows:

- (1) Formally describe the concatenation rules of word segmentation [5]. The emphasis is on the formal description of word segmentation and concatenation rules and their application to word segmentation in Braille Pinyin. However, the word segmentation rules are not perfect, and new rules are still generating, making it more difficult for this method to improve the accuracy of word segmentation.
- (2) Extract word segmentation and concatenation from Braille corpus, use the word segmentation library to segment Chinese words, then implement pinyinization after word segmentation, and then convert to Braille. This method relies on an accurate and large-scale Braille corpus, and the Chinese Common Braille corpus is still continuous improvement [9].
- (3) Directly train word segmentation and concatenation rules from Chinese and Braille's dictionary (unlabelled corpora) through machine learning methods [11]. The Chinese word segmentation is combined with the Braille word segmentation, and the Braille Word Segmentation and Concatenation Rules adjusts the Chinese word segmentation to obtain the final Braille word segmentation, and the final translation accuracy is improved by 3.56% [12]. Satisfactory results were achieved without using the Chinese Common Braille corpus.

Based on the Chinese-Braille word corpus, combined with Method 2 and Method 3, the mutual translation between Chinese and Chinese Common Braille can achieve better results. Firstly, Chinese sentences are organized according to the Braille Word Segmentation and Concatenation Rules;

secondly, Chinese words in Chinese Pinyin are marked (especially the words that are changed in the sentence must be in the same tone); finally, according to the Chinese Common Braille consonant representation rules, the Pinyin syllables are converted into Braille to achieve the "Braille" of Chinese Pinyin and other characters.

In the translation process, there are also some details that need to be coped with. For example, according to the representation rules of the Chinese Common Braille, the comparison table between Chinese Pinyin and the Chinese Common Braille, numerous tables have to be made, such as the comparison table between the numerical symbols and the Chinese Common Braille, the comparison table between the English letters and the Chinese Common Braille, and the comparison table between various symbols and the Chinese Common Braille. When brailing various characters, firstly, it is necessary to solve the problem of adding prompt symbols when Chinese, English, various numbers, symbols, and other text symbols are mixed; secondly, it is necessary to solve various noncontent symbols (such as hyphens), format prompts, etc.

3.1. Chinese-Braille Translation. The Chinese-Braille translation is just the direct translation from Chinese characters into Braille. The sentence can contain non-Chinese characters such as numbers, English, and punctuation marks, but there is a one-to-one correspondence between such information and Braille, where direct conversion can be applied. The accuracy, however, is very high already and will not be discussed here.

The biggest technical problem in Chinese-Braille translation is the Braille Word Segmentation and Concatenation Rules. With the simultaneous development of machine translation technology, Chinese word segmentation technology methods are divided into rule-based, statistical-based, and deep learning [16]. The rule-based method mainly depends on the dictionary, and the more classic ones are the maximum matching. The advantage is that it is simple, easy, and efficient; the disadvantage is that it depends on the accuracy and scale of the dictionary, and the recognition ability of ambiguous words is poor. The statistical methods rely on the corpus and use the corpus context information, word frequency, information entropy, etc., to perform word segmentation. The more classic ones are based on n-gram, maximum entropy model, hidden Markov model (HMMs) [17], conditional random field model (CRF), and so on. The advantage is that the accuracy rate is significantly improved, and the disadvantage is that it relies on a relatively large-scale corpus, which causes bias problems (maximum entropy

model, hidden Markov model) or model complexity and low efficiency (CRF). The N-shortest tokenizer has a better effect than the shortest tokenizer and has a stronger ability to recognize named entities, but the speed is much slower; the CRF (conditional random field) methods usually have an accuracy and recall rate that are higher than 96% and have good new word recognition ability. In recent years, neural networks and deep learning methods are very eye-catching techniques [18]. Deep learning has made outstanding achievements in natural language processing, image recognition (especially medical image recognition [19, 20]), target detection, and so on. Deep learning includes Convolutional Neural Networks (CNNs) to Recurrent Neural Networks (RNNs) to long- and short-term memory neural network model LSTM [21] and improved variations of LSTM. The advantage is that the effect is good, and the disadvantage is that it relies on a large-scale corpus, long training time, and poor interpretability. Of course, there is also a combination of the above-mentioned methods to achieve good results and high efficiency as much as possible.

At present, there is only a Chinese-Braille vocabulary corpus, and there is a lack of a large number of well-known Braille corpora that have been correctly segmented and marked. We, therefore, proposed making full use of the existing Chinese dictionary and word segmentation system, which is conducive to the accuracy of Chinese-Braille translation. Based on the Chinese-Braille vocabulary corpus + HanLP dictionary, this research will use the n-gram language model to segment Chinese sentences and then use the Braille word segmentation normative dictionary to improve the accuracy of word segmentation, thereby improving the accuracy of Chinese-Braille translation.

As shown in Figure 1, the process of Chinese to Chinese Common Braille translation based on the Chinese-Braille word corpus is given. Starting from a Chinese sentence, we firstly use n-gram for training word segmentation (combining HanLP dictionary with Chinese-Braille vocabulary corpus) and then convert Chinese into Pinyin strings; then based on the Chinese-Braille vocabulary corpus, it is converted into a Chinese Common Braille Pinyin string, which fuses word segmentation and concatenation rules in Chinese-Braille word corpus. Finally, the Braille Pinyin string obtained after word segmentation is converted into ASCII Pinyin symbols and Braille symbols.

3.1.1. Chinese Word Segmentation Using n-Gram Language Model [22, 23]. Suppose that the sentence $S = c_1 c_2 \dots c_N$ needs to be divided into Chinese words, and $c_i (1 \leq i \leq N)$ is a single Chinese character. The result after Chinese word segmentation $R = w_1 w_2 \dots w_M$, ($1 \leq M \leq N$).

Assuming that during Chinese word segmentation, the probability of occurrence of the $w_j (1 \leq j \leq M)$ is related to all the previous words, that is, related to the previous $j - 1$ words [24], then the j-gram language model can be expressed as

$$P(w_j | w_1^{j-1}) (j > 1). \quad (1)$$

It can be known from the Bayesian formula that

$$P(w_j | w_1^{j-1}) = \frac{P(w_j)}{P(w_1^{j-1})}. \quad (2)$$

The probability of word w_j can be calculated from

$$P(w_j | w_1, w_2, \dots, w_{j-1}) = P(w_j | w_{j_0+1}, w_{j_0+2}, \dots, w_{j-1}), \quad (3)$$

where $j_0 = \max(j - n, 0)$.

Given a sentence consisting of a sequence of m words $R = (w_1, w_2, \dots, w_M)$, ($1 \leq M \leq N$) that can be denoted as $w_1 w_2 \dots w_M$, and its probability is (4):

$$P(R) = P(w_1) \prod_2^M P(w_j | w_{j_0+1}, w_{j_0+2}, \dots, w_{j-1}). \quad (4)$$

It can be seen from the formula that the probability of forming a sentence is the product of the word probabilities of Chinese word segmentation. Theoretically, the larger the value of n , the better the segmentation effect, but the larger the value of n , the greater the amount of calculation, so that it cannot be calculated; in addition, the sparsity of the data is serious.

The magnitude of the model parameters is an exponential function (N^n) of the model length n , so n cannot be very large. For a word corpus with sufficiently large sample size, the probability can be calculated using the word frequency approximation:

$$P(w_j | w_1^{j-1}) = \frac{P(w_j)}{P(w_{j-n+1}^{j-1})} \approx \frac{\text{count}(w_j)}{\text{count}(w_{j-n+1}^{j-1})}, \quad (5)$$

where $\text{count}(w_j)$ and $\text{count}(w_1^{j-1})$, respectively, represent the number of occurrences of w_j and w_1^{j-1} in the corpus.

According to Markov's hypothesis, the possibility of several words forming a word only depends on one or a limited number of words that appear before it and has nothing to do with the following words. The current value of n is generally 2 or 3.

If it only depends on a word that appears before it, it is called a 2-gram. And if it only depends on the two words that appear before it, it is called a 3-gram.

When $n = 2$, (5) becomes

$$P(w_j | w_1^{j-1}) \approx \frac{\text{count}(w_{j-1}, w_j)}{\text{count}(w_{j-1})}. \quad (6)$$

3.1.2. Data Smoothing Algorithms [25]. There are two serious problems in the approximate calculation of formula (6): the first one is that the probability of words that do not appear is approximated to 0 (data sparsity); the second one is when $\text{count}(w_{j-1}, w_j) = \text{count}(w_1^{j-1})$, then $P(w_j | w_1^{j-1}) = 1$.

Therefore, the data needs to be smoothed. Data smoothing is to appropriately reduce the probability of each word in the sample and adjust the reduced probability value to the words that do not appear so that the probability of all words is not 0 and the sum of the probabilities is still equal to 1. Commonly used data smoothing techniques are Add-delta smoothing, Good-Turing smoothing, combined estimation, simple linear

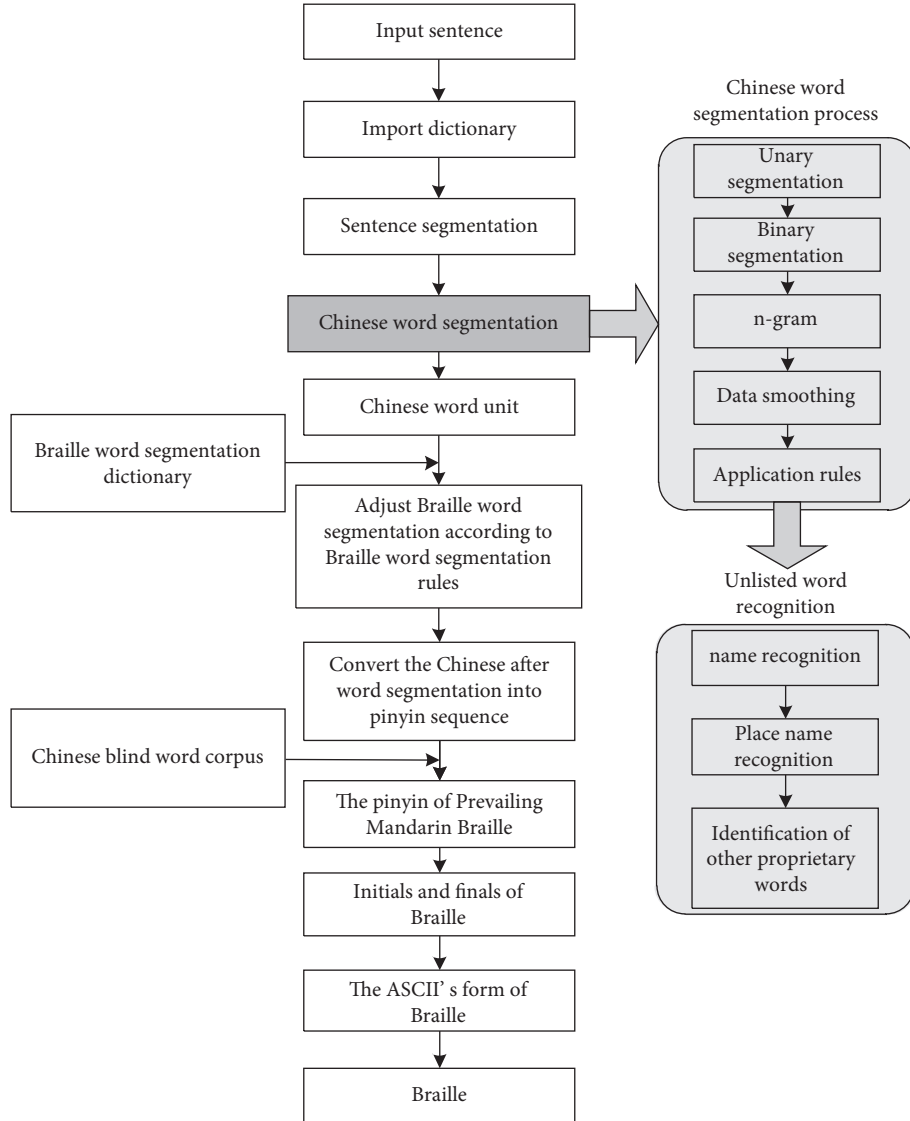


FIGURE 1: The translation process of Chinese-Common Braille.

interpolation, Jelinek-Mercer smoothing, fallback model and Katz smoothing, etc.

This study employs Good-Turing smoothing, which is suitable for large vocabularies to generate multimodal distributions of observations.

N is the size of the original training sample data, and n_x is the number of words that appear x times in the training sample. Then,

$$\begin{aligned}
 N &= \sum_{x=1}^{\infty} n_x x, \\
 N &= \sum_{x=0}^{\infty} n_x x^* = \sum_{x=0}^{\infty} (x+1)n_{x+1}, \\
 x^* &= (x+1) \frac{n_{x+1}}{n_x}.
 \end{aligned} \tag{7}$$

Then, the probability of the word appearing x times in the sample is

$$P_x = \frac{x^*}{N}. \tag{8}$$

Finally, the probability normalization of all words is processed:

$$\widehat{P}_x = \frac{P_x}{\sum P_x}. \tag{9}$$

3.1.3. Chinese Word Segmentation Based on the 2-Gram Model of Word Segmentation and Concatenation Rules. As shown in Figure 1, assuming a Chinese sentence $C = c_1 c_2, \dots, c_n$, C is a sentence, and c_i is a character, based on the Chinese-Braille vocabulary corpus + HanLP dictionary, the specific process is as follows:

- (1) Divide characters: enter Chinese sentences and divide all words of the sentence into independent words. Spaces can be added after each word to

一双鞋	nz	19
一反常态	vl	20
一发	d	143
一发千钧	i	1
一口价	nz	40
一口咬定	vl	24
一口气	d	224
一口茶	nz	1

FIGURE 2: Fragments of unigram model dictionary.

提振@下	2
提振@了	32
提振@人民	2
提振@企业	2
提振@作用	6
提振@信心	12

FIGURE 3: Fragments of bigram model dictionary.

distinguish Chinese, English, and punctuation marks.

- (2) Unary segmentation: use the unary language model dictionary (CoreNatureDictionary) in the dictionary and the maximum matching algorithm to match the characters with the dictionary words, and form a unary word network containing information such as part of speech, word frequency, etc. Combine English characters and numeric characters into atomic words, and words are represented by w , that is, $w_i = c_j \dots c_k$, ($j, k \geq 1$).

The fragment content of the HanLP unary language model dictionary is shown in Figure 2.

The first column is the word, the second column is the type of the word, and the third column is the word frequency; if there is a second type of the word, the fourth column is the second type, the fifth column is the word frequency of this type, and so on.

- (3) Binary segmentation: according to the binary dictionary (CoreNatureDictionary.ngram), we continue to use the maximum matching algorithm to form a word graph (Figure 3), use @ to separate two words, and the probability of appearing as a common word, such as the word after “提振(boost)” is “信心(confidence)” which appears 12 times.
- (4) Chinese word segmentation of the 2-gram model. Use the two dictionaries in HanLP (CoreNatureDictionary and CoreNatureDictionary.ngram). The maximum forward and backward algorithms can be used to segment the sentence to obtain two strings s_1 and s_2 , respectively; if two different word sequences are obtained, the one with the highest probability is selected according to the bigram, which can eliminate part of the ambiguity.
- (5) Apply the above results and apply the rules to identify the spatial nouns.
- (6) Based on a name recognition dictionary, place name dictionary, and proper noun dictionary, use a two-layer HMM (Hidden Markov Model). Taking the word sequence as the observation sequence and the dictionary word probability sequence as the hidden state sequence, when the model predicts the best-hidden state sequence, the Viterbi algorithm is used to identify and match the names of people and places.

- (7) In the above steps, the Chinese sentences are segmented into Chinese words, and the result is $\mathbf{R} = \mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_M$, ($1 \leq M \leq N$).
- (8) Adjust word segmentation based on word segmentation and concatenation rules.

Based on the Chinese-Braille word corpus, the Chinese word segmentation is adjusted according to the Braille Word Segmentation and Concatenation Rules. The length of the word segmentation should be moderate, not too long, or too short. If it is too long and lacks a gap, the touch will easily cause fatigue and affect the effect of “touching and reading;” if it is too short, it is inconvenient to quickly form a concept and affect the speed of “touching and reading.” Because of this, the Braille Word Segmentation and Concatenation Rules has been developed in Braille, which is related to and different from Chinese word partitioning. For example, in the word “蒸馒头(steam steamed buns),” the Chinese segmentation is “蒸/馒头(steam/steamed buns),” and the word in Braille is a monosyllabic verb modifying a two-syllable noun, which needs to be written together. Therefore, Braille word segmentation is more coarse-grained than Chinese word segmentation.

Using the Chinese-Braille vocabulary corpus, a Braille word segmentation dictionary is established, and a unigram language model Braille word segmentation dictionary is obtained. Assume that $\mathbf{L} = \mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_r$, ($1 \leq r \leq N$) is a Braille word segmentation dictionary, and the maximum matching algorithm is used to segment the Chinese sentence C . Because there is no binary language model, its ambiguity is difficult to eliminate.

Using the 2-gram model for Chinese word segmentation, the result is $\mathbf{R} = \mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_M$, ($1 \leq M \leq N$), and \mathbf{w}_i and \mathbf{e}_i are a word of Chinese word segmentation and Braille word segmentation, respectively. Analysis and experiments show that R is fine-grained and disambiguates, while L is coarse-grained but difficult to disambiguate. The two results need to be fused so that the final result R' is coarse-grained and disambiguated.

Definition 1. Concatenated word. Given $\mathbf{w}_{i,i+k} = \mathbf{w}_i \mathbf{w}_{i+1} \dots \mathbf{w}_{i+k}$, the segmentation results in a Chinese sentence $\mathbf{R} = \mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_m$, ($1 \leq m \leq N$), and the Braille segmentation results $\mathbf{L} = \mathbf{e}_1 \mathbf{e}_2 \dots \mathbf{e}_r$, ($1 \leq r \leq N$); if $\mathbf{w}_{i,i+k} = \mathbf{e}_j$, then \mathbf{e}_j is called the concatenated word.

```

R and L are the Chinese word segmentation result and the Braille word segmentation result, respectively, and the input Braille word
segmentation result R' that eliminates ambiguity while meeting the word segmentation and concatenation rules requirement.
for i=1 to m
  for j=1 to r
    if  $w_i = e_j$  then  $R'R' \leftarrow +w_j$ ; break;
    else if  $w_{i+i+k} = e_j$  then  $R'R' \leftarrow +e_j$ ,  $i \leftarrow i+k$ , break;
    else if  $e_{j,j+k} = w_i$  then  $R'R' \leftarrow +w_i$ ,  $j \leftarrow j+k$ , break;
    else if  $w_{i+i+k} = e_{j,j+l}$  then  $R'R' \leftarrow +w_{i+i+k}$ ,  $i \leftarrow i+k$ ,  $j \leftarrow j+l$ , break;
  return R'

```

ALGORITHM 1: Chinese word segmentation results adjusted to word segmentation and concatenation rules.

Idea. By default, the word segmentation is selected from R and placed in R' . When there combines R in L , the word segmentation of L is placed in R' (Algorithm 1).

3.1.4. Chinese to Pinyin. As shown in Figure 1, the previously divided sentences are converted into Pinyin sequences. There are also relatively mature algorithms to realize the conversion of Chinese characters to Pinyin, but the biggest difficulty lies in polyphonic characters. Theoretically, if the pronunciation of the word is unique, it can be directly converted; if the number of pronunciations of the word is or greater than 2, the pronunciation of the word must be determined by the context.

Based on the Chinese sentence segmentation of the HanLP dictionary and the Chinese-Braille dictionary, the Chinese-Braille word corpus was generated. The Chinese-Braille word corpus used the probability to select the Pinyin sequence of words in the Chinese-Braille word corpus as the Pinyin sequence containing polyphonic words. A unigram language model is used for polysyllabic words in Braille dictionaries to reduce the problem of polysyllabic words.

The Pinyin sequence after Chinese word segmentation is still different from the Braille Pinyin sequence. The difference is not in Pinyin but in word segmentation (space position).

3.1.5. Pinyin to Braille Pinyin Sequence. In order to convert Pinyin strings into Braille strings, based on the Chinese-Braille word corpus, a Pinyin-Braille syllable mapping table was established, and the Braille strings were obtained by searching the mapping table and replacing syllables one by one.

3.1.6. Braille Pinyin Sequence to UTF-8. By searching the Chinese Common Braille Pinyin sequence of the Chinese-Braille word corpus, the initials, finals, and Braille ASCII codes are outputted in UTF-8 format after identifying the mapping relationship one by one.


3.2. Braille-Chinese Translation. When translating Braille into Chinese, there are mainly four categories including Chinese characters, English letters, numbers, and punctuation marks that need to be translated. There is a one-to-one mapping relationship between the Chinese Common Braille and English letters, numbers, and punctuation marks. If the

Braille is accurate and there is no ambiguity, it can be directly converted, and the basic implementation is error-free, which will not be discussed here.

As shown in Figure 4, Braille to Chinese translation is the core. In the process of translation, Chinese Pinyin is used as the medium, and the difficulty lies in the homophones. The main process is listed below:

3.2.1. Braille Recognition and Classification. For the input Braille (UTF-8) sentence, we get the corresponding UTF-8 code of each cell of Braille. We then handle punctuation that cut the Braille sequences into Braille sentences.

3.2.2. Braille-Chinese's Pinyin Sequence. When converting Braille to the corresponding Chinese Common Braille Pinyin (initials and finals), the Braille characters have a strict one-to-one correspondence with initials and finals, making the accuracy of this step 100%. We then scan from the beginning to the end of the sentence according to the Chinese-Braille word corpus. The corresponding Chinese Pinyin can be obtained by looking up the Chinese Common Braille Pinyin, and the omitted tones can be supplemented (u2v, Pinyin, initials and finals, and other fields in Table 1).

Exceptions are as follows: (1) when “” appears alone, if the previous cell of Braille is not an initial, then it is a number symbol, and the latter cells are converted to numbers until the empty cell is found; (2) if it does not conform to the arrangement of initials, finals, and tones, and it is not a single syllable, it will be converted according to the English alphabet until an empty cell is found (ended with a concatenated word segmentation).

3.3. Pinyin-word Conversion. Suppose the Chinese Pinyin sequence $\mathbf{S} = c_1 c_2 \cdots c_N$, c_i is the Chinese Pinyin sequence of Braille word segmentation and concatenation. Spaces are used to separate $c_i c_j$.

Based on the Chinese-Braille vocabulary corpus, the maximum matching algorithm is used to convert the Pinyin sequence into Chinese words.

3.4. The Optimal Solution of Words to Form Sentences. When the Pinyin sequence is converted into Chinese words, there is a problem with polyphonic words. As shown in

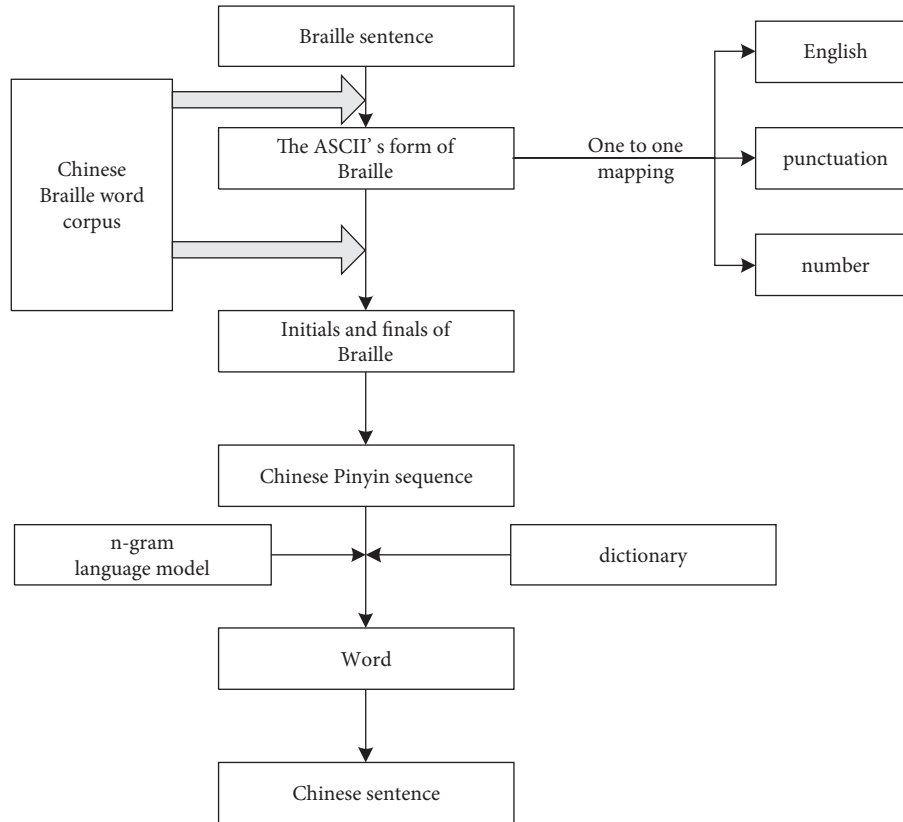


FIGURE 4: The translation process of Chinese common Braille-Chinese.

Table 4, after the previous processing, the Pinyin sequence bu4shi2 can be obtained, which can match “不时 (from time to time)” and “不识 (unknown),” in the Chinese-Braille vocabulary corpus. $L(S)$ is a candidate sentence.

Using the 2-gram language model and HanLP’s binary language model dictionary, the probability of occurrence of the entire word string is calculated. We then take the one with the highest probability as the result of sentence R .

$$R = \max_{l \in L(S)} P(l) = \max_{l \in L(S)} P(c_1) \prod_{k=2}^N P(c_k | c_1, c_2, \dots, c_{k-1}). \quad (10)$$

Finally, the normative spaces of the word segmentation in $c_i c_j$ are deleted and are taken as the output.

4. Improve Braille Word Segmentation and Concatenation Rules

This section will present the experiment in the translation between Chinese and Braille and the fusion of Braille Word Segmentation and Concatenation Rules for the improvement of mutual translation. However, the word segmentation and concatenation rules of Braille are not yet mature [26], while the newly promulgated Chinese Common Braille is still being promoted; therefore, there is a lack of a real Braille corpus. We, therefore, proposed a scheme for improving Braille Word Segmentation and Concatenation Rules through artificial Chinese-Braille vocabulary corpus experiments [27].

The Braille Word Segmentation and Concatenation Rules are still immature. On the basis of the corpus, by improving the Braille word segmentation algorithm, a common unregistered word segmentation can be realized, to experimentally improve the Braille Word Segmentation and Concatenation Rules. Words that do not appear in the Braille dictionary still have some statistical information. Mikolov et al. proposed a method for word segmentation of English phrases [28]. Braille sentence expressions also use spaces (blank cells) for word segmentation. The method of extracting English phrases (similar to unregistered words) can be used to improve Braille Word Segmentation and Concatenation Rules.

Given two Braille words, if there exists a situation where the number of consecutive occurrences of w_i, w_j count (w_i, w_j) is greater than the number of independent occurrences of w_i, w_j count (w_i) or count (w_j) ; then, the two Braille words w_i and w_j are considered to be one Braille word; that is, the Braille words need to be linked together instead of being segmented. A threshold function can be defined accordingly:

$$f(w_i, w_j) = \frac{\text{count}(w_i, w_j)}{\text{count}(w_i) + \text{count}(w_j)}. \quad (11)$$

Then, when $f \geq x$, two consecutive Braille words w_i and w_j will be merged into a new Braille segmentation $w_i w_j$, and the value of x needs to be set through experiments. The word vector is trained through statistical

information such as the number of occurrences of Braille words and the number of simultaneous occurrences between words, to further improve the word segmentation effect and improve the Braille Word Segmentation and Concatenation Rules.

Suppose that X_i represents the number of all Braille word segmentations that appear in the context of the Braille word w_i ; X_{ij} represents the number of times the word w_j appears in the context of the word w_i . Then,

$$X_i = \sum_k X_{ik}. \quad (12)$$

The frequency of Braille segmentation w_j in the context of w_i is

$$P_{ij} = \frac{X_{ij}}{X_i}. \quad (13)$$

We calculate the statistics of the cooccurrence times between Braille words and words in the Braille corpus, and the matrix of Braille word segmentation vector matrix is A . Then,

$$A \in R^{|V| \times d}, \quad (14)$$

where $|V|$ represents the number of Braille words, and d represents the dimension of the Braille word vector.

The above-mentioned Braille word vectors are large in scale, and a relatively simple model can be tested first as a training model, such as the GloVe model. The objective function trained with the GloVe model as the Braille word segmentation vector is

$$J(A) = \sum_{i,j} (A_i^T A_j - \log X_{ij})^2. \quad (15)$$

To remove the low-frequency terms, the above objective function is improved, and the weight terms are added $f(X_{ij})$.

$$f(X_{ij}) = \begin{cases} \left(\frac{X_{ij}}{X_{\max}}\right)^\alpha, & X_{ij} < X_{\max}, \\ 1, & \text{other.} \end{cases} \quad (16)$$

After simplification, the objective function of word vector training is

$$J(A) = \sum_{i,j} f(X_{ij})(A_i^T A_j - \log X_{ij})^2. \quad (17)$$

The above Braille word segmentation training results can effectively improve the Braille Word Segmentation and Concatenation Rules.

Assume that Braille segmentation w_i (or synonym or congener of w_i) and Braille segmentation w_j (or synonym or congener of w_j) occur at high frequency (w_i, w_j), then the Braille segmentation w_i and the Braille segmentation w_j should be a new Braille word. If they do not exist in the corpus, they should be added to the Braille corpus.

Assumption 1. Braille segmentation w_i and w_j can form new words $w_i w_j$ or $w_j w_i$; then, the cosine value of the included angle between the word vectors A_i and A_j corresponding to the Braille segmentation w_i and the Braille segmentation w_j will be close to 1.

If the value obtained by calculation is greater than a certain threshold λ (specified via experiments), it is considered that the Braille segmentation A_{j-1} and the Braille segmentation A_j form a new Braille segmentation $A_{j-1} A_j$. The word segmentation vector is

$$A_{j-1} A_j = A_{j-1:j} = \frac{A_{j-1} + A_j}{|A_{j-1} + A_j|}. \quad (18)$$

The above is to use the existing annotated Braille corpus for training and an experiment to improve the Braille Word Segmentation and Concatenation Rules. For Chinese word segmentation, this method simply uses statistical information and does not use the grammatical and semantic information of words.

5. Results

5.1. Examples of Translation. We tested the effect of translation between Chinese and Chinese Common Braille. The test set is the extracted sentences from books such as ‘‘Chinese Classics Reading (Large Character, Braille Edition)’’ published by China Braille Publishing House and the literary work ‘‘Looking Back-Fragments in Memory.’’ The translation program is shown in Figure 5, and the comparison between the translation result and the human translation is shown in Table 5.

5.2. Evaluation of Chinese-Braille Translation. Published by China Braille Publishing House, ‘‘Chinese Classics Reading (Large Characters, Braille Edition)’’ is a Chinese-Chinese Common Braille book, which belongs to the category of literature, with a small number of proper nouns, such as names of people and places. The Braille in the book is the Chinese Common Braille and is manually translated by Braille experts.

The BLEU (Bilingual Evaluation Understudy) evaluation index can be used to calculate the effect of machine translation [29].

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^N w_n \log P_n\right), \quad (19)$$

where

$$\text{BP} = \begin{cases} 1, & \text{if } c > r, \\ e^{(1-r)/c}, & \text{if } c \leq r, \end{cases} \quad (20)$$

where c is the number of Braille cells of the machine-translated Braille sentence, and r is the number of Braille cells of the Braille sentence translated by a Braille expert.

As can be seen from the translation results, an n-gram is fine-grained, and Braille Word Segmentation and Concatenation Rules are coarse-grained. All those that need to be

TABLE 6: Comparison of BLEU between Chinese-Chinese Common Braille translation.

Method	BLEU%	
	Chinese-Braille translation	Braille-Chinese translation
n-gram	70.97	61.57
n-gram + segmentation concatenation rules	83.16	66.22



FIGURE 6: Error analysis of Chinese-Chinese Common Braille translation.

should be concatenated are not concatenated. As shown in Figure 6, the yellow (light color when printing in black and white) in the figure is the redundant word segmentation for machine translation (should be concatenated here), and purple (dark color when printing in black and white) is the word that should be segmented, or the tones of the words should be marked in machine translation. Therefore, the quality of the translation mainly relies on the word segmentation of Braille. The evaluation metrics of Chinese word segmentation can be used to evaluate Braille word segmentation.

The evaluation metrics are accuracy, recall, precision, and F1. It is assumed that the correct number of words (the number of Braille cells) converted into Braille after segmentation processing is TP; TP + FP is the total number of Braille cells converted into Chinese after segmentation; TP + FN is the total number of Braille cells after expert manual Braille word segmentation [25]. Spaces (blank cells)

are counted in all calculations. When calculating the accuracy rate, we aligned the Braille characters of the machine translation (T text) with the human translation (H text), the total number of Braille characters after alignment is N , and the inconsistency between T and H is called the substitution error (SN). It is called an insertion error (IN) when T has a word that H has not while it is an omission error (ON) when H has a word that T has not.

The Braille to Chinese translation process is similar to the above, and the calculation formula is (21)–(24).

Accuracy A :

$$A = \frac{N - SN - IN - ON}{N}. \quad (21)$$

Precision P :

$$P = \frac{TP}{TP + FP}. \quad (22)$$

TABLE 7: Comparison of effect between Chinese-Chinese Common Braille Translation.

	Segmentation method	A%	P%	R%	F1%
Chinese-Braille translation	n-gram	90.02	93.95	93.02	93.48
	n-gram + segmentation concatenation rules	95.01	96.46	98.38	97.41
Braille-Chinese translation	n-gram	89.41	89.92	91.49	90.70
	n-gram + segmentation concatenation rules	90.15	90.84	92.07	91.45

Recall rate R :

$$R = \frac{TP}{TP + FN}. \quad (23)$$

F1:

$$F1 = \frac{2 \times P \times R}{P + R}. \quad (24)$$

We then conducted experiments based on the electronic copy of the book “Chinese Classics Reading (Large Character, Braille Edition).”

We then tested the effect of Chinese to Braille machine translation. Due to the limited size of the corpus, the training corpus is comprised of the HanLP dictionary and the Chinese-Braille word segmentation and concatenation dictionary generated from the Chinese-Braille word corpus. The corpus of Chinese Braille covers the fields of education and literature. The test set comes from “Chinese Classics Reading (Large Character · Braille Version).”

It should be noted that the content of the test set is consistent with the field of the Chinese-Braille word corpus, but the Chinese-Braille word corpus does not have the training data from “Chinese Classics Reading (Large Character Braille Version).” As shown in Table 7, the accuracy of Chinese to Braille translation is 95.01%, which is 4.99% higher than the traditional method; the F1 value is 97.41%, which is 3.93% higher than the traditional method. In the process of translating from Braille to Chinese, the accuracy rate is 90.15%, which is 0.74% higher than the traditional method; the F1 value is 91.45%, which is 0.75% higher than the traditional method.

6. Discussion

This study proposes a Chinese-Braille translation method that integrates word segmentation and concatenation rules. Firstly, the n-gram language model is used to perform Chinese word segmentation, and then, the Chinese-Braille vocabulary corpus is used to train and generate a Braille word segmentation and concatenation dictionary of a unigram language model to adjust the Braille word segmentation and improve the translation results between Chinese and Chinese Common Braille.

Experiments show that in the field of educational literature, the accuracy of translation from Chinese to Chinese Common Braille has reached 95.01%, and the accuracy of Chinese Common Braille to Chinese translation has reached 90.15%.

This research also has some limitations. One is that the effect of Braille to Chinese translation is not significantly improved. The main reason is that the resource-restrained

corpus makes the model training insufficient, and the situation of typos is serious, especially the single-character polyphonic words (such as he, she, it) that have serious errors. The translation performance failed to improve effectively even after the utilization of the Chinese-Braille word corpus. In future work, we will build a corpus of Braille sentences and use deep learning methods for training, which may result in better performance.

The Braille Word Segmentation and Concatenation Rules is not perfect. At present, word segmentation and concatenation mainly rely on manual annotation by Braille experts. If a large-scale Braille corpus can be established, it is possible to improve the word segmentation and concatenation rules of the Chinese Common Braille through experiments.

Data Availability

The Chinese Common Braille corpus is provided by Zhong Jing-Hua’s team at Beijing Union University. The data has not been fully disclosed.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. H. Zhong, “On the common Chinese braille,” *Chinese Journal of Special Education*, vol. 6, pp. 42–46+41, 2018.
- [2] Y. M. Xiao, J. L. Guo, M. Lv, X. Z. Gao, and J. H. Zhong, “A corpus based quantitative study of the Chinese common braille,” *Chinese Journal of Special Education*, vol. 4, pp. 25–32, 2020.
- [3] C. G. Zhou, B. Q. Zhang, and H. Y. Huang, “The research and realization of braille-Chinese machine translation system based on mixed language model,” *Computer Engineering And Application*, vol. 39, no. 4, pp. 127–130, 2003.
- [4] X. Y. Zhu and T. Bao, “EasyBraille: a translation system for Mandarin and Braille Natural Language Understanding and Machine Translation,” in *Proceedings of the 6th Joint Symposium on Computational Linguistics in China (JSCL-2001)*, pp. 326–331, Beijing, China, January 2001.
- [5] L. Zhuang, T. Bao, and X. Y. Zhu, “The speech and natural language processing technique used in A software system for the blinds,” *Journal of Chinese Information Processing*, vol. 18, no. 4, pp. 72–78, 2004.
- [6] T. Bao and X. Y. Zhu, “Research and implementation of transformation system between Mandarin braille and Chinese,” *Computer Engineering*, vol. 30, no. 20, pp. 45–46+100, 2004.
- [7] Y. Y. Chen, “Blind Chinese translation of word-segmentation processing algorithm,” *Network security technology and Application*, vol. 2, p. 154, 2014.

- [8] H. Q. Li, X. Zh. Fan, L. F. Li, and F. Yang, "Research and implementation of Chinese-braille machine translation system," *Computer Applications*, vol. 22, pp. 3–6, 2002.
- [9] Ch. Yang and L. Che, "Research of transformation system between Mandarin braille and Chinese," *Journal of Beijing Institute of Graphic Communication*, vol. 19, no. 6, pp. 36–38, 2011.
- [10] C. Wang, X. Wang, Y. Qian, and S. Lin, "Accurate Braille-Chinese translation towards efficient Chinese input method for blind people," in *Proceedings of the 5th International Conference on Pervasive Computing and Applications*, pp. 82–87, Maribor, Slovenia, December 2010.
- [11] X. Wang, Y. Yang, H. Liu, and Y. Qian, "Chinese-braille translation based on braille corpus," *International Journal of Advanced Pervasive and Ubiquitous Computing*, vol. 8, no. 2, pp. 56–63, 2016.
- [12] X. Wang, Y. Yang, J. Zhang, and W. H. Y. Jiang, "Chinese to Braille translation based on Braille word segmentation using statistical model," *Journal of Shanghai Jiaotong University*, vol. 22, no. 1, pp. 82–86, 2017.
- [13] X. D. Wang, J. H. Zhong, J. Cai, H. Liu, and Y. L. Qian, "CBConv: service for automatic conversion of Chinese characters into braille with high accuracy," in *Proceedings of the The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 19)*. Association for Computing Machinery, pp. 566–568, New York, NY, USA, 2019.
- [14] J. Cai, X. D. Wang, L. Zh. Tang, X. J. Cui, H. Liu, and Y. L. Qian, "A deep learning method for Chinese-braille conversion based on parallel corpora," *Journal of Chinese Information Processing*, vol. 33, no. 4, pp. 60–67, 2019.
- [15] H. Xiao, "An analysis of the ambiguity of dual-meanings of current Chinese braille and relevant strategies," *Chinese Journal of Special Education*, vol. 5, pp. 43–47+42, 2016.
- [16] L. Tang, Ch. H. Guo, and J. F. Chen, "Review of Chinese Word Segmentation Studies," *Data Analysis And Knowledge Discovery*, vol. 4, pp. pp1–17, 2020.
- [17] Z. Li, R. Wang, T. Zhang, X. Xu, and P. Liang, "Intelligent braille conversion system of Chinese characters based on Markov model," in *Proceedings of the IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference*, pp. 1283–1287, Chengdu, China, March 2019.
- [18] Y. Shimomura, H. Kawabe, H. Nambo, and S. Seto, "Braille translation system using neural machine translation technology I—code conversion," *Advances in Intelligent Systems and Computing*, vol. 1001, pp. 335–345, 2019.
- [19] X. Yu, J. Wang, Q.-Q. Hong, and R. S.-H. Y.-D. Teku, "Transfer learning for medical images analyses: a survey," *Neurocomputing*, vol. 489, pp. 230–254, 2022.
- [20] X. Yu, Q. Zhou, S. Wang, and Y. D. Zhang, "A systematic survey of deep learning in breast cancer," *International Journal of Intelligent Systems*, vol. 37, no. 1, pp. 152–216, 2022.
- [21] C. Zh, *A Research on Key Methods in Tibetan-Chinese (Chinese-Tibetan) Machine Translation under Low-Resource Condition*, Ph.D. Dissertation, Qinghai Normal Univ, Qinghai, China, 2020.
- [22] Y. L. Wu, G. Wei, and H. Z Li, "A word segmentation algorithm for Chinese language based on N-gram models and machine learning," *Journal of Electronics and Information Technology*, vol. 23, no. 11, pp. 1148–1153, 2011.
- [23] L. Zh. Feng, G. J. Yang, X. Xu, and Y. H. Xu, "Bi-direction matching Chinese word segmentation based on N-gram statistical model," *Journal of Applied Sport Management*, vol. 39, no. 4, pp. 633–643, 2020.
- [24] J. Zheng, *NLP Principles and Practice of Chinese Natural Language Processing*, pp. 164–166, Electronic Industry Press, China, 2017.
- [25] R. X. Jiang, S. Y. Huang, L. Z. Duan, and L. J. Luo, "Research on new word recognition based on rules and N-Gram algorithm," *Modern Electronics Technique*, vol. 42, no. 4, pp. 166–170, 2019.
- [26] S. H. Ju, M. H. Lu, J. X. Zhang, Ch. L. Liu, and Q. Xu, "Rules for word segmentation and link writing in Chinese-braille: advances and prospects," *Chinese Journal of Special Education*, vol. 3, pp. 37–40, 2019.
- [27] X. X. Shen and X. Y. Li, "Improving Chinese word segmentation via unsupervised learning," *Journal of Chinese Computer Systems*, vol. 38, no. 4, pp. 744–748, 2017.
- [28] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pp. 3111–3119, Red Hook, NY, USA, December 2013.
- [29] K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "A method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL 02)*, pp. 311–318, Philadelphia, PA, USA, July 2002.

Research Article

A Novel Deep Learning Network and Its Application for Pulmonary Nodule Segmentation

Dechuan Lu,¹ Junfeng Chu,¹ Rongrong Zhao,² Yuanpeng Zhang,³ and Guangyu Tian ²

¹Cancer Center, Jiangdu People's Hospital, Yangzhou, Jiangsu, China

²Department of Oncology, Jiangdu People's Hospital, Yangzhou, Jiangsu, China

³Department of Medical Informatics, Nantong University, Nantong, Jiangsu, China

Correspondence should be addressed to Guangyu Tian; 1817061027@stmail.ntu.edu.cn

Received 19 March 2022; Revised 17 April 2022; Accepted 19 April 2022; Published 17 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Dechuan Lu 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.

Pulmonary nodules are the early manifestation of lung cancer, which appear as circular shadow of no more than 3 cm on the computed tomography (CT) image. Accurate segmentation of the contours of pulmonary nodules can help doctors improve the efficiency of diagnosis. Deep learning has achieved great success in computer vision. In this study, we propose a novel network for pulmonary nodule segmentation from CT images based on U-NET. The proposed network has two merits: one is that it introduces dense connection to transfer and utilize features. Additionally, the problem of gradient disappearance can be avoided. The second is that it introduces a new loss function which is tolerance on the pixels near the borders of the nodule. Experimental results show that the proposed network at least achieves 1% improvement compared with other state-of-art networks in terms of different criteria.

1. Introduction

Lung cancer is the most deadly cancer in the world, which is characterized by high malignancy and indiscernibility [1]. Most cases of lung cancer are diagnosed when the cancer has already metastasized. Pulmonary nodules are the early manifestation of lung cancer, which appear as circular shadow of no more than 3 cm on the computed tomography (CT) image [1]. Pulmonary nodules in the early stage can be divided into solitary pulmonary nodules (SPNs), juxtapleural nodules (JPNs), and ground glass opacity (GGO) [1]. Due to such complex diversity, the shape, size, density, location, and other characteristics of pulmonary nodules are quite different, which may lead to misdiagnosis. Therefore, accurate segmentation of the contours of pulmonary nodules can help doctors improve the efficiency of diagnosis.

In recent years, there are many segmentation methods for pulmonary nodules [2–5], which can be divided into two categories: the traditional unsupervised segmentation method and the machine learning-based segmentation method. Among the traditional unsupervised segmentation methods,

the morphological method, threshold segmentation method, and clustering method are commonly used [6–10]. Although these methods are quick and simple, problems such as under-segmentation or over-segmentation still exist. Morphological methods can remove the edge burr of pulmonary nodules, but the parameters involved in the operation are not easy to control. The threshold segmentation method is not ideal in the segmentation of vascular adhesion pulmonary nodules. Sun et al. [11] used the maximum expectation algorithm and mean shift method to extract pulmonary nodules and achieved good results. However, the segmentation effect of this method was not ideal when the number of nodules was greater than or equal to 2. Although the gray threshold method proposed by Armato et al. [12] improves the segmentation accuracy, this method is time-consuming, limited, and inconvenient to use. Kanazawa et al. [13] used fuzzy clustering algorithm to extract lung and pulmonary vascular regions but lost 3D spatial feature information. Miwa et al. [14] proposed an algorithm called variable N-quoit filter for automatic recognition of pathological shadow candidates, which requires excessive manual operation and low automation.

Recently, deep learning-based methods have achieved great success in computer vision [15]. Deep learning can automatically extract features from training data and can produce fewer false judgments compared to traditional fine segmentation methods. In 2015, Long et al. [16] proposed fully convolutional network (FCN) to achieve end-to-end pixel-level prediction, which has become a trend in biomedical image segmentation. Researchers have also rapidly applied FCN to the field of medical imaging, and FCN has shown good performance in some high-intensity lesion segmentation. Based on FCN, Ronneberger et al. [17] proposed the U-NET for medical image segmentation, which is more suitable for biomedical image data processing with small data volume and can obtain better segmentation results. Due to U-NET's excellent performance in biomedical image segmentation tasks, in the following years, citations of the original U-NET literature proliferated. In addition, a number of improvements based on U-NET are also emerging. For example, Tong et al. [18] introduced the idea of residual network into U-NET to improve the segmentation accuracy of pulmonary nodules. Liu et al. [19] proposed a cascaded dual-path residual network segmentation method for pulmonary nodules, which was slightly more accurate than human experts. Zhong et al. [20] introduced the dense connection into U-NET which can not only strengthen the transmission and utilization of features but also avoid the vanishing gradient problem. Hou et al. [21] combined 3D-UNET and fully connected conditional random fields to improve the segmentation accuracy of pulmonary nodules. In 3D-UNET, the spatial and context information of pulmonary nodules was integrated to extract the different resolution characteristics. Moreover, in fully connected conditional random fields, the relationships between pixels were considered to optimize the previously rough segmentation with a step of resegmentation.

Although existing U-NET-based works have won great achievements for pulmonary nodule segmentation, there are still some existing shortcomings that should be further addressed.

- (1) The imbalance problem always exists in pulmonary nodule segmentation due to small size of pulmonary nodules. Therefore, how to choose a good loss function to overcome imbalance problems should be carefully considered.
- (2) Very discriminant features are very significant in pulmonary nodule segmentation. How to extract features from tiny targets is also very important.

In this study, based on U-NET, we propose a new network DENSE-UNET for pulmonary nodule segmentation. This new network enhances the transmission and utilization of the features, which can effectively alleviate the class imbalance problem and has a great improvement in the segmentation of small target regions such as pulmonary nodules. The main contributions of this paper are as follows:

- (1) A dense connection from DENSE-NET is introduced to U-NET to combine the features between the upper and lower convolutional layers. In view of the

difficulty of feature extraction in some small target regions, the dense connection strategy can enhance the transmission and utilization of features in the network and solve the problem of gradient disappearance.

- (2) To avoid forcing the network to rigorously learn from an imprecise ground truth which often leads to over-fitting problems, we introduce a new loss function which is tolerant on the pixels near the borders of the nodule.

Different from U-NET, the proposed network uses dense connection to combine the features between the upper and lower convolutional layers. Additionally, a new loss function is used, which is tolerance on the pixels near the borders of the nodule.

The remaining sections are arranged as follows. In Section 2, we give the structure and loss function of the proposed DENSE-UNET. In Section 3, we report the experimental results. Lastly, we conclude the whole work and indicate our future works.

2. DENSE-UNET

2.1. Network Structure. In classical U-NET [17], the features extracted from each layer are usually learned only once, and there is a lack of connection between features of different layers. Therefore, the classical U-NET has low utilization for features, which affects the final segmentation accuracy. Unlike U-NET, DENSE-NET uses dense connection to combine the features of the current layer in the network with those of all previous layers and transmit the resulting features to all subsequent layers. In this cascading way, each layer in DENSE-NET can learn the features of the previous layers, which can not only strengthen the transmission of features by the network, so as to achieve feature reuse, but also alleviate the problem of gradient disappearance in the network. Suppose o_l is the output of the l -th layer in DENSE-NET, and thus we have

$$o_l = H_l([o_{l-1}, o_{l-2}, \dots, o_0]), \quad (1)$$

where H_l is a nonlinear function of the l -th layer and $[o_{l-1}, o_{l-2}, \dots, o_0]$ represents feature fusion of different layers.

In this study, by virtue of the concept of dense connections from DENSE-NET, we design the structure of dense connection, as shown in Figure 1. We see that each dense connection module mainly contains two 3×3 convolution layers and two feature fusion operations. For the feature graph of the input dense connection module, after each convolution operation, the generated feature graph will be fused with the original feature graph to form a new feature graph, and finally the feature graph will be input to the next dense connection module.

In addition, a batch normalization (BN) layer and a rectified linear unit (Relu) activation layer are added behind each convolution layer to improve the performance of the network. The batch normalization layer is proposed to solve the problem that the training effect of network is easily

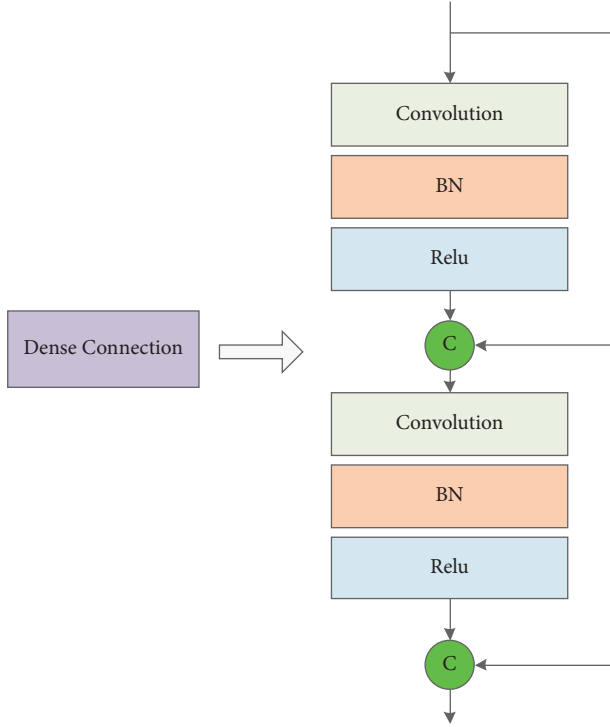


FIGURE 1: Structure of dense connection.

affected by the initial data distribution and the model generalization ability is poor. The batch normalization layer is first used to normalize the input data as

$$\begin{aligned}\mu_B &= \frac{1}{m} \sum_{i=1}^m x_i, \\ \sigma_B &= \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, \\ \hat{x}_i &= \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}\end{aligned}\quad (2)$$

where m is the batch size, σ_B is the mean value, σ_B^2 is the deviation, and ε is a smoothing factor to avoid zero denominators.

The above operation changes the distribution of features that the network learns. In order to avoid the effect of network learning being affected by the change of feature distribution, the normalized data need to be transformed and reconstructed by

$$y_i = \gamma \hat{x}_i + \beta, \quad (3)$$

where γ and β are learnable refactoring parameters. y_i is the output of batch normalized input data on the network.

The rectified linear unit activation layer is responsible for mapping the input to the output of the neuron in the neural network, which introduces nonlinear factors into the network, thus improving the nonlinear expression ability of the network. This layer is defined as

$$f(x) = \max(0, x). \quad (4)$$

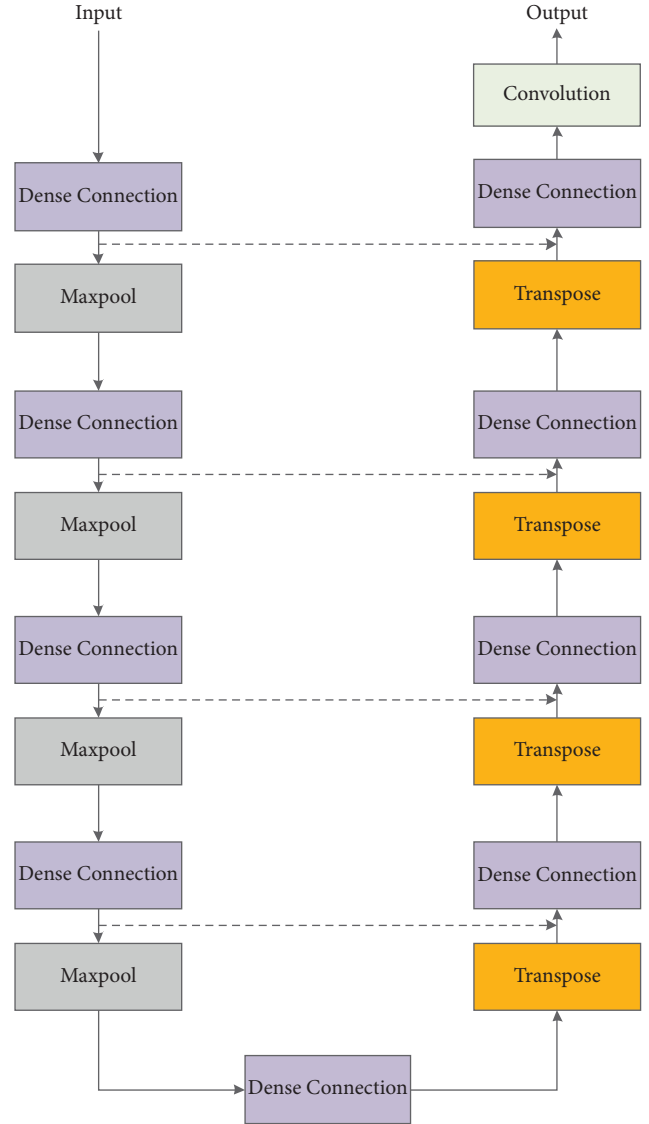


FIGURE 2: Structure of DENSE-UNET.

Based on the classical U-NET and combined with the dense connection module, we design an improved DENSE-UNET model, as shown in Figure 2. DENSE-UNET consists of an encoder, decoder, classifier, and skip connection. The encoder contains intense connection module and maximum pooling layer, where the dense connection module by convolution layer is used to extract image semantics, and the maximum pooling layer is used for the downsampling operation, which aims at reducing network computation and increasing the receptive field, so as to improve the robustness of image features. For the original input image, two convolution operations will be carried out through the dense connection module to obtain the 64×64 feature map, and then the size of the feature map is halved by pooling. Finally, after four times of convolution and pooling, the feature map of 4×4 size is obtained.

The decoder includes a dense connection module and a deconvolution layer, in which the deconvolution layer is used for the upsampling of the feature map to recover the

resolution. After each deconvolution, the size of the feature map will be doubled, and finally the feature map with the same size as the original input image can be obtained. In addition, the encoder and the decoder are linked by a dense connection module.

The classifier consists of a 1×1 convolution layer and a sigmoid activation layer, where the 1×1 convolution layer is used to reduce the number of feature maps, and the sigmoid activation layer is used to determine the class of each pixel in the final feature map so as to output the final segmentation result. The skip connection combines the simple features with the deep features in the network so as to obtain more fine segmentation results. Table 1 shows the parameter setting of each layer in DENSE-UNET.

2.2. Loss Function. In many segmentation networks, a common problem is arising, that is, precision of the boundaries of pulmonary nodules is often not available. This is caused not only by the loss function itself but also by the masks provided by the training dataset, which in many cases is not very accurate. Forcing the network to rigorously learn from an imprecise ground truth often leads to over-fitting problems. Therefore, inspired by the mean square error (MSE), we introduce a new loss function which is tolerant on the pixels near the borders of the nodule. To be specific, the proposed loss function measures the loss between one pixel and all the pixels around the corresponding one in the ground truth, inside a specific area. Then, it takes the minimum of those values. The radius of that area is $\Delta \leq f(d_{nod})$, where d_{nod} represents the diameter of the nodule and f represents a function that is defined in (5). From (5), we see that f actually reflects the relationship between the diameter of nodules and the radius of the specific area.

$$f(d_{nod}) = \begin{cases} 0 & \text{if } \frac{d_{nod}}{n} < 64 \\ 1 & \text{if } \frac{d_{nod}}{n} < 16 \\ 2 & \text{otherwise} \end{cases}, \quad (5)$$

where n is the total number of elements of the image and the unit of distance is in pixels. Furthermore, it is coupled with an exponential preventing under-segmentation. That is to say, the proposed loss function is defined as

$$L(g_i, p_i) = \frac{1}{n} \sum_{i=1}^n (g_i - p_i)^2 \sigma^{(g_i - p_i)}, \quad (6)$$

where g_i is the ground truth of the i -th pixel, p_i is the prediction result of the i -th pixel, and σ is a hyperparameter that is determined by users.

3. Experimental Results

3.1. Datasets. The training and validation sets are collected from the public dataset LIDC-IDRI which is sponsored and provided by the National Cancer Institute of America [22].

TABLE 1: Parameters in each layer.

Layers	Size of feature map	Size/step of convolution
Input	64×64	—
Dense connection	64×64	$[3 \times 3 \text{ Conv-64}] \times 2$
Max pooling	32×32	$2 \times 2/2$
Dense connection	32×32	$[3 \times 3 \text{ Conv-96}] \times 2$
Max pooling	16×16	$2 \times 2/2$
Dense connection	16×16	$[3 \times 3 \text{ Conv-128}] \times 2$
Max pooling	8×8	$2 \times 2/2$
Dense connection	8×8	$[3 \times 3 \text{ Conv-256}] \times 2$
Max pooling	4×4	$2 \times 2/2$
Dense connection	4×4	$[3 \times 3 \text{ Conv-512}] \times 2$
Max pooling	8×8	$2 \times 2/2$
Dense connection	8×8	$[3 \times 3 \text{ Conv-256}] \times 2$
Max pooling	16×16	$2 \times 2/2$
Dense connection	16×16	$[3 \times 3 \text{ Conv-128}] \times 2$
Max pooling	32×32	$2 \times 2/2$
Dense connection	32×32	$[3 \times 3 \text{ Conv-96}] \times 2$
Max pooling	64×64	$2 \times 2/2$
Dense connection	64×64	$[3 \times 3 \text{ Conv-64}] \times 2$
Max pooling	64×64	$1 \times 1 \text{ Conv}$

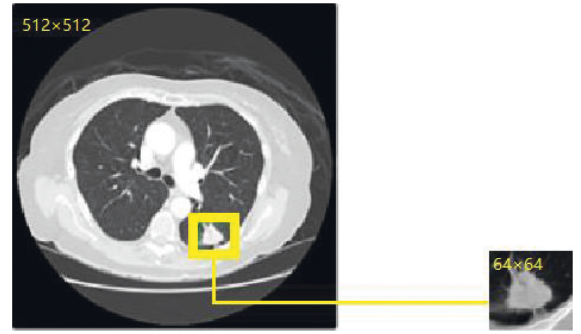


FIGURE 3: Example of original CT image reduction.

LIDC-IDRI contains CT images of 1018 patients. The nodule segmentation of each CT image is guided by 4 radiologists. The testing set is collected from the People's Hospital of Jiangdu which contains 100 patients.

In this study, we select 4000 CT images (slice < 3 mm) from LIDC-IDRI as training samples and 500 CT images (slice < 3 mm) from LIDC-IDRI as validation samples. We also select 200 CT images (slice < 3 mm) from Jiangdu People's Hospital as testing samples. The ground truth of the testing samples is obtained by manual segmentation.

Since the sizes of pulmonary nodules are far less than those of the background, the imbalanced number of positive and negative samples will seriously affect the training process of the neural network and ultimately affect the segmentation performance of the network. Therefore, in this study, the original CT image samples containing pulmonary nodules are clipped to reduce the interference of other lung tissues on the experimental results. The size of the original CT image is 512×512 . We cut the original CT image into an image with a size of 64×64 , as shown in Figure 3.

3.2. Settings. To evaluate the segmentation performance of pulmonary nodules, the Dice similarity coefficient (Dice),

precision, and recall were used as evaluation indexes, which are defined as follows:

$$\begin{aligned} \text{Dice} &= \frac{2 \sum_{i=1}^N g_i p_i}{\sum_{i=1}^N g_i + \sum_{i=1}^N p_i}, \\ \text{precision} &= \frac{\sum_{i=1}^N g_i p_i}{\sum_{i=1}^N p_i p_i}, \\ \text{recall} &= \frac{\sum_{i=1}^N g_i p_i}{\sum_{i=1}^N g_i g_i}, \end{aligned} \quad (7)$$

where g_i is the ground truth of the i -th pixel and p_i is the prediction result of the i -th pixel. From the above definitions, we see that Dice is used to measure the degree of similarity between the predicted results and the real results. Precision refers to the proportion of the total number of pixels correctly predicted as pulmonary nodules to the total number of pixels predicted as pulmonary nodules, and recall refers to the proportion of the total number of pixels correctly predicted as pulmonary nodules to the total number of pixels actually. The higher their value, the better the segmentation result.

To highlight the performance of our proposed network DENSE-UNET, FCN_32s [16], SegNet [23], U-NET [17], and U-NET with batch normalization (BN-U-NET) are introduced. All of these benchmarking networks follow the default settings. As for DENSE-UNET, Adam is adopted as the optimizer, the batch size is set to 32, the learning rate is set to $1e-6$, the momentum parameter is set to 0.8, and the number of training iterations is set to 50.

The experimental environment of this study is built based on the deep learning framework TensorFlow combined with Python. The computer configuration is as follows: the operating system is Windows 10, the processor is Intel Core i7-9700, the graphics card is NVIDIA RTX 2080TI of 11 GB memory size, and the system memory size is 32 GB.

3.3. Segmentation Results and Analysis. Table 2 shows the segmentation results of different networks on the testing set. From Table 2, we see that FCN_32s and SegNet perform worse than the other three kinds of models, and this is because FCN_32s and SegNet did not make full use of the extracted image features of each layer of the network. As for the features extracted by the encoder, FCN_32s and SegNet only carried out the upsampling to resize to the input image size, while ignoring the multiscale features on the space and the connection between the pixel location and classification. Therefore, FCN_32s and SegNet generate coarse segmentation results.

Unlike FCN_32s and SegNet, skip connection is introduced into U-NET, U-NET + BN, and DENSE-UNET based on the U-NET structure, which enables the network to combine the simple features of the shallow layer with the abstract features of the high layer. Since the number of high-quality medical images is small and the contrast between target region and background is low, it is difficult to extract features from them. Therefore, such kind of network is more

TABLE 2: Segmentation results of different networks in terms of Dice, precision, and recall.

Networks	Dice	Precision	Recall
FCN_32s	0.6885	0.7025	0.6781
SegNet	0.6944	0.7214	0.6841
U-NET	0.7211	0.7225	0.7234
BN-U-NET	0.7320	0.7454	0.7148
DENSE-UNET	0.7442	0.7551	0.7254

suitable for medical image segmentation. By virtue of skip connection, simple features are used for accurate pixel positioning and abstract features are used for accurate pixel classification. That is to say, the two kinds of features are combined to help the network obtain more precise segmentation performance. In our experiments, our network has Dice of 0.7442, precision of 0.7551, and recall of 0.7254, which performs better than U-NET and BN-U-NET. This is because on the one hand, the improved DENSE-UNET not only used the skip connection modes but also fused the features extracted from the adjacent network layers. Such structure not only strengthened the transmission and utilization of features but also can alleviate gradient disappearance problems. Therefore, the improved DENSE-UNET has a stronger learning ability for the small or fuzzy boundary pulmonary nodules in the image. On the other hand, the improved DENSE-UNET can effectively unify the distribution of training data through batch normalization operation, thus improving the convergence speed of the network, improving the optimization performance of the network, and obtaining more accurate segmentation results.

4. Conclusion and Future Work

Accurate segmentation of pulmonary nodules is very important for the early diagnosis of lung cancer. Based on the U-NET structure, we proposed a novel network termed as DENSE-UNET to segment pulmonary nodules from CT images. The improved DENSE-UNET introduced the dense connection module between the convolutional layers, integrated the features of the upper and lower layers of the network, further strengthened the transmission and utilization of the features of the network, effectively improved the segmentation performance of the network, and solved the problem of gradient disappearance. At the same time, the improved mixed loss function is used to help the network optimize the samples that are difficult to learn stably and targeted, so as to solve the class imbalance problem faced by the network in the training process. The experimental data in this study are from the public database LIDC-IDRI and the Jiangdu People's Hospital. By comparing with the experimental results of FCN_32S, SEGNET, U-NET, and BN-U-NET, it is shown that the improved DENSE-UNET can effectively distinguish lung nodules from the background region, can achieve accurate segmentation of lung nodules, and has good segmentation performance.

However, the algorithm in this paper still has some limitations. As pulmonary nodules exist in multiple CT sections, the method in this paper only focuses on the lesion

area in a single section and ignores the connection between adjacent section images, which inevitably leads to some misdetection or missed detection for small pulmonary nodules. Therefore, using the pixel connection between pulmonary nodules in adjacent sections to improve the segmentation accuracy will be the focus of the next research.

Data Availability

The data are available on LIDC-IDRI which is sponsored and provided by the National Cancer Institute of America. The code can be accessed by sending e-mail to the corresponding author.

Ethical Approval

The studies involving human participants were reviewed and approved by ethics committee of Jiangdu People's Hospital.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant (BK20201441).

References

- [1] J. Zhang, Y. Xia, H. Cui, and Y. Zhang, "Pulmonary nodule detection in medical images: a survey," *Biomedical Signal Processing and Control*, vol. 43, pp. 138–147, 2018.
- [2] Y. Qin, H. Zheng, X. Huang, J. Yang, and Y.-M. Zhu, "Pulmonary nodule segmentation with CT sample synthesis using adversarial networks," *Medical Physics*, vol. 46, no. 3, pp. 1218–1229, 2019.
- [3] H. A. Gietema and C. M. Schaefer-Prokop, W. P. T. M. Mali, G. Groenewegen, and M. Prokop, Pulmonary nodules: interscan variability of semiautomated volume measurements with multisection CT-influence of inspiration level, nodule size, and segmentation performance," *Radiology*, vol. 245, no. 3, pp. 888–894, 2007.
- [4] J. Ding, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pp. 559–567, Springer, Berlin, Germany, 2017.
- [5] A. Nibali, Z. He, and D. Wollersheim, "Pulmonary nodule classification with deep residual networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 10, pp. 1799–1808, 2017.
- [6] W. Wu, H. Hu, J. Gong, X. Li, G. Huang, and S. Nie, "Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis," *Physics in Medicine and Biology*, vol. 64, no. 3, Article ID 035017, 2019.
- [7] H. Liu, F. Geng, Q. Guo, C. Zhang, and C. Zhang, "A fast weak-supervised pulmonary nodule segmentation method based on modified self-adaptive FCM algorithm," *Soft Computing*, vol. 22, no. 12, pp. 3983–3995, 2018.
- [8] M. B. Tavakoli, M. Orooji, M. Teimouri, and R. Shahabifar, "Segmentation of the pulmonary nodule and the attached vessels in the CT scan of the chest using morphological features and topological skeleton of the nodule," *IET Image Processing*, vol. 14, no. 8, pp. 1520–1528, 2020.
- [9] 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 ID 105330, 2020.
- [10] Y. Zhang, J. Li, X. Zhou et al., "A view-reduction based multi-view TSK fuzzy system and its application for textile color classification," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–11, 2019.
- [11] S. S. Sun, H. Li, and X. R. Hou, "Pulmonary nodule segmentation based on EM and Mean-shift," *J Image Grap*, vol. 14, no. 10, pp. 2016–2022, 2009.
- [12] S. G. Armato, M. L. Giger, C. J. Moran, J. T. Blackburn, K. Doi, and H. MacMahon, "Computerized detection of pulmonary nodules on CT scans," *RadioGraphics*, vol. 19, no. 5, pp. 1303–1311, 1999.
- [13] K. Kanazawa, Y. Kawata, N. Niki et al., "Computer-aided diagnosis for pulmonary nodules based on helical CT images," *Computerized Medical Imaging and Graphics*, vol. 22, no. 2, pp. 157–167, 1998.
- [14] T. Miwa, J.-i. Kako, S. Yamamoto et al., "Automatic detection of lung cancers in chest CT images by the variable N-quoit filter," *Systems and Computers in Japan*, vol. 33, no. 1, pp. 53–63, 2002.
- [15] 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.
- [16] 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 (CVPR)*, pp. 3431–3440, Boston, MA, USA, June, 2015.
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science*, vol. 9351, pp. 234–241, 2015.
- [18] G. Tong, Y. Li, H. Chen, Q. Zhang, and H. Jiang, "Improved U-NET network for pulmonary nodules segmentation," *Optik*, vol. 174, pp. 460–469, 2018.
- [19] H. Liu, H. Cao, E. Song et al., "A cascaded dual-pathway residual network for lung nodule segmentation in CT images," *Physica Medica*, vol. 63, pp. 112–121, 2019.
- [20] S. Zhong, X. Guo, and Y. Zheng, "Improved U-NET network for lung nodule segmentation," *Comput Eng Appl*, vol. 56, no. 17, pp. 203–209, 2020.
- [21] T. Hou, J. Zhao, Y. Qiang, S. Wang, and P. Wang, "Pulmonary nodules segmentation based on CRF 3D-UNet structure," *Comput Eng Des*, vol. 41, no. 6, pp. 1663–1669, 2020.
- [22] S. G. Armato, G. McLennan, L. Bidaut et al., "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on CT scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011.
- [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.

Research Article

Cost Early-Warning Model System of Large-Scale Construction Project

Jingyi Dai and Dandan Ke 

School of Resources and Architectural Engineering, Gannan University of Science and Technology, Ganzhou 341000, Jiangxi, China

Correspondence should be addressed to Dandan Ke; 9320050229@jxust.edu.cn

Received 14 March 2022; Revised 11 April 2022; Accepted 12 April 2022; Published 10 May 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Jingyi Dai and Dandan Ke. 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.

China's construction industry has been suffering from high cost, high efficiency, and maladjustment of management. The traditional management mode makes the construction project face the risk of cost overruns and delays, which cannot achieve the effect of cost control. Therefore, by summarizing the characteristics of construction projects and the classification of costs, this paper analyzes the foundation of earned value management method in construction and combines it with cost management; in addition, the actual development of Company A is taken as an example, where a cost early-warning system suitable for construction projects is constructed, including the design of the organizational structure, operation platform, and early-warning model of this system, which can improve the cost management level of construction enterprises to a certain extent.

1. Introduction

With China's economy entering a new normal and structural adjustment steadily, the growth target of the total output in the construction industry from 2016 to 2021 is only set at 7%, which reveals that the country's expectation for the construction industry has shifted from the rapid growth in scale in the past to paying attention to the quality of construction projects. In the critical period of transformation and development, the sooner a construction enterprise abandons the growth model driven by low cost in the past and turns to the pursuit of more reasonable internal control process and more accurate cost management process, the more it can seize the opportunity to realize the growth of profit [1].

With the development of new buildings, project management is becoming more and more difficult. For a long time, due to the backward technical means and the lack of modern information management platform in China, the traditional management mode often leads to the risk of cost overruns and schedule delays for most projects, which cannot achieve the effect of cost control [2, 3]. In cost management, the deviation between the budgeted cost and

the actual cost is viewed in isolation, the construction progress is not taken into account, the cost management and progress monitoring are not effectively combined, and the management efficiency and cost management effect need to be improved. In addition, the cost management system is imperfect, and most of the cost management systems adopt timing analysis, which cannot realize the timely warning of the cost, which is more of an after-the-fact analysis that makes suggestions on the future cost management, lacking a whole-process dynamic cost early-warning system [4]. The division of responsibility for cost management is unclear, the enthusiasm of all staff for cost management is not mobilized, the participation is low, the coordination among various departments is less, and there is a lack of effective communication. When there are cost management problems, there are many cases where departments shirk their responsibilities. The responsible person cannot be found.

The theory of project cost management can already meet the development needs of daily work, but the research and development of this software still has a lot of room for development due to the complicated projects, huge data processing, multiparty collaboration, and information

sharing [5]. Therefore, this paper makes use of the earned value management method to design a system that can realize cost early warning, which combines cost management with progress monitoring and assigns responsibility to people by distributing early-warning instructions, so as to strengthen employees' awareness of cost management, improve the cost management level of enterprises, and increase the company's economic benefits.

2. Introduction of Construction Project Cost

2.1. Characteristics of Construction Projects. The construction project is a fixed asset investment led by the construction unit and implemented by the construction enterprise. Its process takes effect from the date of bidding and ends with the expiration of the warranty period. Compared with ordinary projects, it also has similar characteristics of definite objectives, irreversible management process, and limited by external environmental conditions. Besides, it has its own uniqueness: (1) large investment in construction projects, (2) long construction plan period, (3) more uncertain factors and risks, (4) large number of participants, etc. [6, 7].

Figure 1 shows the management of the whole construction process including (1) management in the decision-making stage, (2) management in the implementation stage, and (3) management in the use (operation) stage [8, 9]. Among them, the implementation stage of construction projects includes design preparation, design, construction preparation, construction, and acceptance [10].

2.2. Composition of Construction Project Cost. There are many classification methods of construction cost, which are generally classified according to different standards and different application scopes. At present, the main classification methods are the cost valuation quota method and the project completion degree method [11].

2.2.1. Classification of Cost Valuation Quota Method

- (i) Budget cost: the average social cost or enterprise cost calculated according to the physical quantity of construction and installation projects and the budget quota and charging standard formulated by the country or region or enterprise are determined by analysis, prediction, collection, and calculation based on the working drawing estimate
- (ii) Target cost: according to the requirements of the enterprise itself, such as the provisions of the internal contract, the standard cost, also known as the target cost, is determined by combining the technical characteristics, physical and geographical characteristics, labor quality, and equipment conditions of the construction project
- (iii) Actual cost: the sum of all expenses actually incurred during the construction of the project that can be included in the cost expenditure

2.2.2. Classification of Project Completion Degree Method

- (i) Current period construction cost represents all construction costs incurred in the construction of the project during the cost calculation
- (ii) Current period completed construction cost represents the partial project cost of which all the contents specified in the budget quota have been completed during the cost calculation period
- (iii) Uncompleted construction cost represents the partial project cost of the unfinished budget quota during the cost calculation period
- (iv) Completion construction cost represents the cost of the completed unit project during the whole construction period from commencement to completion

3. Application Basis of Earned Value Management in Construction

Earned value management is a management method that integrates scope, schedule, and resources to objectively measure project performance [12, 13]. On the basis of work breakdown, a set of methods combining progress monitoring and cost management is established; the project schedule, progress, and cost budget are determined according to the project schedule; and the cost budget is allocated to monitor the progress of the project [14]. Due to the inconsistency between the measurement units of schedule and cost, earned value management method considers converting time scale into monetary scale according to calculation rules and adopting the same index system to measure the performance of project schedule and cost at the same time so as to solve the difficulty that schedule and cost cannot be checked synchronously [15–17]. It is also possible to collect and sort out the data on work progress and cost related to the project at an appropriate time and analyze the next plan of construction implementation and cost management measures according to the results.

3.1. Basic Indicators. As shown in Figure 2, earned value management has three basic indicators, namely, BCWS (budgeted cost of work scheduled) [18, 19], BCWP (budgeted cost of work performed) [20–22], and ACWP (actual cost of work performed) [23, 24].

Earned value method will measure and analyze various indicators in cost control so that it can monitor the status of the project in real time, know the relevant cost information at the first time, know the progress trend of the project, and constantly adjust the corresponding measures during the project implementation to ensure that the project cost is within the controllable normal range. The measurement relationship of each index is shown in Table 1.

4. Design of Cost Early-Warning System

Company A was founded in 1956, and its construction scope covers many fields such as housing construction, municipal engineering, mechanical, and electrical

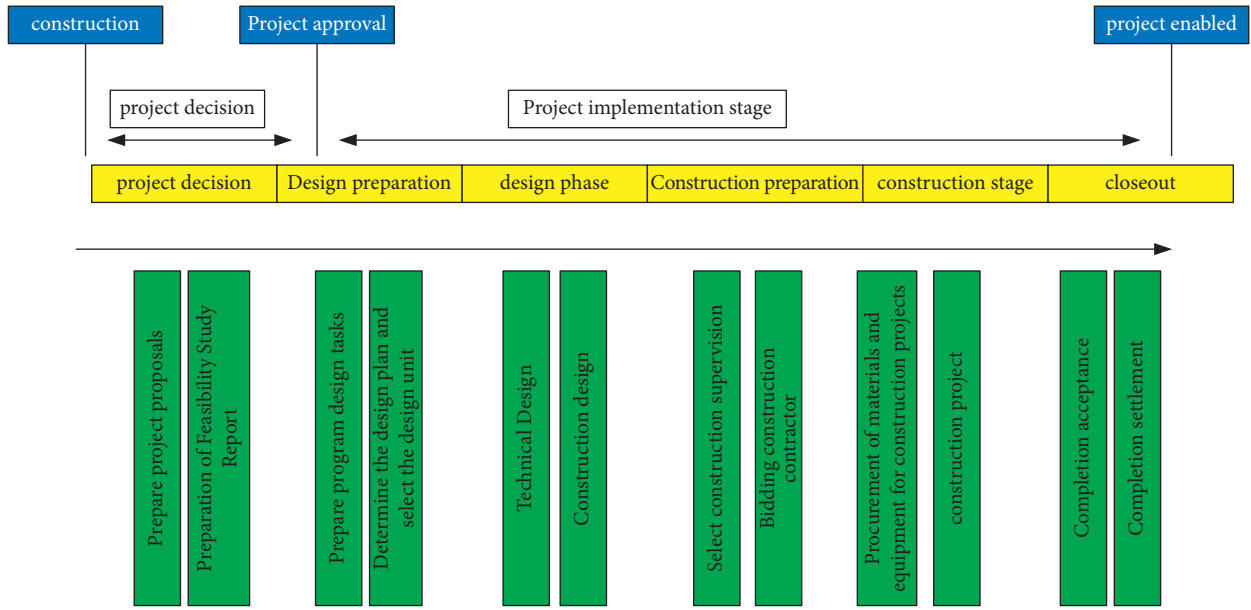


FIGURE 1: Implementation stage of the construction project.

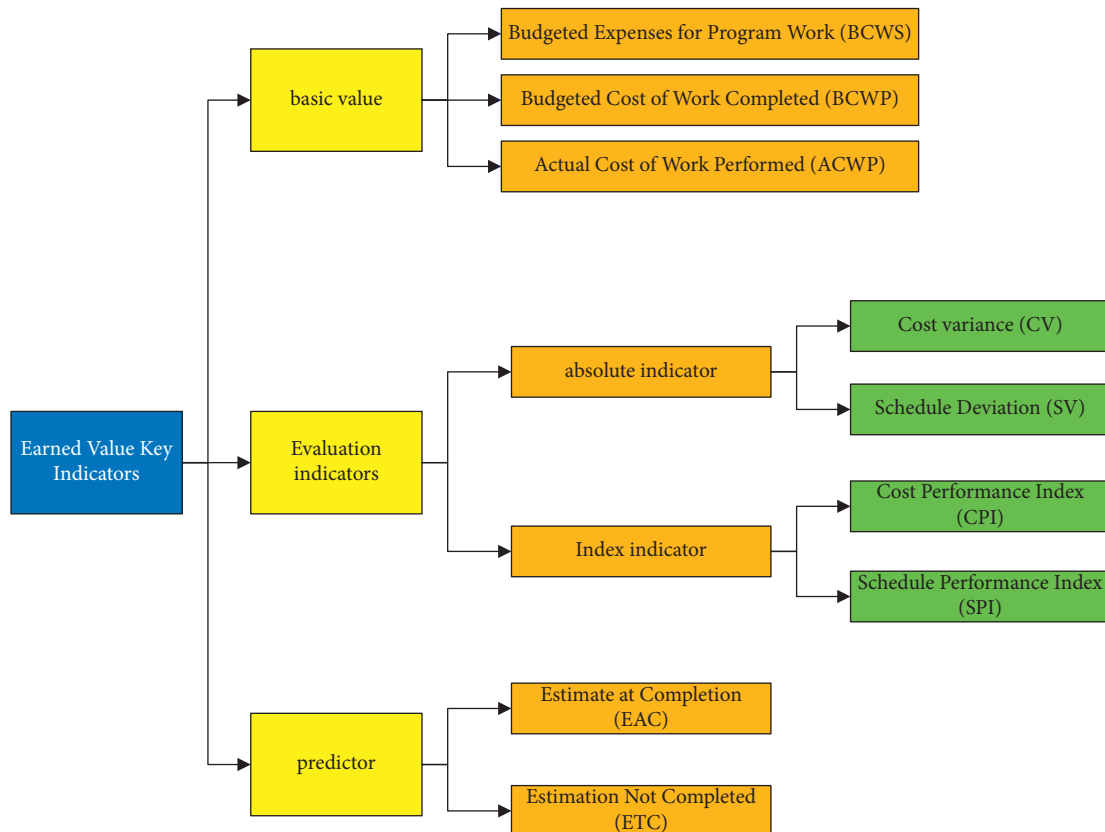


FIGURE 2: Main parameters in EVM.

engineering and participates in the construction of EPC project and PPP project. By taking Company A as an example, in this paper, the cost management of its construction project is analyzed, and a cost early-warning system is built. Through the analysis of the main problems

existing in the cost management of the construction project of Company A, some prominent problems are exposed in its cost management: therefore, it is necessary to introduce the cost early-warning system into the cost management.

TABLE 1: Evaluation index of earned value management.

Index	Calculation formula	Judging procedure
CV	$CV = BCWP - ACWP$	Measure the level of cost management. If $CV < 0$, it means that the project operation cost exceeds the expenditure; If $CV \geq 0$, it means that the project operation cost saves expenses.
SV	$SV = BCWP - BCWS$	Measure the progress of the project. If $SV < 0$, it means that the project progress is delayed; If $SV \geq 0$, it means that the project progress is ahead of schedule.
CPI	$CPI = BCWP/ACWP$	Measures the execution level of the project budget. If $CPI < 1$, it means that the cost exceeds the expenditure; If $CPI \geq 1$, it means cost savings.
SPI	$SPI = BCWP/BCWS$	Measure the actual execution level of the project workload. If $SPI < 1$, the actual progress is insufficient; If $SPI \geq 1$, it means that the actual progress is ahead.

4.1. Organizational Structure Design of Cost Early-Warning System

4.1.1. *Establish a Cost Management Center.* Company A's cost management system is imperfect, and the division of rights, responsibilities, and benefits is unclear, which leads to the widespread phenomenon that the management departments of Company A pay insufficient attention to cost management. Most departments lack communication, and cost management mainly focuses on the comparison of the difference between budget cost and actual cost, neglecting to find effective strategies, which restricts the development of the construction industry [25–28]. Therefore, it is urgent to strengthen the attention of each management department, establish an effective internal control mechanism of enterprises, and cultivate the consciousness of cost management among all employees. It is suggested that Company A set up a cost management center and choose a new method, which can respectively deploy personnel from the presettlement management center, the engineering management department, the human resources department, the finance department, and the audit department to perform the following main duties:

- (1) Formulate cost management standards: including compiling the standard manual of cost management, drawing up the detailed implementation rules for full participation, and further improving the construction of the cost management system and internal control system.
- (2) Determine the cost early-warning index: according to the characteristics of the construction project and the type of "Responsibility Letter for Construction Project Management Objectives," the feasible target cost is determined, which provides the basis for the process of early warning.
- (3) Establish the cost early-warning system: including inputting the original data information of completed projects such as target cost, actual cost, cost of each progress node, and construction progress to form a database and inputting the existing projects under construction into the system according to the construction progress so as to facilitate tracking the cost progress of new projects. In addition, exchange and

learn from the more advanced cost management measures of similar enterprises and enrich the cases in the database for reference in future work.

- (4) Monitor and evaluate the process: from the comparison and bidding of suppliers in the construction preparation stage to the monitoring of the authenticity and rationality of cost produced in the construction stage, the analysis of cost change, the identification of early warning, and the risk assessment of settlement in the final stage, the whole process of cost management can be dynamically monitored and evaluated.
- (5) Organize the cost analysis meeting: hold a cost analysis meeting for the projects that have shut down, and if necessary, adopt the field investigation method to control the latest progress of the project, eliminate abnormal factors, and timely report the projects that are really abnormal to the department leaders.

4.1.2. *Organizational Reconstruction.* In order to keep the organizational structure of a company's management layer separated from the construction layer unchanged, a cost management center is set up, and the project department submits the cost information of construction to the cost management center for review and then submits it to the corresponding departments for retention and recording. Company A's organizational reconstruction is shown in Figure 3:

4.2. Operation Design of Cost Early-Warning System

4.2.1. *Systematic Operation Platform.* The operation of Company A's cost early-warning system depends on the support and connection of each platform, and its framework should at least include prearrangement, current situation evaluation, trend prediction, and regulation scheme design.

(1) *Prearrangement.* the operating platform of the system should take cost accounting as the starting point, refer to the cost of each subitem listed in the "Construction Project Cost Analysis and Cost Control Table" for accounting. The cost

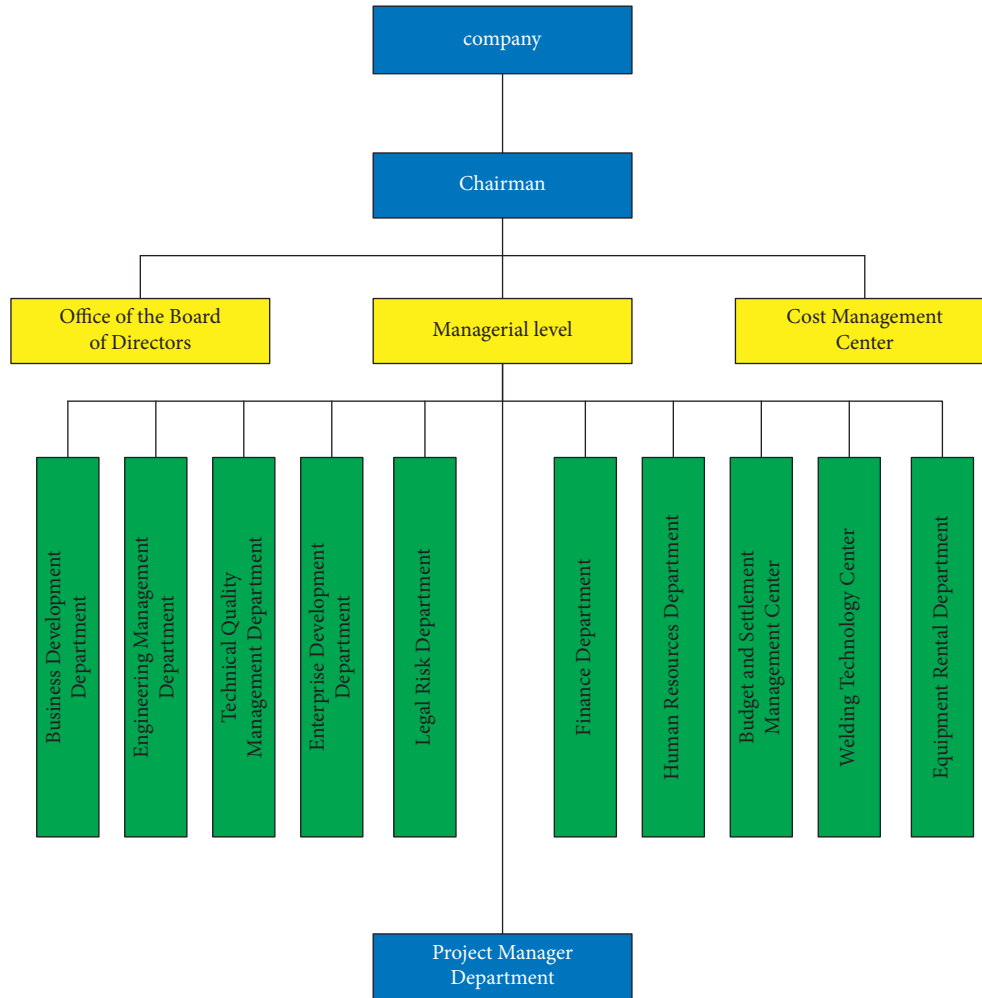


FIGURE 3: Organizational reconstruction of Company A.

analysis meeting suggests to be organized in time by the cost management center when the construction project is abnormal. The project leader, project management department, and finance department should be invited to participate together, and the cost and output income of the current month need to be compared with the statements and budgets of the previous period, respectively. After the longitudinal analysis, the cumulative amount of the whole project should be analyzed by focusing on the projects with large profits or losses in the process. In addition, the advanced cost management system can provide more accurate and timely feedback on abnormal situations and transmit them to the relevant responsible persons, thus reducing the information transmission time in the middle and effectively improving the work efficiency.

(2) *Evaluation of current situation.* cost early-warning system should work out a reasonable floating range based on the budgeted cost, which can comprehensively consider the cost preference of the project management department. It should also consider the sustainable profitability, capital turnover efficiency, and tax planning of the enterprise in combination with the management status and strategic objectives of the whole company so as to effectively guarantee the

operational safety of the enterprise. In addition, outside the reasonable floating range, it is necessary to set the corresponding sensitive range according to different degrees, which can trigger the corresponding early-warning signal in time and facilitate the relevant responsible person to handle it in time.

(3) *Trend prediction.* Company A should establish a systematic mechanism of early-warning display, prediction, and warning at the same time so as to analyze and predict the development trend of future cost through different sensitive intervals of current situation assessment. In addition, the trend should be quickly transmitted to relevant responsible persons to provide data support and guidance for their next construction plan as well as to pay attention to the feedback information and the solutions in time. Moreover, a database of coping strategies corresponding to warning needs to be established to enrich the contents of the database while warning, which can provide reference for the cost management and can be extended to other enterprises in the industry.

(4) *Regulation scheme design.* in the research of cost early-warning system, Company A needs to comprehensively consider factors such as safety production risks and

engineering quality risks faced by construction enterprises. The cost problems caused by the complexity and variability of the external environment cannot be solved naturally during the construction process. It is necessary to deeply explore the root causes of cost deviation and find solutions. At the same time, when the enterprises are regulating and controlling the system to solve the existing cost problems, new risks and problems may arise. Therefore, Company A should combine existing and contingent environmental planning measures to solve various difficulties and fully consider the possible adverse effects of existing measures to select the appropriate scheme.

4.2.2. System Operation Program. Company A's operation program of cost early-warning system is divided into four parts: early-warning preparation, cost monitoring and signaling, problem handling, and summary, which is a cycle of continuous self-improvement [29]:

- (1) Early-warning preparation, that is to determine the possible arrival time of the crisis: including early-warning knowledge reserve, early-warning elements identification and collection, and early-warning index system establishment. First, the establishment of early-warning index system needs to input the existing data of completed projects and calculate the total construction cost and its ratio. It is necessary to distinguish the project department, the nature of the project, and the letters of responsibility so as to determine the early-warning index system of Company A; secondly, enter the "Construction Project Cost Analysis and Cost Control Table" of each project; and finally, the construction in progress is entered into the system together with the contract amount and the existing cost.
- (2) Cost monitoring and signaling: when the reasonable floating range determined by the project is exceeded, the system will send out a general warning prompt, which will be passed to every member of the cost management center for attention and sent to the project leader. Cost monitoring is divided into two parts: cost operation status monitoring and cost deviation analysis rainbow early-warning chart. Two modes of real-time monitoring and timing monitoring are adopted for monitoring. Real-time monitoring is carried out at every operation with cost. After this operation is completed, the computer software will automatically draw a rainbow early-warning chart, which is convenient for operators to judge. While the timing monitoring is similar to the original mode of Company A, after the quarterly financial statements are prepared, it can be seen that Company A monitors the cost operation status mainly by real-time monitoring, supplemented by regular monitoring. When the rainbow warning chart shows that it exceeds the corresponding pre-control line, it will trigger the warning system and release the corresponding warning signal to the relevant responsible person.
- (3) Problem handling: when the cost early-warning system sends an early-warning signal, the cost management center starts to analyze the cause. At this time, the person in charge of the project needs to write written materials and report them to the cost management center for explanation. If users go beyond the warning line to the area for shutdown and rectification, it will automatically trigger a severe warning. In principle, no new cost should be added, and the project leader should be immediately notified to hold a cost analysis meeting and discuss the causes and countermeasures together. Any subsequent cost can only be added with the account number of the person, and the cost can only be added normally after the early warning is lifted. After investigating the reasons, it will search the matching solution from the system and implement the action plan to remove the warning signal. When the cause of the problem is not in the system reserve plan, it will seek solutions from outside actively, including communication with owners, government agencies, and suppliers, as well as investigation and reference from other construction enterprises.
- (4) Summary: after the early warning is lifted, the early-warning treatment plan will be recorded, summarized, and entered. The summary of early warning is generally divided into three types. First, analysis of the cause, which is meant to describe the causes of this crisis. Second, conclusion, where Company A can start from the operation such as early-warning identification and crisis handling or evaluate from the macrostructure such as the layout structure of the whole early-warning system and the establishment of early-warning indicators; finally, rectification. After comprehensively classifying the problems, timely put forward rectification opinions and implement them.

The flow chart of the cost early-warning system is shown in Figure 4.

4.3. Establishment of Cost Early-Warning Model

4.3.1. Selection of Indicators. The premise of combining cost management with schedule monitoring is the need for an accurate schedule plan. The work content of the project and the sequence constitute the key factors of the schedule. According to the requirements of the construction budget and scheme, the decomposition structure diagram of the construction operation needs to be constructed. In addition, it should decompose the construction into different task structures and then into different task units according to their interrelation and logical order. Task units need to be integrated into daily production and operation activities to guide the construction projects to be completed according to the schedule. Figure 5 is a typical diagram of work breakdown structure.

According to the chart, it is more appropriate to bring earned value indicators into the cost early-warning system of

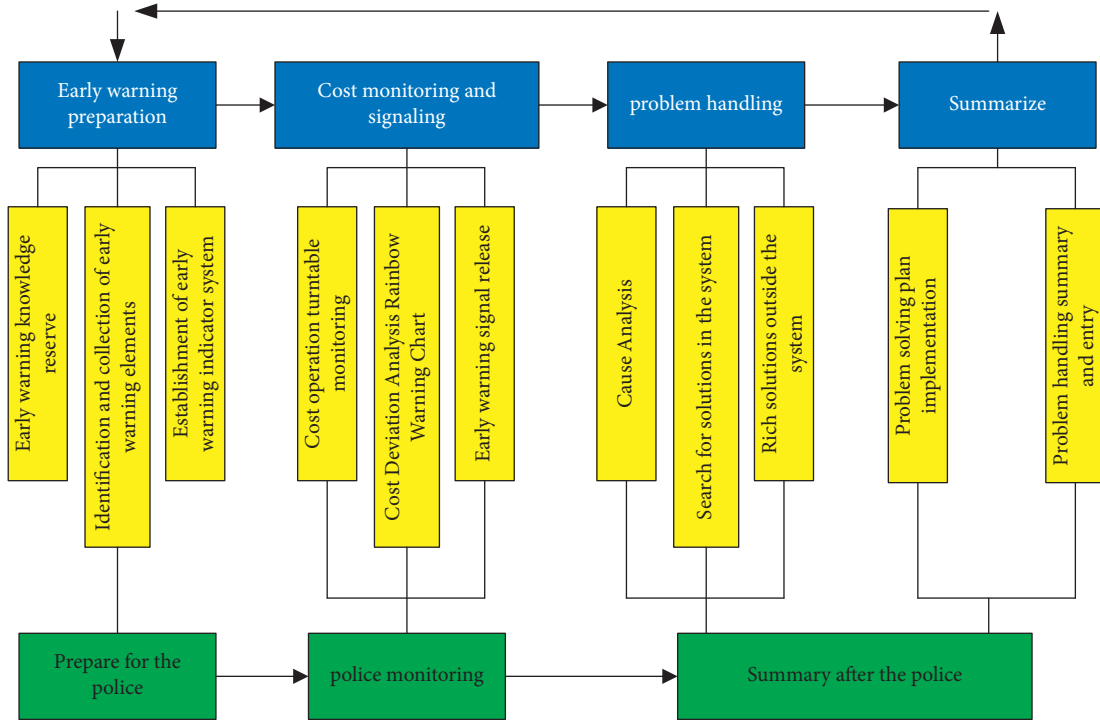


FIGURE 4: Design of cost early-warning system.

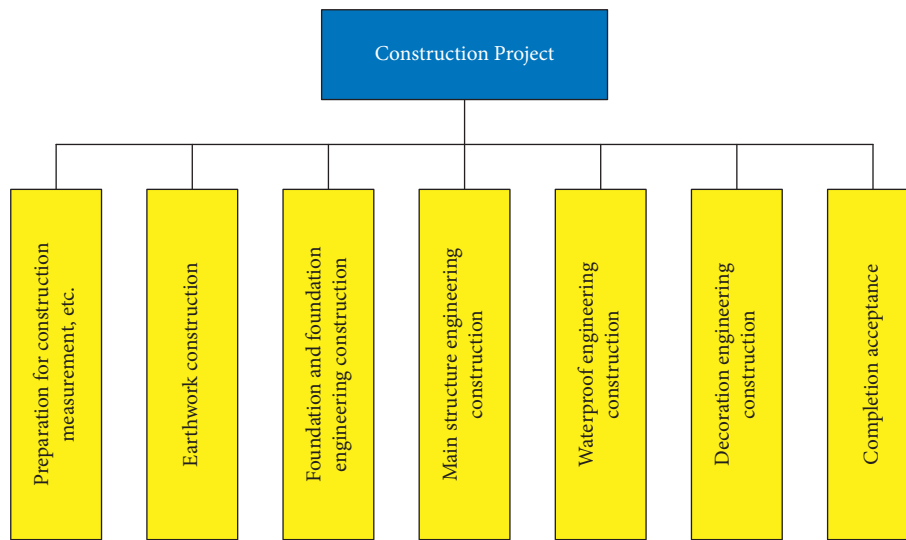


FIGURE 5: Work breakdown structure diagram.

Company A after defining the work content and scope of the construction project.

4.3.2. *Cordon Setting*. There are two kinds of deviations in the process of construction progress, namely, positive deviation and negative deviation. The positive deviation indicates the cost saving or the advance of progress. However, it may also be the result of the decline of engineering quality, or the change of operation plan, which seems to be uncertain and will bring hidden dangers to the next construction progress. Negative deviation often brings the waste of cost or

backward progress, but unexpected situations also happen from time to time. For the convenience of later construction, the overbudget expenditure that often happens in the early stage needs specific analysis. Therefore, it needs to comprehensively consider various factors in the construction process.

The cost early-warning system based on deviation analysis is suitable for dynamic monitoring of construction project cost management where the results presented by this system can provide timely and accurate cost dynamic information to relevant responsible persons and provide data support for the follow-up progress of construction projects.

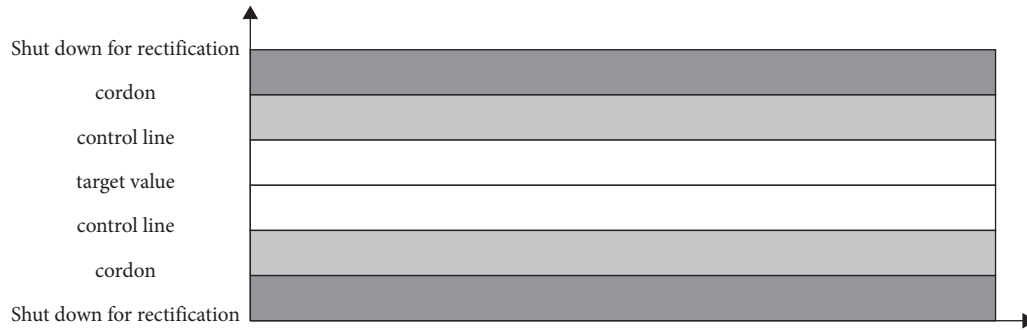


FIGURE 6: Rainbow diagram.

The deviations between the actual schedule and the planned schedule and between the actual cost consumption and the budget consumption cannot cover the complicated actual situation. If the deviation value is extremely small, the correction cost will be high; if the deviation value is large, then active measures should be taken. In addition, other deviations have no corresponding correction; therefore, it is necessary to delimit the range of deviation and establish early-warning system for deviation analysis.

Rainbow early-warning system can realize the above functions. The application of the rainbow early-warning system needs to first determine the deviation unit and early-warning level, then bring different project data into the system for measurement and judgment, and visually display the results, which is convenient for the responsible person to control and take action in time.

Studies have shown that there is no clear specification for the demarcation of the cost warning line. Coupled with the particularity of large size and tight construction period, there is not enough theoretical support in this field [30, 31]. The research on financial early warning follows the principle of “half-and-half division” and delimits the interval according to the cost management requirements of Company A. As shown in Figure 6, with the target value as the control center line, a control line is divided on both sides of the target value, then 0.50 standard deviation is taken as the unit to form a control area of standard value, which is called the target area (blank area). Afterwards, take the target value as the control center line, draw a warning line on each side of the blank area, and form an area with a difference of 0.5 standard deviation from the boundary line of the target area, which is called the warning area (light grey area). When reaching the alert zone, a general warning signal requires the cost management center to contact the project department in time so as to understand the progress of the project and avoid the potential risks in the construction process. In addition, a cost analysis meeting will be organized if necessary. Outside the two light grey areas, it is the shutdown and rectification area (dark grey area). When reaching the shutdown and rectification zone, it is considered that the current scheme cannot meet the actual needs. If the cost is out of control, a severe shutdown warning signal will be sent immediately, and the cost management center should organize a cost analysis meeting to eliminate the problems before resuming the construction of the project.

4.3.3. Cost Warning Boundary. Because the deviation of cost does not involve whether the work is in the critical path, the monitoring of cost early warning is better than that of progress early warning. The dynamic early warning of engineering construction project cost can be adjusted according to the tracked deviation to determine whether measures need to be taken and whether the cost plan needs to be adjusted.

Set the deviation strength value of the cost as follows:

$$ai = \frac{BCWP * ACWP}{ACWP} \times 10 \times \frac{CV}{ACWP} \times 100\%. \quad (1)$$

Assuming that the control range of allowable fluctuation of the project cost is $\pm\delta\pm\delta$; the boundary for monitoring and warning signals is $\pm\epsilon$, which is divided into the following four situations for analysis.

- (1) If the cost deviation intensity ai falls beyond $+\delta$, the cost plan needs to be revised
- (2) If the cost deviation intensity ai falls beyond $-\delta$, it is necessary to consider whether the cost balance has an impact on the project quality based on the project situation
- (3) If the cost deviation intensity ai falls between $+\delta$ and $+\epsilon$, an early warning is needed to find out the key factors leading to the deviation and take measures to reduce the cost
- (4) In other cases, it is considered that the cost is within the controllable range, so no treatment is required.

5. Conclusion

For enterprises, choosing the best cost management system according to their actual situation and the degree of management refinement can promote the management level of enterprises. This paper attempts to set up a cost early-warning system for construction projects by setting up a cost control center. The results show that the earned value management method is helpful to establish the key points of cost early warning, set the warning line index, and establish the cost warning boundary so as to meet the actual situation of the company and construct an effective construction project cost early-warning system. The cost early-warning system can realize the dynamic control of the whole process

of the cost before, during, and after the project, where the cost status more intuitively and in the whole process is displayed and warned, which is convenient to compare and analyze different construction projects.

Data Availability

The data set can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Research Project on Accurate Control of Engineering Elements of Urban Large Complex Building Based on BIM (No. GJJ218512), Jiangxi Provincial Department of Education, and Hubei Provincial Education Science Project in 2018 "Research on the construction of practical Teaching System based on innovation Ability Cultivation," and the project number is 2018GB152.

References

- [1] J. Xu, "Construction project cost management model based on big data," *Journal of Physics: Conference Series*, vol. 1852, no. 2, 2021.
- [2] U. S. Igwe and S. F. Mohamed, "Recent technologies in construction; A novel search for total cost management of construction projects," *IOP Conference Series: Materials Science and Engineering*, vol. 884, no. 1, Article ID 012041, 2020.
- [3] T. G. K. Vasista, "Towards innovate methods of construction cost management and control," *Civil engineering and urban planning: International Journal*, vol. 4, no. 1, 2017.
- [4] F. Mohammed, S. Al-Zwainy, R. Amer, and T. Khaleel, "Reviewing of the simulation models in cost management of the construction projects," *Civil Engineering Journal*, vol. 2, no. 11, 2016.
- [5] C. Liu and J. Jiang, "Application of computer technology in construction cost management," *Bio-Technology: An Indian Journal*, vol. 10, no. 20, 2014.
- [6] L. Xie, Y. Chen, and R. Chang, "Scheduling optimization of prefabricated construction projects by genetic algorithm," *Applied Sciences*, vol. 11, no. 12, p. 5531, 2021.
- [7] W.-Y. Wang, Y. Wang, and R. R. Sudhakarapandian, "Diagnosis index system setup for implementation status management in large-scale construction projects," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5531449, 9 pages, 2021.
- [8] P. Fatemeh and H. Behzadan Amir, "Core competencies for construction project management: literature review and content analysis," *Journal of Civil Engineering Education*, vol. 147, no. 4, 2021.
- [9] C. Hu, J. Wang, and M. Yuan, "Uncertain time-resource-cost trade-off models for construction project schedule," *KSCE Journal of Civil Engineering*, vol. 25, 2021.
- [10] A. Ayman and ElR. Khaled, "Multiobjective optimization model for planning repetitive construction projects," *Journal of Construction Engineering and Management*, vol. 147, no. 7, 2021.
- [11] M. Çevikbaş and Z. Işık, "An overarching review on delay analyses in construction projects," *Buildings*, vol. 11, no. 3, p. 109, 2021.
- [12] M. Zhao and X. Zi, "Using Earned Value Management with exponential smoothing technique to forecast project cost," *Journal of Physics: Conference Series*, vol. 1955, no. 1, 2021.
- [13] M. Seyed Taha Hossein, N. Siamak, and B. Morteza, "Directed earned value management based on ordered fuzzy numbers," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 5, 2021.
- [14] F. Yu, X. Chen, C. A. Cory, Z. Yang, and Y. Hu, "An active construction dynamic schedule management model: using the fuzzy earned value management and BP neural network," *KSCE Journal of Civil Engineering*, vol. 25, 2021.
- [15] W. Hakami, M. A. Shameri, and B. R. Aldhubhani, "Influences of earned value management on construction project's performance in Yemen," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 12, 2020.
- [16] G. Liu and H. Jiang, "Performance monitoring of project earned value considering scope and quality," *KSCE Journal of Civil Engineering*, vol. 24, no. 4, 2020.
- [17] S. Patil, R. Kavuru, and M. Ruhina Begum, "Analysis of construction project using earned value - a case study," *Journal of Critical Reviews*, vol. 7, no. 4, 2020.
- [18] M. Bagherpour, M. Khaje Zadeh, M. Amin, and X. Deng, "Interpretive structural modeling in earned value management," *Journal of Civil Engineering and Management*, vol. 26, no. 6, 2020.
- [19] S. Soltan and M. Ashrafi, "Predicting project duration and cost, and selecting the best action plan using statistical methods for earned value management," *Journal of Project Management*, vol. 5, no. 3, 2020.
- [20] M. Durga Sruthi and A. Aravindan, "Performance measurement of schedule and cost analysis by using earned value management for a residential building[J]," *Materials Today Proceedings*, vol. 33, 2020.
- [21] P. Velumani, N. Nampoothiri, and R. Aparnadevi, "Prediction of construction project duration and cost using earned value management," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1S3, 2019.
- [22] E. Amin, S. M. Mousavi, and V. Mohagheghi, "A new interval type-2 fuzzy approach for analyzing and monitoring the performance of megaprojects based on earned value analysis (with a case study)," *Neural Computing & Applications*, vol. 31, no. 9, 2019.
- [23] O. P. Edem, "The earned value management (EVM) model unveiling the mask," *Journal of Emerging Trends in Engineering and Applied Sciences*, vol. 10, no. 4, 2019.
- [24] J. Ge-di, Y. Wang, and L. I. Jing-jing, "KFFM-based large-scale project earned value management completion prediction," in *Proceedings of the 2019 International Conference on Informatics, Control and Robotics*, pp. 436–441, ICICR, Shanghai, China, June 2019.
- [25] C. Babu Kapuganti, K. V. G. D. Balaji, and T. Santhosh Kumar, "Comparison of project monitoring and controlling methods: earned value management (EVM) & earned duration management (EDM)," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6, 2019.
- [26] M. Amin, B. Morteza, and J. S. Ahmed, "Grey earned value management: theory and applications," *IEEE Transactions on Engineering Management*, vol. 68, 2019.
- [27] F. Elghaish and S. Abrishami, "A centralised cost management system: exploiting EVM and ABC within IPD, Engineering," *Construction and Architectural Management*, vol. 28, no. 2, pp. 549–569, 2020.

- [28] F. Elghaish and S. Abrishami, "A centralised cost management system: exploiting EVM and ABC within IPD," *Engineering Construction and Architectural Management*, vol. 28, 2020.
- [29] M. Glogovac and J. Filipovic, "Quality costs in practice and an analysis of the factors affecting quality cost management," *Total Quality Management and Business Excellence*, vol. 29, no. 13-14, 2018.
- [30] T. Iryna and O. Oksana, "Methodical approaches to the cost management of industrial enterprises," *Modern Economics*, vol. 7, 2018.
- [31] D. Liu, X. Zhang, C. Gao, Mo Yang, li Qi, and Li Mu, "Cost management system of electric power engineering project based on project management theory," *Journal of Intelligent and Fuzzy Systems*, vol. 34, no. 2, 2018.

Research Article

Interactive Design of Personalized Website Search Interface Based on Visual Communication

Zhen Xu ¹ and Shan Wang²

¹College of Art, The Tourism College of Changchun University, Changchun 130607, Jilin, China

²The Language Set, Changchun Experimental Middle School, Changchun 130117, Jilin, China

Correspondence should be addressed to Zhen Xu; xz@tccu.edu.cn

Received 23 February 2022; Revised 9 March 2022; Accepted 15 April 2022; Published 9 May 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Zhen Xu and Shan 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.

Aiming at the problems of low user satisfaction and long search time in the traditional interactive design method of the personalized website search interface, a personalized website search interface interactive design method based on visual communication is proposed. Under the analysis of personalized website users' search behavior, the interactive personalized website search interface is designed through a navigation module, search module, link module, interactive layout module, and visual rendering module; in the visual rendering module, the advanced texture mapping method is used to render the personalized website search interface; on the personalized website search interface, the disturbance function is imported along the normal vector, the simplified new normal vector is intelligently calculated through the concave convex texture mapping algorithm, the normal vector is solved to generate the intersection point of the high-precision interface, the illumination brightness value of each pixel of the interface is intelligently calculated, the visual communication rendering model is constructed, and the visual communication effect of the interface is improved. The simulation results show that the website interface search time of this method is within 4.9 s and the user satisfaction is up to 100%, indicating that the interaction effect of the personalized website search interface designed by this method is good.

1. Introduction

Since entering the 21st century, the information industry has developed rapidly. The Internet has been applied in all aspects of people's life and occupies a very important position. At present, with the rapid development of information technology, for most people, the main source of external information is browsing the web. Therefore, designing a personalized website search interface with a good user experience has become particularly important. At the same time, it has attracted more and more attention from Internet enterprises. Good user experience, easy-to-use and interesting interaction effects, and comfortable and easy-to-read visual interface are important factors to win user loyalty for the website. In people's life, the demand for Internet products is growing and the number of websites is also increasing rapidly. However, the quality of websites is also

mixed [1]. In order to increase the visits of websites, attract users, and bring economic benefits, major websites have tried every means to make various distinctive personalized websites. In addition, in recent years, web technologies such as H5 and CSS3 have developed rapidly and more special effects have been applied in the design of personalized website search interface. The expression forms of personalized website search interface are more rich and diverse, which is the result of the progress of technology and the blooming of products [2]. At the same time, in the design of a personalized website search interface, some people deliberately use a large number of dynamic effects, exquisite pictures, and many technical special effects in order to excessively pursue the interface effect of the website but ignore too much in the use quality of the product itself, such as the human-computer operation interaction of the website and the friendliness of the operation experience of the

website; thus, there are many so-called “vase” Internet products on the market. In this case, there will be many problems when users use the website. For example, the loading time when users open the website is too long due to the excessive use of materials on the website; when the user uses the corresponding module functions, the interactive interface is too complex or too cool, which makes the user unexpected or even surprised when operating, contrary to the user’s own expectations when using the product; in addition, there are many problems, such as complex operation steps of personalized website functions, illogical use process, and so on [3]. Such problems occur when users make complaints about the use of Tucao or even quit the website, which leads to the loss of users of personalized websites [4]. This kind of problem occurs in personalized website design mainly because designers do not pay enough attention to the interface interaction design in website design when designing personalized websites, resulting in incomplete consideration in designing website products, so as to design bad websites and bring bad user experience to users; at the same time, the personalized website products designed and developed are in a dilemma and cannot be revitalized until they are eliminated by the market. Internet products are in this situation due to the failure of design. How to design a search interface with more reasonable interaction and a good experience in personalized website design is particularly important [5].

Literature [6] proposed a multisensory visual interaction interface design method based on a fuzzy median filter. Based on the description of the main contents of the multisensory visual interaction interface, the edge of the page image is extracted, the smooth image signal is obtained by calculating the smoothing function, the page image is processed by wavelet transform, and the maximum value of the page image after wavelet transform is detected. The edge points of the image are determined according to the information obtained after the wavelet transform, the gradient vector of the image is calculated, the local maximum of the gradient vector modulus is used to determine the edge of the page image, the threshold is calculated according to the noise variance and the number of pixels of the page image, and the wavelet threshold is used to denoise the page image to improve the definition of the page image; finally, the generation of multisensory visual interaction interface under unconscious behavior is realized. Document [7] proposes an interface interaction design method based on big data processing technology, constructs the database of an interface interaction system, uses the hierarchical structure design method of process constraints for interface information interaction and big data fusion, and uses the fuzzy clustering method for information clustering of interface retrieval database. The program scheduling and cross compilation of the interface are carried out under the control of the Linux kernel source code. The interface interaction design system mainly includes process management, program control, and internal file management modules. Combined with big data processing technology, the optimization design of the interface interaction system is realized. The test results show that the designed interface interaction system has good big data information processing and scheduling ability, and the recall of data is good.

However, the user satisfaction with the above-given two methods is low. Document [8] proposed the design method of an HTML5 mobile interactive interface system based on Yipai 360. According to the overall framework of the Yipai 360 system hardware platform, the hardware structure is designed. According to the animation control, interactive setting, social application, and data application modules of Yipai 360pc, the mobile terminal configuration is adjusted in time through the setting of the panel and the Bluetooth/RS-485 gateway module is designed; we realize the bidirectional conversion between the signal and RS-485 signal, analyze the PLC data acquisition status in the protocol data, send the data packet to Yipai 360pc, control the digital or analog input and output according to the programmable logic controller, and complete the interface interaction design with the support of the execution and evaluation of the connection between Yipai 360 and HTML5 mobile interactive interface. Literature [9] proposed an information interface interaction design method based on unconscious cognition, analyzed unconscious cognitive behavior and information interaction efficiency by means of literature, investigation, and research, guided information interface interaction design by constructing an unconscious cognitive behavior model, and provided a theoretical basis for information interface design; information interface design based on user unconscious cognition is an important method to optimize the user experience and improve the usability and interaction efficiency of information interface, which provides a new idea and method for the study of information interface design. However, the above two methods take a long time to search the interface, resulting in low interface search efficiency.

Aiming at the problems of low user satisfaction and long search time in traditional methods, this paper proposes an interactive design method of personalized website search interface based on visual communication, uses an advanced texture mapping method to render the personalized website search interface, solves the normal vector to generate the intersection of a high-precision interface, and obtains the illumination brightness value of each pixel of the generated interface; we build a visual communication rendering model to improve the visual communication effect of the interface. Under this method, the website interface search time is within 4.9 s and the user satisfaction is up to 100%, which not only solves the problems existing in the traditional methods but also lays a foundation for improving the effect of the website user experience.

2. Analysis of Search Behavior of Personalized Website Users

Before the interactive design of the personalized website search interface, firstly, the user search behavior of a personalized website is analyzed. Users usually get the use mode of something from their daily use experience, and they will continue to use it instead of going deep into the specific principle. The great success of search engines has changed the behavior mode of users [10]. In the Internet experiment, when users are allowed to find ways to solve problems on the web page where they know at will, they will go to a search engine website in 85% of the cases. Users look for answers

through search engines rather than good websites. This change in their behavior makes the website focus on building “high-viscosity” websites instead of improving the optimization of personalized websites and improving the ranking in search engines. The change in users’ behavior of using personalized website search interfaces has promoted the change in personalized website function, especially the personalized website directly facing users [11]. Now, search has become an essential and important function of personalized websites. The commodity search method of Internet users when shopping is shown in Figure 1.

The proportion of users using search engines to search for goods and on-site search for goods is 27.1% and 20.6%, respectively, further highlighting the importance of search in the selection of online shopping goods [12].

In interface development and design, designers design products according to their own understanding of products (i.e., design psychological model), users use products according to their own understanding of products (i.e., user psychological model), and the platform for designers to communicate with users is the system. The user mode determines users’ understanding of products, and the design mode determines whether product operation methods are easy to learn and use [13]. When designing products, designers must consider the user’s psychological mode and design the products from the perspective that users can understand, so that the balance between the design mode and user mode is finally reflected in the product interface [14]. The design pattern, user pattern, and system representation are shown in Figure 2.

Users have a strong monopoly on the use of personalized websites. They will not blindly and passively accept information but actively obtain information, which also determines that users interact more autonomously and frequently when using personalized websites. It is found that most users browse the website interface with an “F” shaped path, that is, users first browse horizontally at the top of the website interface, and then the horizontal browsing distance will be shortened as the user’s line of sight moves down. Finally, users quickly browse the vertical area on the left side of the interface, as shown in Figure 3 [15]. Of course, this “F” mode does not represent the browsing behavior of all users. If the information and picture content of interest to users appear below, “F” mode will also become “e,” as shown in Figure 4.

Therefore, when arranging the information priority of the website interface of colleges and universities, we can follow the visual browsing rule of users from left to right and from top to bottom and put the main information and key interaction on the upper left of the website interface, so as to meet the information needs of users in time and bring users a good information interaction experience [16].

3. Interactive Design of Personalized Website Search Interface Based on Visual Communication

3.1. Design of the Overall Architecture. Based on the analysis of personalized website user search behavior, this paper designs the personalized website search interface

interactively through the navigation module, search module, link module, interactive layout module, and visual rendering module [17]. The overall framework is shown in Figure 5.

3.2. Navigation Module Design. Navigation is the directory of a personalized website search interface. It helps users understand their position in the personalized website and the overall structure of the website. It also guides users on where to go in the personalized website, so that users can quickly find the content and information they need. At the same time, it helps users walk freely through the site to find the content and functional elements they need. The design of navigation directly affects whether the information content of the personalized website search interface can be searched by users and browsed effectively [18]. For each kind of information, the more paths to find, the more likely it is to be read. This is also a good way to improve the browsing volume of a personalized website search interface. Therefore, the optimization of personalized website search interface navigation design can significantly promote the usability of the website. This paper uses the method of global navigation design to design personalized website search interface navigation. Global navigation mainly refers to mastering the path of the whole website and having a unified main navigation to control, as shown in Figure 6. Generally, there is global navigation on every page in the website [19]. No matter which page the user is on, accessing any other page can be realized through global navigation.

3.3. Search Module Design. The success of search engines has changed users’ habits of using web pages. Search has become an important behavior of users in the website. Unless the website is really small and well organized, each page should have a search view or a link to a search page. The design of the search should be concise and follow the general formula of search: an input box, a button, and the word “search.” We should avoid using too fancy design and words. At the same time, for novice users, descriptive text can be added to the search bar to inform users of the keyword content that can be input. It is necessary to avoid indicative descriptive text similar to “input keyword” because even novice users who use the website for the first time know the function of the search bar [20]. In addition, if there is a possibility of confusing the search scope and content, you need to write it out in the search bar. The design of a personalized website search bar is shown in Figure 7.

3.4. Link Module Design. The interaction between users and personalized websites is mainly completed through links. According to the user’s usage habits, the buttons and links that can be clicked at the obvious signs can improve the usability of the web page. The buttons in the web page are clickable, which is obtained from the user’s experience [21]. Therefore, the marking of text links is a place to pay attention to in the principle of ease of use. For the navigation of personalized websites or all links in websites, you can remove the underline design and just change the color or

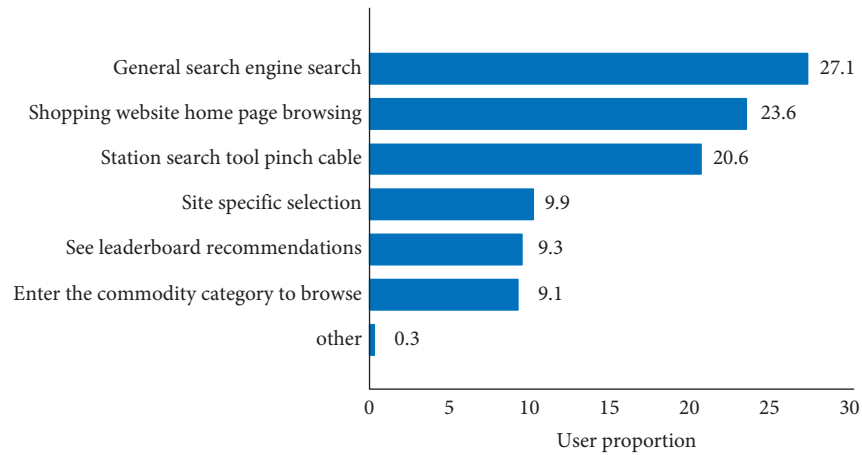


FIGURE 1: Commodity search method when Internet users shop.

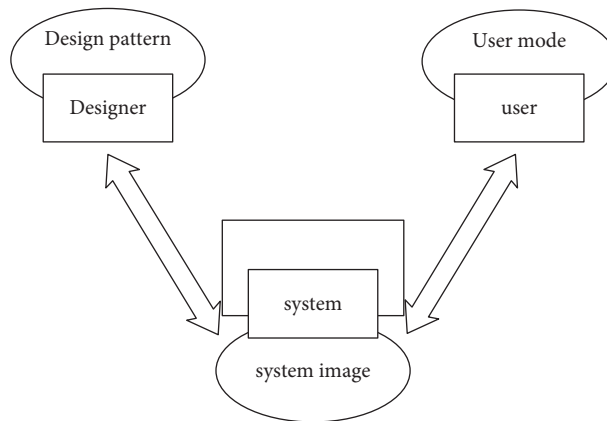


FIGURE 2: Design pattern, user pattern, and system representation.

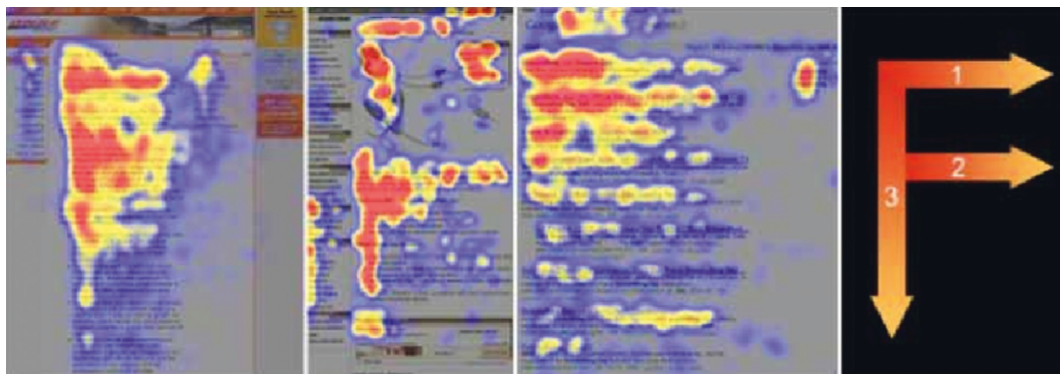


FIGURE 3: "F" browsing mode.

add an underline when the mouse moves to the link text. For the mixed arrangement of ordinary text and hyperlink text, it is necessary to clearly distinguish the difference between the hypertext link and ordinary text [22]. In addition, you can design text links as button icons.

3.5. *Interactive Layout Module Design.* As an expressive visual language, personalized website search interface design pays special attention to the interactive layout of the

interface. The layout of the web interface directly affects the convenience of users using the interface information. A reasonable interactive layout will enable users to quickly find the core content and services. On the contrary, they do not know how to obtain the required information, or how to browse to get the corresponding service, and then the user will choose to leave [23].

Although the personalized website search interface layout does not attract the user's visual attention to a certain location or object as other elements such as color and

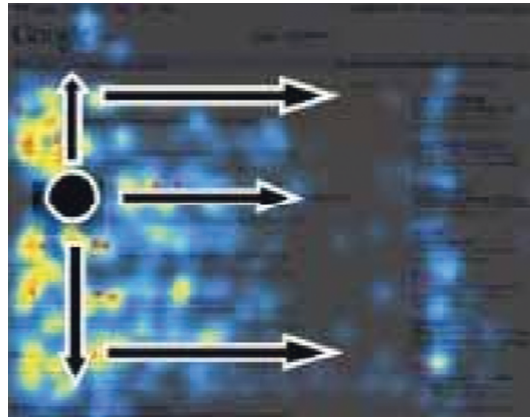


FIGURE 4: “e” browsing mode.

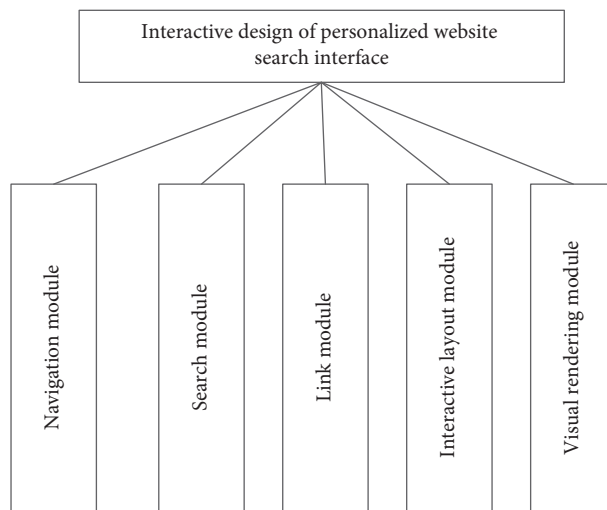


FIGURE 5: Overall design architecture.

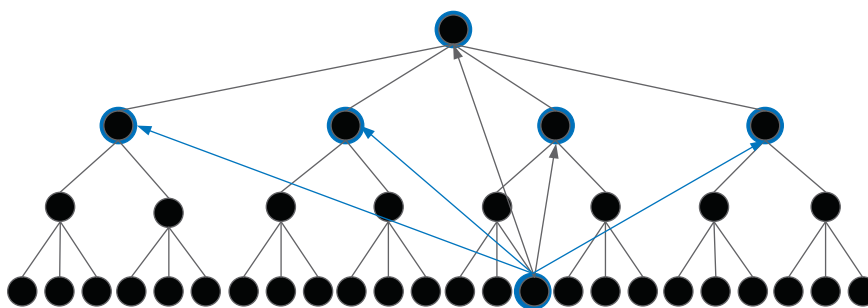


FIGURE 6: Global navigation.

graphics, a good interface layout often becomes a prerequisite for attracting the user’s attention and trying to choose and the user does not tend to be in a chaotic state. It takes time and energy to pay attention to an element or content in a layout without a sense of stability. Therefore, a good interactive layout first ensures that the user’s visual attention is stimulated. To define whether an interactive layout is chaotic, it is first necessary to ensure that the user’s visual weight on the layout reaches a certain balance; when the elements in the interface are gathered together, the visual weight is

formed. The visual weight is virtual and obtained through the user’s visual perception. Generally speaking, in order to achieve the balance of the interface layout, these visual weights must be offset by a weight with equal and opposite weight; otherwise, the layout will show an unstable state. A balanced layout can make the shift of sight operate in an orderly manner within a reasonable range without psychological burden and pressure [24]. A balanced layout design makes users feel stable and simple, which is often more attractive. The interface layout between the Mint

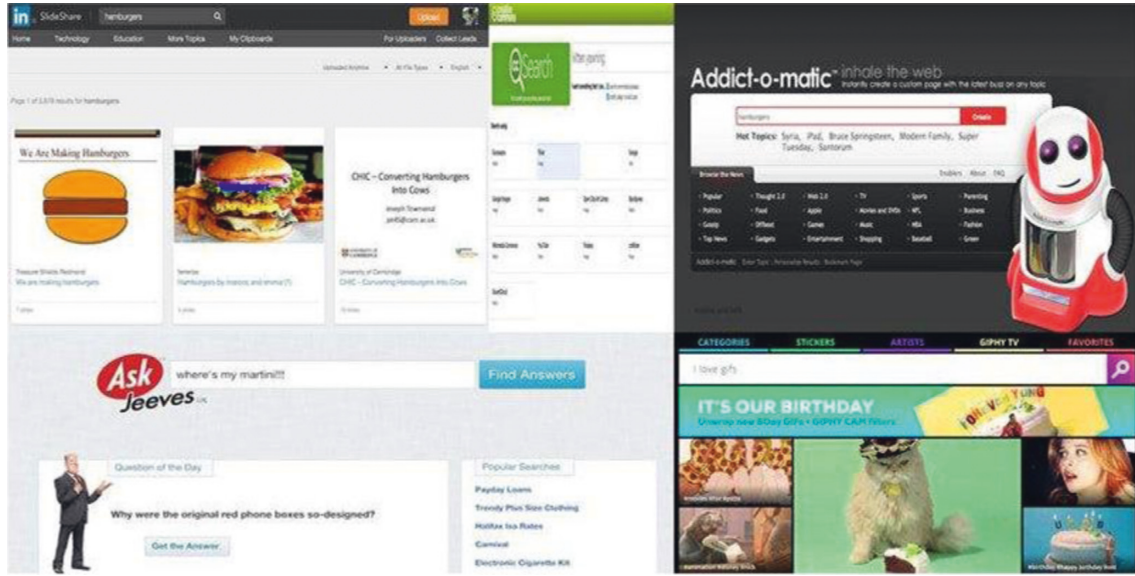


FIGURE 7: Personalized website search bar design.

Wheels website and the Dallas Baptist University website is shown in Figure 8.

As shown in Figure 8, in the Mint Wheels web interface, the logo is placed on the central axis of the interface and on the top to help establish a symmetrical balance. Moreover, the visual weight on both sides of the central axis is the same and the elements on both sides also form a corresponding relationship [25]. This makes the web interface layout delicate and concise and complements its content. However, such a symmetrical visual weight balance will inevitably appear rigid and lifeless, so the concept of balance cannot be achieved only by symmetry. When the elements on both sides are asymmetric, they can also achieve the balance of visual weight. As shown in Figure 8, the Dallas Baptist University web interface is particularly asymmetric compared with the Mint Wheels web interface, but visually, the logo balances the search box, the loose large content area on the right side of the interface balances the tight small content area on the right side, but the asymmetric arrangement of elements achieves the balance of weight, which will greatly enhance the possibility of users' visual selective attention [26].

3.6. Design of the Visual Rendering Module. The personalized website search interface includes two elements: graphics and text, and these two visual elements have one thing in common, that is, these two elements are the media to convey page information to users through visual color [27]. Color is a highly stimulating and powerful design element. Color can convey a kind of information to the browser of the interface, so as to reflect the user's psychological feelings and stimulate the user's psychological activities. Therefore, color is often easier to attract the user's attention and attract the user's attention than other elements.

Because the perception of the light wave by human eyes has a length range and different colors have different light wave lengths, people's perceptions of different colors will be

different. In general, the length of human perception of light wave ranges from 400 to 700 μm ; when the color wavelength is perceived by the human eye at both ends of the light wave range, i.e., 400 μm (red) or 700 μm (purple), the brightness of the color will weaken. Too strong or too weak brightness of the color will cause people's visual fatigue, so when selecting the color, one must try to choose the light wave length suitable for people's vision to create a comfortable interactive environment. If people look at a position for a long time, they will feel that their sight is becoming more and more blurred. This phenomenon is called visual residue [28]. Therefore, in order to improve the visual quality of the personalized website search interface and improve the visual communication effect of the interface, when designing the personalized website search interface, the personalized website search interface is rendered through the advanced texture mapping method, which includes concave-convex and normal texture mapping algorithms:

The flow of the bump texture mapping algorithm is as follows.

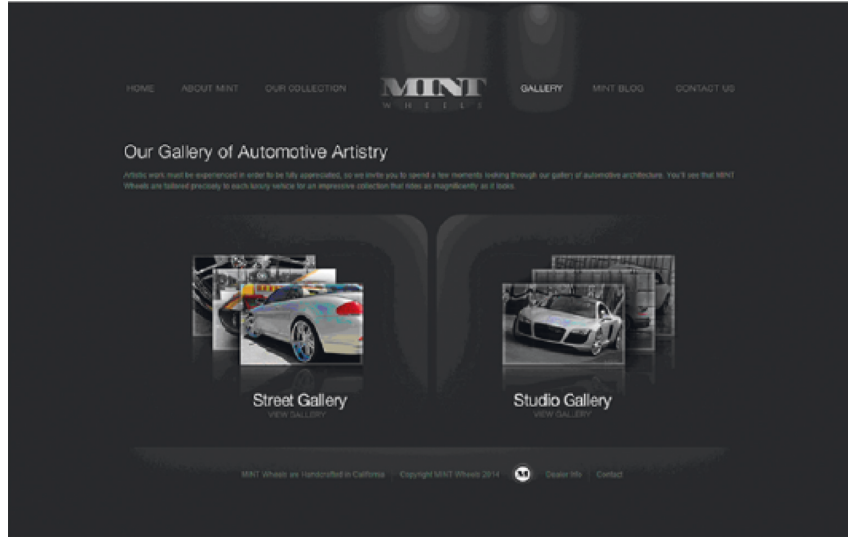
Step 1. Suppose that the web page crawling function P of the personalized website search interface is as follows:

$$P = P(x, y), \quad (1)$$

where (x, y) represents texture coordinates [29].

Step 2. Calculate the partial derivative of formula (1) to obtain P_x and P_y ; then, the normal vector E of point 1 on the personalized website search interface is shown in the following formula:

$$E = \frac{(P_x \times P_y)}{|P_x \times P_y|}. \quad (2)$$



(a)



(b)

FIGURE 8: Interface layout of (a) Mint Wheels website and (b) Dallas Baptist University website.

Step 3. On the personalized website search interface, each point imports a small disturbance function along the normal vector, as shown in the following formula:

$$P' = P(x, y) + S(x, y) \cdot E. \quad (3)$$

Here, the surface equation of the new personalized website search interface is described by P' and the disturbance function is described by $S(x, y)$.

Calculate the partial derivative of P' in x, y direction, as shown in the following formulas:

$$P'_x = P'_x + S_x \cdot E + S(x, y) \cdot E_x, \quad (4)$$

$$P'_y = P'_y + S_y \cdot E + S(x, y) \cdot E_y. \quad (5)$$

After simplification, the new normal vector of each point in the personalized website search interface is obtained, as shown in the following formula:

$$\begin{aligned} E' &= P'_X \times P'_W = P_X \times P_Y + S_X(E \times P_Y) + S_Y(P_X \times E) \\ &= E + S_X(E \times P_Y) + S_Y(P_X \times E), \end{aligned} \quad (6)$$

where the disturbance factor is described by $S_X(E \times P_Y)$ and $S_X(P_Y \times E)$. Under the action of light, the surface of a personalized website search interface will produce an uneven rendering effect [30].

In order to improve the effect of the bump texture mapping algorithm, it is improved by normal texture mapping, as shown in Figure 9.

As can be seen from Figure 9, in order to solve the normal vector, emit rays from the low to high interface, generate the intersection of a high-precision interface, generate the illumination brightness value of each pixel of the interface, and construct the visual communication rendering model. The expression is

$$FB = E(P_X' \times P_W')P(x, y). \quad (7)$$

Therefore, the rendering effect of the personalized website search interface is improved.

4. Simulation Experiment Analysis

4.1. Experimental Design. In order to verify the effectiveness of the interactive design method of personalized website search interface based on visual communication in practical application, a simulation experiment is carried out. The experiment adopts the general configuration of the current mainstream PC, and the compilation and running environment adopt common tools. Specific parameters are shown in Table 1.

Under the above-given experimental environment, this paper selects an automobile website as the experimental object for the experimental test. The experimental object is shown in Figure 10.

4.2. Experimental Analysis

4.2.1. Website Interface Search Time. The interactive design method of the personalized website search interface based on visual communication proposed in this paper, the multisensory visual interactive interface design method based on the fuzzy median filter proposed in the literature [6], and the interface interactive design method based on big data processing technology proposed in interface literature [7] are used to search the automobile website interface and test the time consumed by the three search methods. The test results are shown in Table 2.

According to Table 2, when the number of experiments is 50, the website interface search time of the method in [6] is 11.7 s, the website interface search time of the method in [7] is 18.6 s, and the website interface search time of our method is 3.7 s; when the number of experiments is 100, the website interface search time of the method in [6] is 14.5 s, the website interface search time of the method in [7] is 24.6 s, and the website interface search time of our method is 4.9 s; the interactive design method of the personalized website search interface based on visual communication proposed in this paper takes 4.9 s to

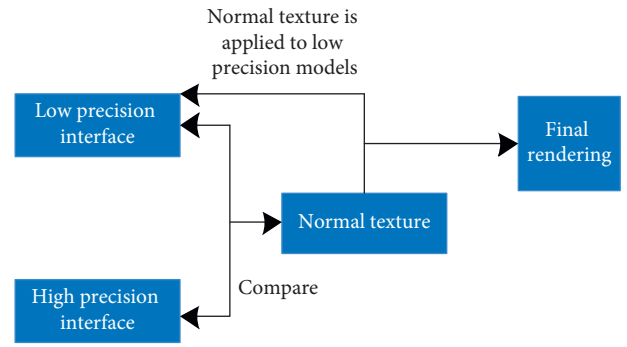


FIGURE 9: Normal texture mapping structure.

search the personalized website interface, which is shorter than the multisensory visual interactive interface design method based on fuzzy median filter proposed in the literature [6] and the interface interactive design method based on big data processing technology proposed in the literature [7]. This is because this method solves the normal vector by emitting rays from the low to high interface, generates the intersection point of the high-precision interface, obtains the illumination brightness value of each pixel of the generated interface, constructs a visual communication rendering model, improves the rendering efficiency of the personalized website search interface, and effectively improves the website interface search efficiency of this method.

4.2.2. User Satisfaction. In order to further verify the effectiveness of this method, the personalized website search interface interactive design method based on visual communication proposed in this paper, the multisensory visual interactive interface design method based on fuzzy median filter proposed in [6], and the interface interactive design method based on large data processing technology proposed in [7] are used for personalized website interface search. The satisfaction of users is tested, and the test results are shown in Figure 11.

According to Figure 11, when the number of experiments is 20, the user satisfaction of the method in [6] is 80%, the user satisfaction of the method in [7] is 68%, and the user satisfaction of the method in this paper is as high as 92%; when the number of experiments is 60, the user satisfaction of the method in [6] is 81%, that of the method in [7] is 69%, and that of our method is as high as 99%; after applying the interactive design method of personalized website search interface based on visual communication proposed in this paper, the user satisfaction is up to 100%, which shows that the interactive effect of personalized website search interface designed by our method is good. This is because this method uses the advanced texture mapping method to render the personalized website search interface, improve the visual communication effect of the interface, and improve the user satisfaction.

TABLE 1: Experimental environment parameters.

Serial number	Name	Parameters
1	Central processing unit	Intel (R) Core (TM) 2 Duo CPU E7500@2.93 GHz 2.94 GHz
2	Installed memory	8.00 G
3	Display adapter	NVIDIA GeForce GTX 550 Ti
4	Development tool	Visual Studio 2010
5	Language and corresponding library	C/C++, OpenCV 2.4.3, Qt4.7.4, Matlab R2013a
6	System environment	Windows 7

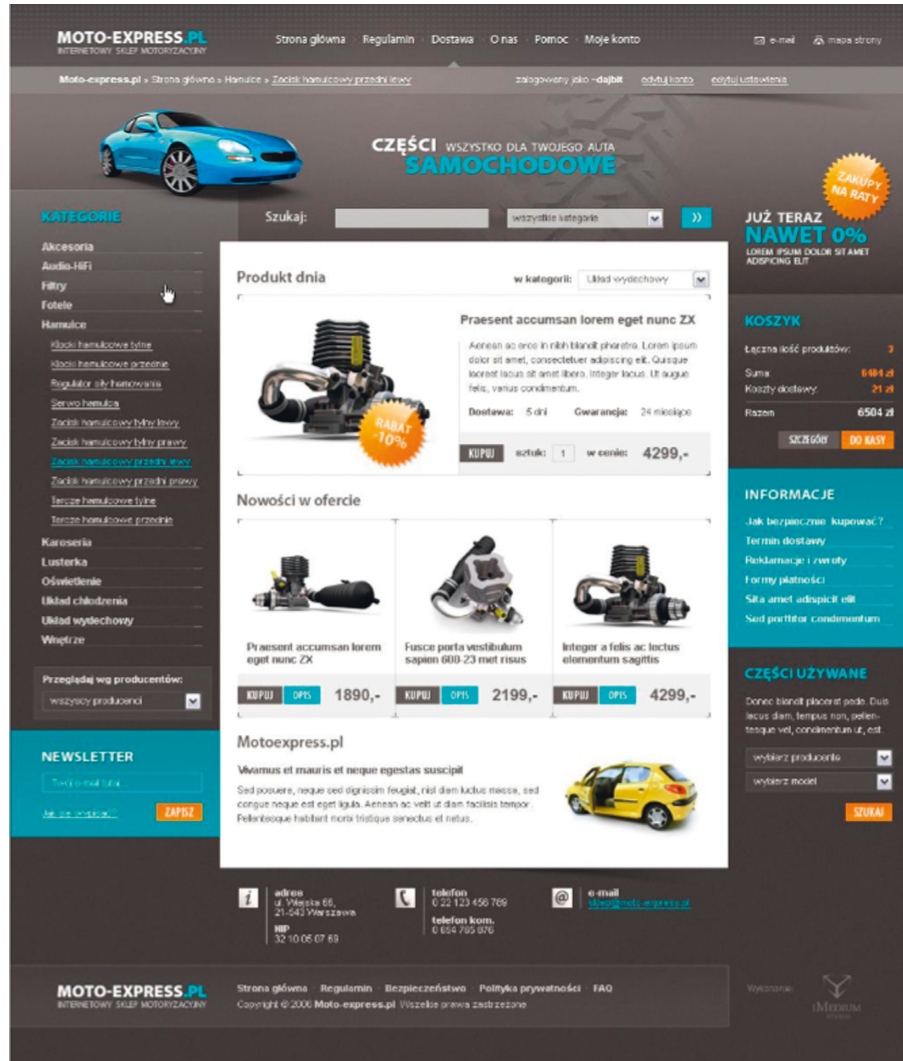


FIGURE 10: Experimental object.

TABLE 2: Website interface search time of three methods.

Number of experiments/time	This paper's method	Method in literature [6]	Method in literature [7]
10	2.6	10.6	15.6
20	2.9	10.7	17.2
30	3.1	10.8	18.6
40	3.5	11.5	18.2
50	3.7	11.7	18.6
60	3.9	11.9	19.2
70	4.1	12.5	19.8
80	4.5	12.8	20.1
90	4.6	13.2	23.5
100	4.9	14.5	24.6

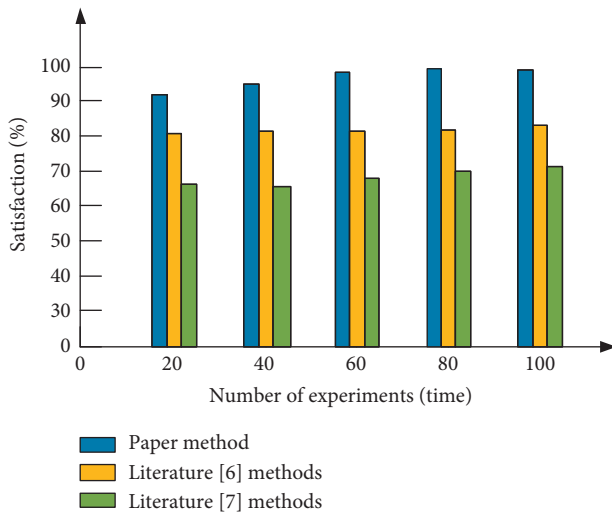


FIGURE 11: Comparison results of user satisfaction.

5. Conclusion

This paper proposes an interactive design method of a personalized website search interface based on visual communication and uses an advanced texture mapping method to render the personalized website search interface; through the concave-convex texture mapping algorithm, we intelligently calculate the simplified new normal vector, solve the normal vector to generate the intersection of the high-precision interface, intelligently calculate and generate the illumination brightness value of each pixel of the interface, build the visual communication rendering model, and improve the visual communication effect of the interface. The following conclusions are drawn through experiments:

- (1) The interactive design method of personalized website search interface based on visual communication proposed in this paper takes 4.9 s to search the personalized website interface, and the personalized website interface search time of this method is short
- (2) When the number of experiments is 60, the user satisfaction of this method is as high as 99%; after applying the interactive design method of personalized website search interface based on visual communication proposed in this paper, the user satisfaction is up to 100%, which shows that the interactive effect of personalized website search interface designed by this method is good

This paper has achieved high user satisfaction, but the search time still needs to be improved, and further research is needed in the future.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was the phased research result of the 2021 Jilin Province Vocational Education Scientific Research Project “The Innovative Practice of” Curriculum Ideology and Politics in Art Majors of Vocational Colleges,” under project number 2021XHY217.

References

- [1] L. Pang and L. Pang, “Research on B2C website interface interaction design based on Internet user experience,” *Digital communication world*, vol. 168, no. 12, p. 154, 2018.
- [2] Y. Du, X. Wang, and J. Liu, “Interactive design of website interface of high-quality resource sharing course,” *Journal of Huzhou Normal University*, vol. 2, pp. 110–112, 2021.
- [3] L. Li, “On interactive design of mobile app interface for user experience,” *Science and technology innovation guide*, vol. 16, no. 35, p. 143, 2019.
- [4] X. Pan, “Interactive design of mobile phone interface,” *Research progress of human-computer interaction*, vol. 1, no. 1, pp. 8–15, 2018.
- [5] K. Shen, B. Shao, and H. Chen, “On the web interface detail design of University Library Website Based on user experience,” *Library Forum*, vol. 2, pp. 107–110, 2021.
- [6] C. Li, W. Jiang, and W. Li, “Simulation of multi sensory visual interactive interface generation under unconscious behavior,” *Computer Simulation*, vol. 37, no. 9, pp. 231–234, 2020.
- [7] H. Li, “Research on interface interaction design based on big data processing technology,” *Modern electronic technology*, vol. 42, no. 1, pp. 38–41, 2019.
- [8] J. Kou, “Design of HTML5 mobile interactive interface system based on Yipai 360,” *Electronic design engineering*, vol. 28, no. 22, pp. 78–82, 2020.
- [9] W. Yang, “Information interface interaction design based on unconscious cognition,” *Western leather*, vol. 41, no. 18, pp. 52–53, 2019.
- [10] J. Liu and Z. Zhao, “Design of B2C shopping website web interface visual display system based on improved webml modeling,” *Modern electronic technology*, vol. 44, no. 8, pp. 40–48, 2021.
- [11] Y. Chen and J. Du, “Design and implementation of automatic generation module of management interface based on Django framework,” *Automation and instrumentation*, vol. 18, no. 5, pp. 109–111, 2018, 115.
- [12] T. Tang and Z. Shang, “Research on interface interaction design of university portal website,” *China National Expo*, vol. 18, no. 2, pp. 102–103, 2018.
- [13] Y. Song and P. Feng, “Design and implementation of a software for drawing graphics and automatically generating C language code based on visual interface,” *Computer applications and software*, vol. 35, no. 9, pp. 123–125, 2018.
- [14] J. Liu, “Research on interactive information visualization design based on personalized requirements in app,” *Western leather*, vol. 42, no. 6, pp. 77–82, 2020.
- [15] L. Zhao, L. Zhang, Z. Wang, and G. Tian, “Design of multi view human-computer interaction interface for dispatching automation system,” *Power system automation*, vol. 42, no. 6, pp. 86–91, 2018.
- [16] R. Raleigh, “Visual interactive interface design of smart home based on graphic features,” *Electronic design engineering*, vol. 28, no. 13, pp. 164–168, 2020.
- [17] B. Yan, X. Wu, and K. Tang, “Research on color coding of intelligent vehicle interactive interface based on visual

- behavior,” *Industrial engineering design*, vol. 2, no. 6, pp. 111–115, 2020.
- [18] Y. Wei, Y. Liu, X. Zhou, and H. Y. Liu, “Research on optimal design of Web interactive interface jumping visual flow based on random forest algorithm,” *Packaging Engineering*, vol. 42, no. 6, pp. 92–97, 2021.
- [19] S. Huang, “Automatic optimization design of human-computer interaction interface of NC machine tool based on human visual characteristics,” *Automation and instrumentation*, vol. 18, no. 7, pp. 186–189, 2019.
- [20] N. Zhao, “Research on the expression of color matching relationship between digital app interaction and visual design of UI interface,” *The aurora borealis*, vol. 22, no. 9, pp. 118–119, 2019.
- [21] Z. You, “Research on visual communication method of human-computer interaction interface based on Analytic Hierarchy Process,” *Journal of Zhoukou Normal University*, vol. 36, no. 5, pp. 95–99, 2019.
- [22] E. Guo, “Analysis of visual communication elements based on enterprise website interface design,” *Electronic Commerce*, vol. 18, no. 11, pp. 38–40, 2020.
- [23] X. Feng, “Research on visual communication design of e-commerce web interface in the “Internet +” era,” *The Art Book*, vol. 18, no. 6, pp. 129–130, 2018.
- [24] L. Mao, “Research on Web page visual communication design based on user personalized characteristics,” *Modern electronic technology*, vol. 41, no. 13, pp. 159–162, 2018.
- [25] Y. Li, “Research on visual communication design based on graphic beautification technology,” *Modern electronic technology*, vol. 18, no. 16, pp. 65–69, 2018.
- [26] W. Zeng, “Research on visual communication design based on mobile Internet,” *Packaging world*, vol. 18, no. 6, pp. 34–35, 2018.
- [27] G. Guo, “Application of visual communication design in mobile UI interface design,” *The Art Book*, vol. 17, no. 06, pp. 76–77, 2020.
- [28] A. Fu, “Application and design of auxiliary graphics in visual communication system,” *Fashionable Colour*, vol. 18, no. 11, pp. 62–64, 2020.
- [29] C. Ye, “Application of visual communication design in mobile UI interface design,” *Computer knowledge and technology: Academic Edition*, vol. 15, no. 9, pp. 308–310, 2019.
- [30] B. Yang and H. Gao, “Visual interaction design of digital learning resources,” *Journal of Gansu Radio and Television University*, vol. 16, no. 3, pp. 56–59, 2021.

Research Article

A Novel Bitcoin and Gold Prices Prediction Method Using an LSTM-P Neural Network Model

Xinchen Zhang ¹, Linghao Zhang ¹, Qincheng Zhou,² and Xu Jin ²

¹*School of Telecommunications and Information Engineering, Nanjing University of Posts and Tele-Communications, Nanjing 210046, China*

²*School of Science, Nanjing University of Posts and Telecommunications, Nanjing 210046, China*

Correspondence should be addressed to Xinchen Zhang; b19011721@njupt.edu.cn and Xu Jin; jinxu@njupt.edu.cn

Received 15 March 2022; Revised 12 April 2022; Accepted 12 April 2022; Published 5 May 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Xinchen 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.

As a result of the fast growth of financial technology and artificial intelligence around the world, quantitative algorithms are now being employed in many classic futures and stock trading, as well as hot digital currency trades, among other applications today. Using the historical price series of Bitcoin and gold from 9/11/2016 to 9/10/2021, we investigate an LSTM-P neural network model for predicting the values of Bitcoin and gold in this research. We first employ a noise reduction approach based on the wavelet transform to smooth the fluctuations of the price data, which has been shown to increase the accuracy of subsequent predictions. Second, we apply a wavelet transform to diminish the influence of high-frequency noise components on prices. Third, in the price prediction model, we develop an optimized LSTM prediction model (LSPM-P) and train it using historical price data for gold and Bitcoin to make accurate predictions. As a consequence of our model, we have a high degree of accuracy when projecting future pricing. In addition, our LSTM-P model outperforms both the conventional LSTM models and other time series forecasting models in terms of accuracy and precision.

1. Introduction

The financial markets have a magical air that attracts a varied spectrum of investors to take part in the game of chance. In recent years, the rapid rise of financial technology and artificial intelligence has led to their widespread use across a broad range of sectors, including methods such as machine learning and deep learning, among others. When it comes to the interpretation and prediction of temporal data, deep learning algorithms have made significant strides. The use of quantitative algorithms is becoming more popular in conventional financial markets, such as futures trading and stock trading, as well as in other financial markets, such as digital currency exchanges. It is the systematic process of executing automated pre-programmed trading instructions while accounting for factors such as volume, price, and time in order to maximize profit and minimize risk. When it comes to trading in

the financial market, it employs complex mathematical algorithms to increase the efficiency of trading choices. As a result of using sophisticated algorithms in combination with mathematical models and human supervision, it makes trading decisions on exchanges. Furthermore, according to the Financial Times, quantitative trading is increasingly being utilized to trade digital currencies (e.g., Bitcoin) in addition to the traditional hedge fund and futures trading. Numerous quantitative strategies have benefitted from the high volatility of digital currencies, which has resulted in an increase in the rate of return. In general, we believe that a quantitative strategy involves a thorough examination of market characteristics, market sentiment, and other varied information in order to obtain maximum returns with relatively tolerable risk. Based on research, according to the study, the Algorithmic Trading Market was valued at USD 11.66 billion in 2020 and is predicted to expand to USD 26.27 billion by 2028, growing

at a compound annual growth rate (CAGR) of 10.7 percent between 2021 and 2028.

According to our beliefs, the price movement of an investment is the consequence of a dynamic game in which a number of different people participate, and we think that although the market is nonlinear, there are some patterns that may be seen. We might infer from the principles that underpin technical analysis that history is a repeating cycle. Numerous time series of the financial market have been discovered to include stylized facts of the financial market, and there are many stylized facts of financial market indexes to be found [1, 2] as well. As a result, in this paper, we will discuss ways for forecasting market prices.

The time series analysis method is one of the quantitative forecasting approaches that are now accessible on the market. The creation and inference of statistical models, as well as the most optimal time series prediction, control, and filtering, are all discussed in detail. Time series analysis may be separated into two forms: deterministic change analysis and random change analysis. Deterministic change analysis is the more common of the two types. Trend change analysis, cycle change analysis, and random change analysis are all types of deterministic change analysis, while random change analysis includes AR, MA, and the ARMA model. Some forms of random change analysis are covered in this section of the manual.

Traditionally used time series forecasting models, such as the ARIMA [3] and Vector Autoregressive Model (VAR) [4], have some drawbacks due to the fact that the investment market is a nonlinear nonhomogeneous system. Machine learning, on the other hand, thanks to its improved performance, can carry out efficient excavation in order to get prospective information from the market. In recent years, the application of machine learning in market forecasting has grown in importance, and it has largely exceeded the prediction accuracy of classical time series forecasting models in terms of accuracy.

Price prediction models based on neural networks, such as recurrent neural networks (RNN) [5], convolutional neural networks (CNN), multilayer perceptron (MLP) [6, 7], and long short-term memory (LSTM) [8], have proven to be the most widely used among machine learning approaches in recent years. For the period between March 11, 2014, and March 31, 2019, Soylemez used a multilayer artificial neural network technique to estimate gold prices, utilizing parameters such as Brent oil prices, the VIX index, the Dow Jones index, and the US Dollar index [9]. In order to forecast the stock prices of selected companies, a long short-term memory model is utilized to analyze the daily stock price movement and returns of different sectors based on their prior values [10]. Using a DAE LSTM model [11], Sanghyuk's study suggests that the proposed approach may be used to predict future stock prices. A short-term memory network (LSTM) on the basis of a convolutional neural network has been added to considerably improve the prediction capacity of gold volatility. However, despite the fact that deep learning has shown substantial success in time series prediction, particularly in the financial industry,

building and selecting the best-suited strategy for a researcher is a time-consuming and challenging procedure.

Due to the fact that prices are impacted by a wide range of macro- and microeconomic variables (for example, investor attitude), price changes may be quite severe. The process of dramatic fluctuations generates a huge quantity of noisy data, which is also created throughout the process. As a result, when a single machine learning model is used to anticipate prices, there is a significant amount of prediction error. As a result, we wish to increase the accuracy of predictions by the combined usage and refinement of models, as well as by introducing some novel concepts (e.g., LSTM-P).

The following are the particular contributions that we have made:

- (1) Achieve noise reduction by the application of the wavelet transform: this paper presents a novel noise reduction strategy based on the wavelet transform that is used to smooth the fluctuations of price data and increase the accuracy of future forecasts in order to reduce the impact of high-frequency noise components on pricing.
- (2) Develop an LSTM prediction model that is as accurate as possible (LSPM-P): to anticipate future gold and Bitcoin prices, we create an upgraded LSTM prediction model (LSPM-P) and train it using past gold and Bitcoin price data. This is the second stage. If we look at the conclusion of our model, we can see that it has a high degree of accuracy when it comes to anticipating future prices. As previously stated, our LSTM-P model outperformed both conventional and other time series forecasting models in terms of accuracy and precision, as well as other time series forecasting models.

2. Related Works

2.1. Long Short-Term Memory. It was Hochreiter and Schmidhuber who first proposed the long short-term memory (LSTM) neural network in 1997, and it has subsequently gained widespread acceptance. Afterwards, it was fine-tuned over a period of many years. Since the introduction of the "gate" component into the LSTM framework, it has become more popular. Using the "gate" structure, it is possible to choose the most effective feature for processing from a huge number of characteristics and achieve the aim of controlling the flow of information [12]. LSTM neural networks include numerous neurons called storage units in the hidden layer, and each storage unit has three "gates." This class of "gates" is referred to as forgetting gates (f_t), input gates (i_t), and output gates (o_t). One of the capabilities of an LSTM neural network is the ability to control the transmission of input data; additionally, the independence between the memory storage unit output and result output can be maintained, allowing the sequence to retain important information during transmission as well as the ability to retain longer-term memories, among other things.

Thus, it is advised that the LSTM neural network can be used for time series prediction in the financial industry.

The following sections detail the processes involved in training an LSTM neural network. First, we can figure out what the network topology and loss function of the LSTM are. Second, initialize each parameter, and then, using the loss function, determine the accuracy rate of the model as a percentage. If it is unable to achieve the required precision, we will need to make changes to the settings. The weight and the size condition for updating bias terms should be that they minimize the loss of the stated objective function in the training sample, which is referred to as the pass loss in the training sample. When comparing and contrasting parameters, the function collects the gradient information, which is subsequently used to update parameters in accordance with the model's learning rate. As a general rule, we shall choose the most appropriate optimization procedure to update the gradient value. Consequently, when the gradient achieves the needed accuracy, the parameters of the model can be computed, the LSTM model can be finished, and the model may be used for prediction or classification.

Because of its memory units and "gates," the LSTM is capable of dealing with issues that the RNN was unable to handle. It can also carry out long-term memory storage of essential information in a quick and efficient way, and it can learn successfully via these mechanisms. As a consequence, a substantial amount of research has been conducted on financial time series modeling as a result of these developments. In spite of the fact that the neural network's revenue distribution closely mirrored that of Bitcoin, Wang et al. concluded that more advanced learning algorithms such as deep cyclic neural networks, RNN, and LSTM may give superior forecast accuracy [13]. When estimating Bitcoin, Wu et al. compared the ARIMA time series model with the LSTM deep learning model, which she found to be superior. As a result, the average absolute error of the LSTM model was much lower than that of the ARIMA model. The results show that the LSTM model was significantly accurate in forecasting the price of special currency [14]. The sophisticated temporal aspects of machine learning methodologies led Marendra et al. to discover that cyclic neural networks and long short-term memory neural networks were more accurate at forecasting Bitcoin prices than traditional multilayer perceptron (MLP) neural networks [15, 16]. This study is built on an LSTM network as a basis, since it is effective at learning the correlation of time in a sequence and has a broad variety of applications in the field of time series prediction.

2.2. Attention Mechanism. The amount of information that a neural network is capable of storing is referred to as the network capacity of the particular neural network under consideration. The increase of the number of neurons in the network will lead to the corresponding increase in network capacity. Because the capacity of a network increases with its size, the number of neurons increases as well, making the network more complex overall. As a consequence, the bigger the number of parameters in the neural network, the more

complex the network. It is not suggested to increase the complexity of the model to improve its expressiveness due to the limitations in the processing capability of the computer. The outcome has been the development of an attention mechanism that is inspired by the mechanics of the human brain, which may be used to cope with the challenges associated with information overload. It is no longer necessary for the attention mechanism to pay attention to data characteristics from all areas; instead, it only needs to pay attention to data characteristics from critical locations, resulting in significant time savings. In addition to saving a substantial amount of time by removing an enormous number of pointless calculations, this may improve the accuracy and generalization of the model by increasing its generalization and accuracy [17].

According to the attention mechanism, data with dynamic change characteristics may be captured more efficiently, resulting in a more accurate correlation analysis. It is suggested that the attention method should be introduced into the prediction of time series data prediction. Wang et al. used the LSTM model, which is based on attention, to forecast the day when the SSE 50 daily closing price will be reached. In this study, he looked at a single-factor LSTM model that did not include the attention mechanism and contrasted it with the conventional LSTM model. Aside from that, he looked at two other versions of the attention mechanism, one with and one without the decoding process, and he compared them to one another. Although the attention-LSTM model without the decoding process trained more quickly than the model with the decoding process, the prediction accuracy of the model without the decoding process was not as good as that of the model with the decoding process [18]. Wang et al. proposed a random recursive network called the CLVSA model. Researchers claim that their technique may be used to predict likely variations in raw financial transaction data. Based on deep LSTM and attention mechanisms, Zheng and Xu have developed a financial data forecasting strategy that is based on deep LSTM and attention mechanisms [19]. This study used daily data as well as time-sharing data to investigate the influence of capital flow variations on stock trend changes, with the finding that the self-attention model is enhanced as a consequence. As a result of the experiments, the suggested technique was able to improve the accuracy of trend judgment to 63.04 percent and gain 6.562 percent in the two-month backtest experiment, demonstrating that the model has a certain level of efficacy and practicability in the prediction of stock price trends. LSTM has encountered a number of obstacles in the area of financial time series prediction. Improving the LSTM model is essential for producing more accurate results in the highly dynamic field of investment forecasting. According to the article, this model is referred to as the LSTM-P model.

3. Bitcoin and Gold Prices Prediction

3.1. Data. The data employed in our model for the empirical application consist only of historical price series for Bitcoin and gold, which were collected between September 2016 and

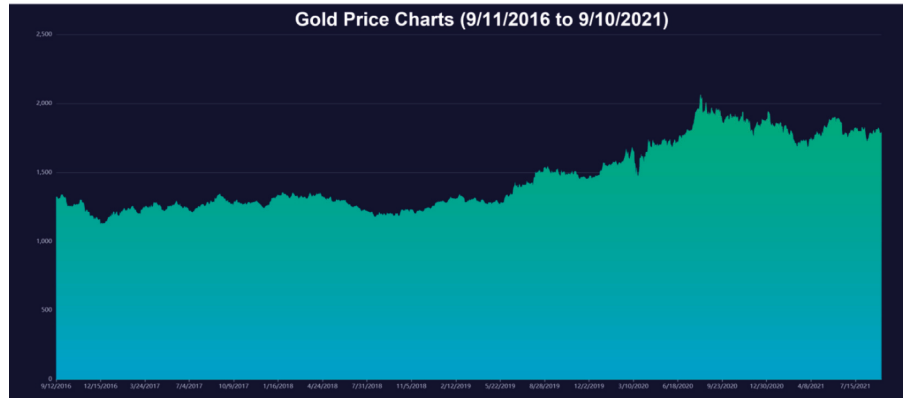


FIGURE 1: Gold daily price.

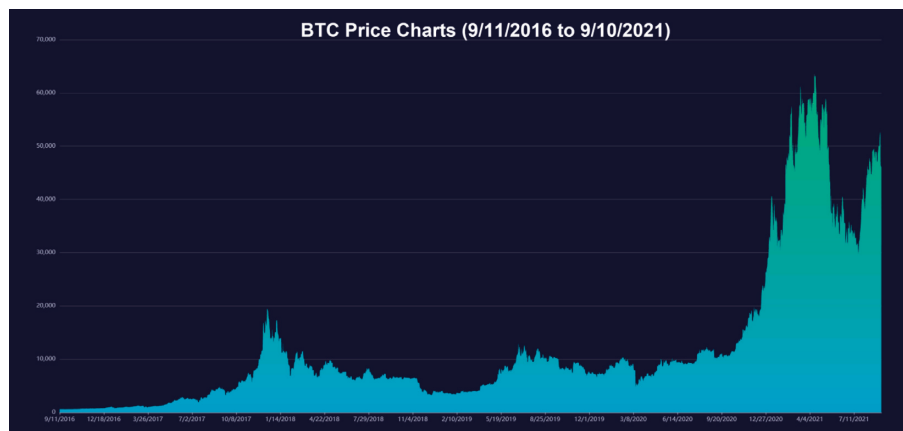


FIGURE 2: BTC daily price.

September 2021. Gold daily prices (in US dollars per troy ounce) are sourced from the London Bullion Market Association, while Bitcoin daily prices (in US dollars per Bitcoin) are sourced from the Nasdaq Stock Market. Take a look at Figures 1 and 2.

Note: Bitcoin can be traded every day, so the data are continuous. However, gold has a difference between trading days and nontrading days and the data are not continuous. We first do a smoothed interpolation of the historical price of gold to facilitate the forecasting model. Moreover, we consider the nontrading day scenario in the trading strategy.

3.2. Flow of Our Work. Considering the background information and restricted conditions, the article consists of three parts. The first one is data preprocessing part. We use interpolation fitting and wavelet transform noise reduction for Bitcoin and gold historical price data, in order to get higher accuracy in the later time series prediction. Then, we use a modified LSTM-Plus (LSTM-P) neural network for training and prediction. LSTM-P is characterized by keeping only one control gate in the original LSTM model and adding cellular connections to the candidate hidden states and control gate to improve prediction accuracy. At last, we make a sensitivity analysis. In order to avoid complicated

descriptions and intuitively reflect our work process, the specific flowchart of the full article can be referred to in Figure 3.

For the purposes of this work, Python 3.8 is used for data preparation and processing, with the programs NumPy, TensorFlow, and Keras serving as support. These tools are used to complete the model construction process. In addition, we utilize SPSS25 to evaluate the data.

3.3. Model Preparation

3.3.1. General Assumptions and Justifications. To make the issue easier to understand, we make the following fundamental assumptions, each of which is supported by appropriate evidence.

Assumption 1. Prices reflect all information.

Justification: this is the basic principle of technical analysis. Whether it is fundamental information or investor behavior and sentiment, it is ultimately reflected in price changes. Investors' reasons and thinking for making trading strategies are like the hidden layer of a neural network. Thus, we have no need to overthink and we just need to know the results, which is one of the reasons why we choose a neural network to forecast.

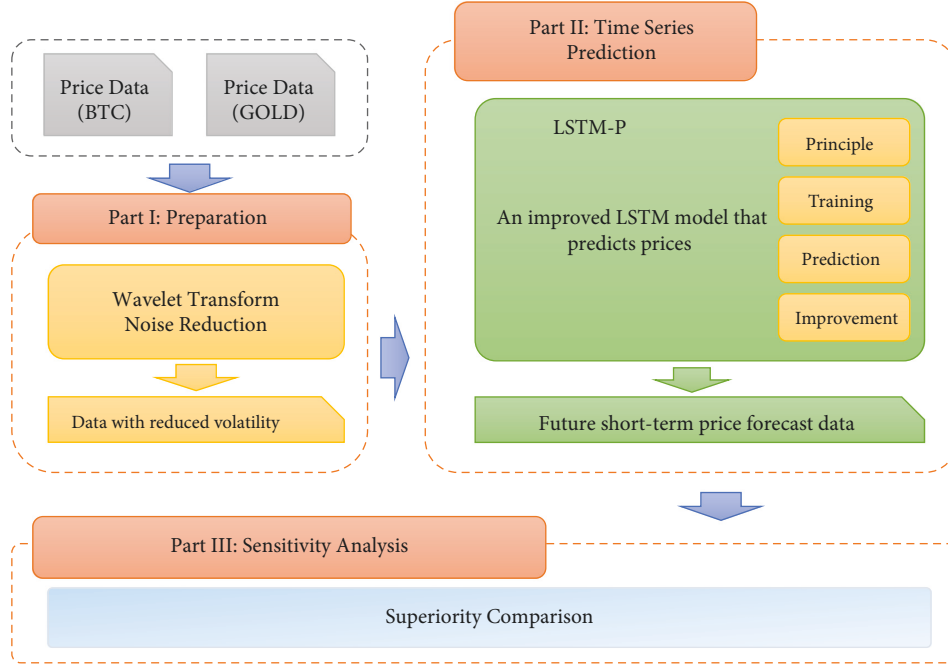


FIGURE 3: Flowchart of our work.

Assumption 2. We do not consider high-frequency trading and shorting trading.

Justification: there are high transaction costs, and we only have daily price data, so we only consider swing trading and the trading frequency will not be too large. Moreover, there is no shorting mechanism, only spot buy and sell operations are allowed.

3.3.2. Data Noise Reduction. Bitcoin and gold price data are trend-like time series data, and we usually differentiate the data first. However, because of various factors, the volatility of gold and Bitcoin prices is very sharp (e.g., Bitcoin), leading to large price fluctuations in the short term and generating a lot of white noise. Therefore, the different methods will not be good for the trend prediction results. We need to perform volatility smoothing on the time series of prices to some extent, so that the trending characteristics can be retained and the noise is reduced.

If you are looking for a method that can be used for time-frequency analysis, wavelet analysis is one that you should consider. Data scaling and panning are used in this method; as a result, it is capable of examining both time-domain and frequency-domain components of a signal at the same time. The wavelet technique [6] may be used to split a noisy signal into numerous scales and denoise the signal while keeping its integrity. This is true regardless of the frequency content of the signal. As a result of its adaptive characteristics, wavelet analysis is particularly well suited for problem signals that are smooth and nonlinear in nature. Liu et al. used the wavelet transform to their data in order to reduce the amount of noisy data they had. Mallick et al. proposed a novel approach for increasing the prediction accuracy of noisy multivariate time series that are based on the wavelet-denoising algorithm and multiple echo state networks.

When it comes to precision, it exceeds both the Fourier transform and the window Fourier transform together.

If a time series function $f(t) \in L^2(\mathbb{R})$ and when the following conditions are met,

$$C_W = \int_0^{\infty} \frac{|\psi(\omega)|^2}{\omega} d\omega < \infty, \quad (1)$$

where $\psi(t)$ is the mother wavelet function and $\psi(\omega)$ is the Fourier transform of the mother wavelet function, then we can get

$$W(a, \tau) = \int_{-\infty}^{+\infty} f(t) \bar{\psi}_{a,\tau}(t) dt, \quad (2)$$

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} a^{-2} W(a, \tau) \psi_{a,\tau}(t) da d\tau,$$

where the displacement and scaling of the fundamental wavelet is

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right), \quad (3)$$

where a is the scale factor, τ is the translation variable, and equation (3) is the continuous wavelet transform. Next, the signal is decomposed (equation (4)) and reconstructed (equation (5)):

$$c_k^j = \sum_n H(n-2k) c_n^{j-1},$$

$$d_k^j = \sum_n H(n-2k) d_n^{j-1}, \quad (4)$$

$$c_n^{j-1} = \sum_n h(n-2k) c_k^j + \sum_n g(n-2k) d_k^j. \quad (5)$$

The low-pass filter is $H(n)$, and the low-pass impulse response is $h(n)$. The high-pass filter is $G(n)$, and the high-

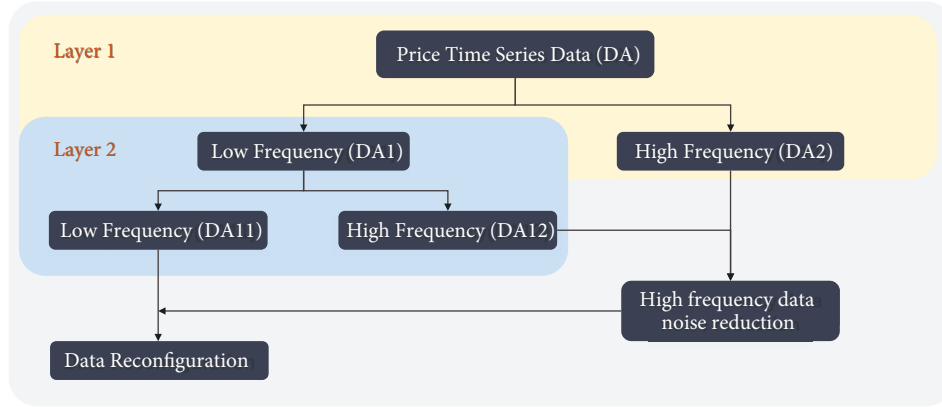


FIGURE 4: Two-layer wavelet noise reduction schematic.

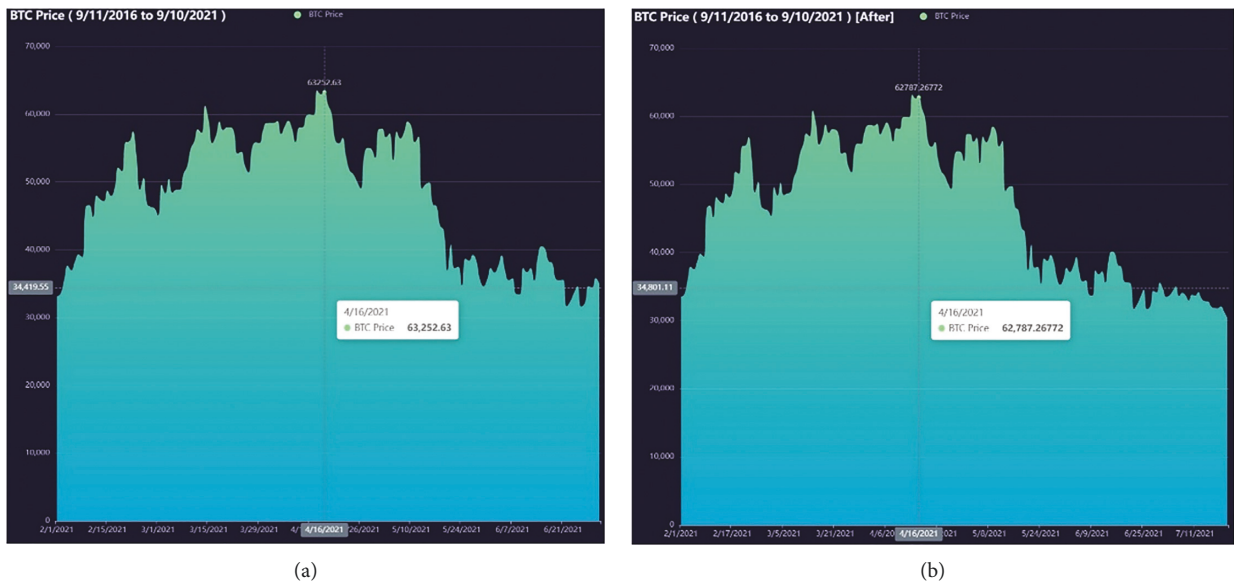


FIGURE 5: BTC before and after noise reduction. (a) Example of data before noise reduction. (b) Example of data after noise reduction.

pass impulse response is $g(n)$. The low-frequency signal is c_k^j , which contains less noise, and the high-frequency signal is d_k^j , which has more noise.

Because the deconstructed high-frequency data include practically all of the noise, the high-frequency signal must first be denoised according to the threshold value and then the low-frequency and denoised high-frequency sections must be rebuilt from the low-frequency data.

The reconstructed data are the denoised data, which finally complete the whole denoising process. Figure 4 shows the whole process (taking 2-layer decomposition as an example).

On the basis of the foregoing, it can be concluded that, because the trending characteristics of the low-frequency part of the price time series data are the most important information, it is possible to remove noise from the data by performing noise reduction on the price time series data in order to improve the accuracy of prediction by the neural network model by using neural networks.

3.3.3. Data Analysis after Preprocessing. We choose db8 wavelets in the original signal decomposition process and performed a 6-layer decomposition of the signal. The figures below show the data after noise reduction by wavelet transformation (using Bitcoin as an example). We find that the processed data have a smoothing effect in time regions with high short-term fluctuations. Small fluctuations have been largely eliminated and almost no longer contain noisy data. Figure 5 shows the comparison of data before and after processing in a time region with high Bitcoin fluctuations. Figure 6 shows a before-and-after comparison of all Bitcoin data.

The error rates before and after data processing are shown in Figure 7.

Taking Bitcoin as an example, we find that the average error rate of the data after noise reduction by wavelet transform is 0.03. We then differentiate the data before and after processing to get the increase and decrease of each day relative to the previous day. Calculating their statistical

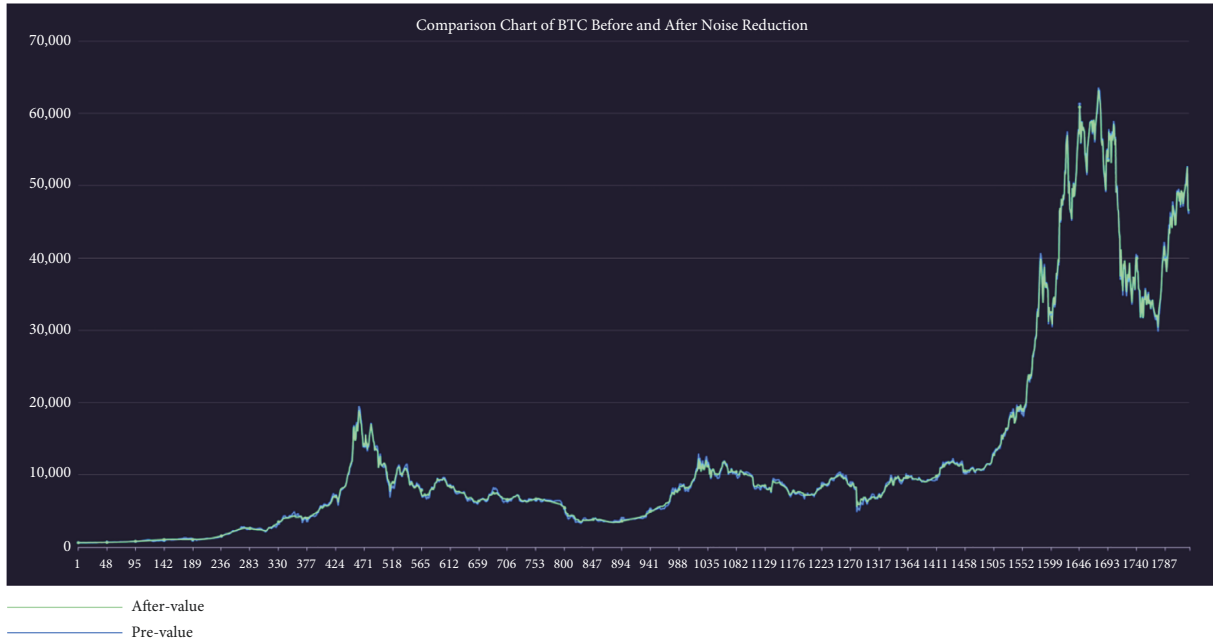


FIGURE 6: BTC before and after noise reduction.

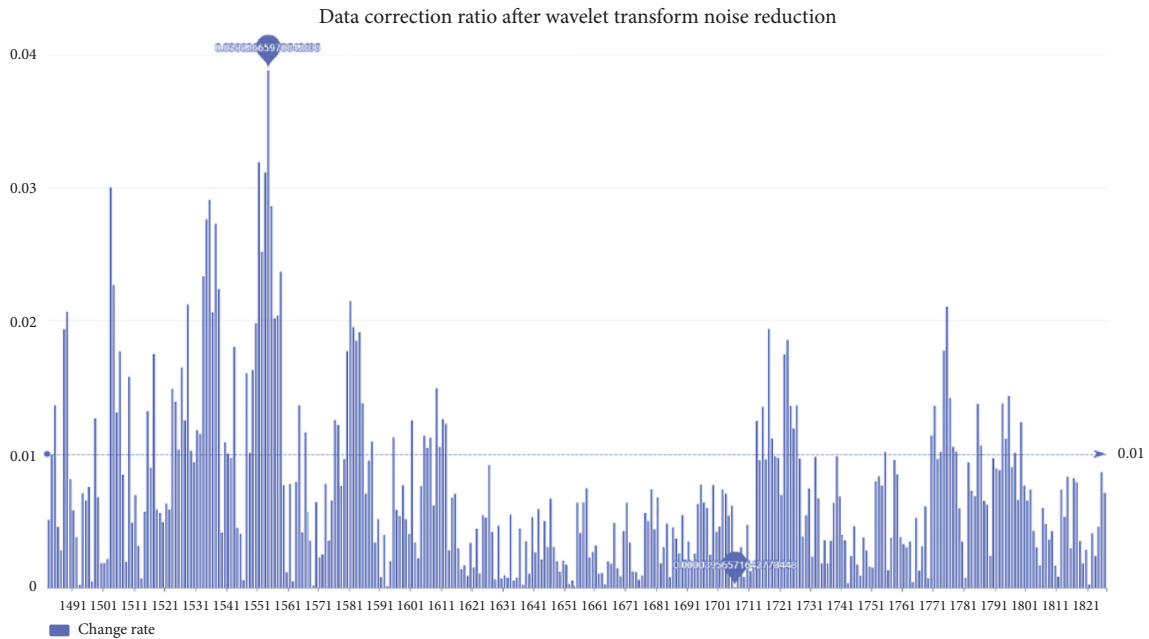


FIGURE 7: Error rate of prices data.

TABLE 1: Statistical characteristics of the data before and after processing.

	Mean	Variances	MAD	SNR	MAD
Meta data	0.00324	0.00172	0.02665	—	—
Processed data	0.00261	0.00074	0.01532	110.19721	0.40392

features (Table 1), we find that the variance of the processed data is much lower, achieving the effect of smoothing fluctuations. In addition, the larger the value of the signal-to-

noise ratio (SNR), the smaller the noise of the data after noise reduction, and the smaller RMSE indicates that the noise reduction process changes the trend of the original data less.

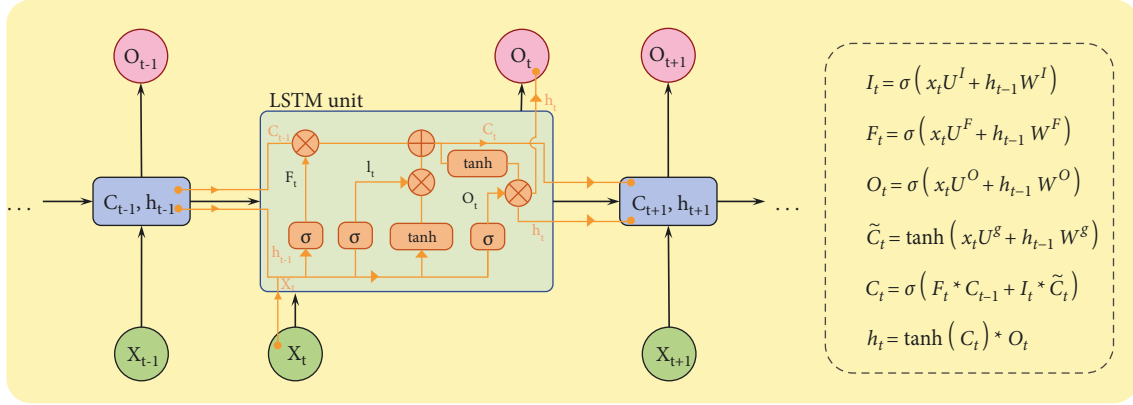


FIGURE 8: Structure of a typical LSTM storage cell.

3.4. Time Series Prediction (LSTM-P)

3.4.1. Description of the Model. When it comes to time series prediction, it is widely accepted that LSTM (long short-term memory) neural networks do very well in this area. A higher memory function than other creatures may be found in its cell structure. LSTM networks filter input in order to maintain and update the state of storage cells. This is accomplished via the use of gating mechanisms. Gating in the model may be divided into three types: in-going gates, forgetting gates, and output gates. The storage cell has three sigmoid layers and one tanh layer, and the model also has three sigmoid layers and one tanh layer, with three sigmoid levels and one tanh layer. LSTM storage cells are shown in Figure 8 as having a common structural layout.

LSTM models are very advanced when it comes to time series prediction, but they still have a number of shortcomings that must be addressed. The training efficiency and prediction accuracy of a model will both suffer if the number of layers in the model is excessive, as in the case of a multilayer model. Because of this, we are striving to strengthen the existing model by adding both the mature model and some other superior models [20, 21], which may not only simplify the model but also increase prediction accuracy as a result of the improved models.

We create a new LSTM model called LSTM-Plus (LSTM-P), which has been proven to improve the prediction accuracy. LSTM-P has the following features:

- (i) Coupled input gate and forgetting gate.
- (ii) The previous moment cell state is added to the gating of the coupled input and forgetting gates. That is, the more important part of the model is the candidate's hidden state of the previous moment, which allows the gating to more accurately select the information that needs to be input versus forgotten.
- (iii) A constant column vector β is subtracted before the coupled gating activation (the value of β depends on the predicted dataset, usually set to 1). This will allow accumulating slightly more information than the amount of information forgotten, making it easier to analyze the time series.

Figure 9 depicts the structure of the enhanced LSTM-P model.

The neurons of the improved LSTM-P model are specifically interpreted as follows:

$$\begin{aligned}
 f_t &= (W_f [c_{t-1}, h_{t-1}, x_t] + b_f), \\
 \tilde{C} &= \tanh(W_c [h_{t-1}, x_t] + b_c), \\
 C_t &= \sigma(f_t) * C_{t-1} + (1 - \sigma(f_t - \beta)) * \tilde{C}_t, \\
 h_t &= C_t.
 \end{aligned} \tag{6}$$

3.4.2. Training of Our Model. We train the Bitcoin and gold price data separately and do incremental training on each other. We divide the dataset in a 4:1 ratio, with a total of 1826 pieces of data. We use the first 50 days of prices to predict the last 5 days of prices. After determining the input data for the model, we train and predict a 2-layer LSTM-P model.

Several parameters were changed during the process, such as the epoch used to compare the loss function and the value of the overfitting dropout function. In addition, we do a comparison between the LSTM-P model and the LSTM model prediction results.

For our final fixed epoch is equal to 30, as shown in Figure 10, the loss function is less than 0.05 and the final MAPE is less than 0.05.

In addition, the comparison shows that the LSTM-P model we use has a faster convergence rate of the decreasing loss function than the general LSTM model. When the dataset is not denormalized, the value of the loss function (MAPE) of our LSTM-P model is 0.0481 after 30 training sessions, which is better than that of the LSTM at 0.0608 (Table 2).

3.4.3. Results of Our Model. Figures 11 and 12 show the predicted data for Bitcoin and gold prices compared to the source data. We define an indicator error rate (ER, equation (7)) to compare the degree of error of the predicted data to the original data.

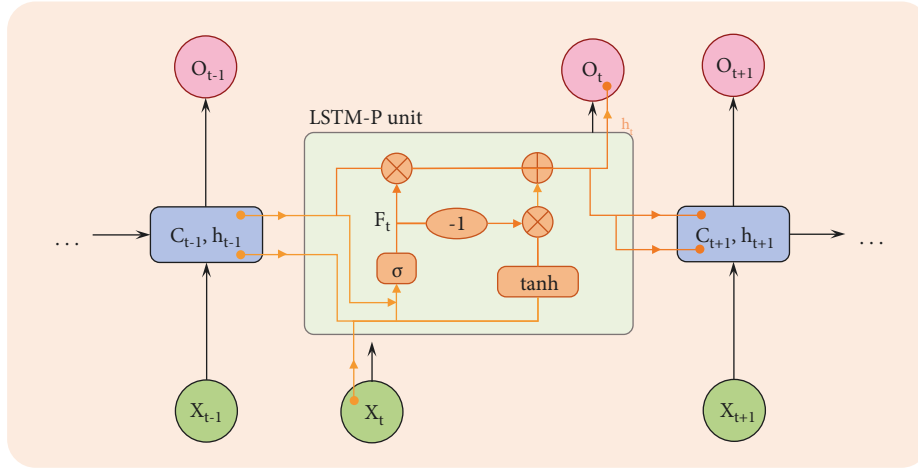


FIGURE 9: Structure of the LSTM-P model.

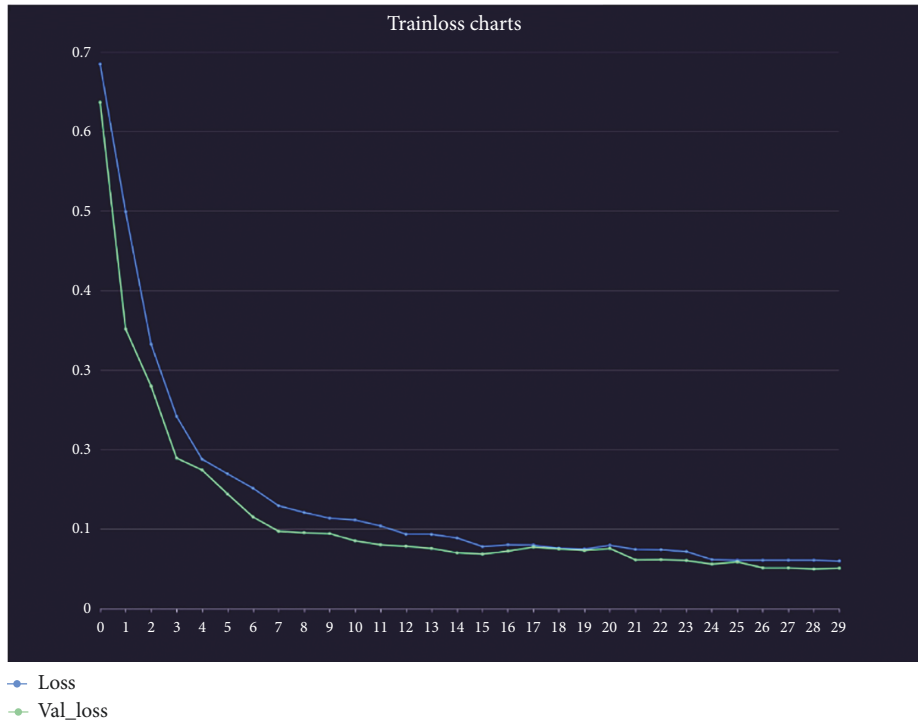


FIGURE 10: Train loss charts.

TABLE 2: Statistical characteristics of the data before and after processing.

Model	MSE	RMSE	R^2	MAPE (%)
LSTM	1.3539	1.1687	0.7973	6.08
LSTM-P	0.9453	1.0069	0.8862	4.81

$$ER = \frac{(\text{value} - \text{metavalue})}{\text{metavalue}} \quad (7)$$

We calculate the error rate and depict it with scatter plots in Figures 11 and 12.

Figures 13 and 14 show the forecast data for Bitcoin and gold for some time periods.

Based on the error rate analysis and the results of prediction above, we conclude that our prediction model is valid. We find the following conclusions:

- (i) The average error rate for gold price prediction is 0.0094, which is smaller than Bitcoin's 0.0382.

Note: this is also in line with common sense. As Bitcoin's volatility is much greater than that of gold, even after our wavelet noise reduction, Bitcoin's price prediction still has more errors than gold's.

- (ii) Because of the limited amount of data available before training, the estimates for the two underlying investments exhibit large errors in the first few hundred days.

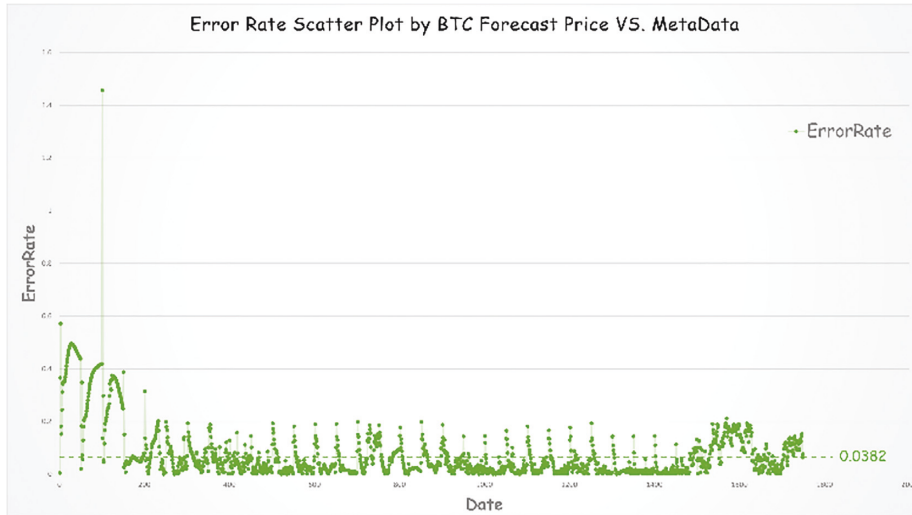


FIGURE 11: Error rate scatter plot by BTC forecast price vs. metadata.

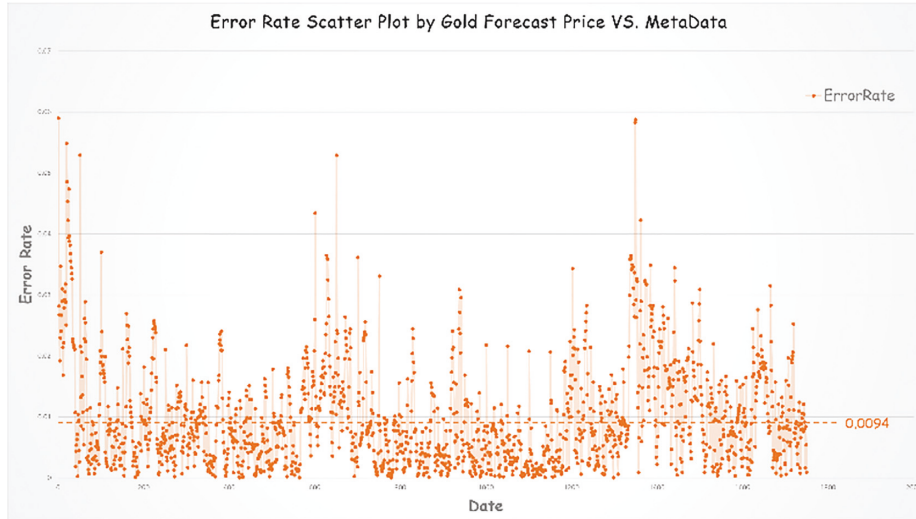


FIGURE 12: Error rate scatter plot by gold forecast price vs. metadata.

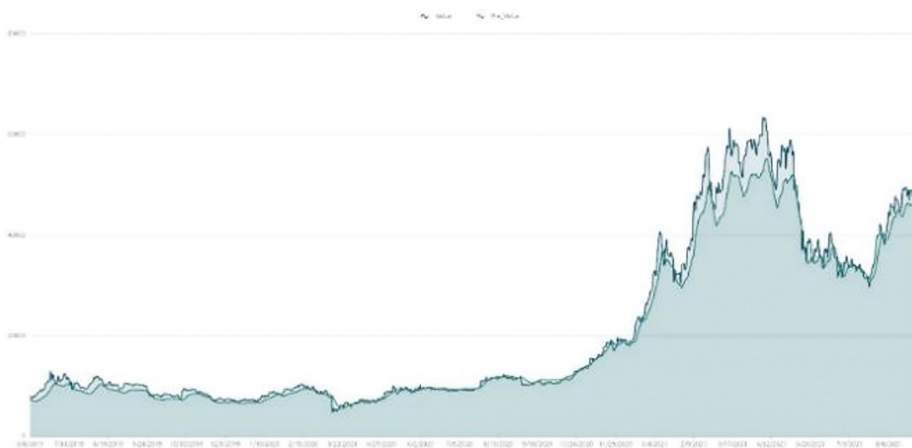


FIGURE 13: Forecast data for Bitcoin.

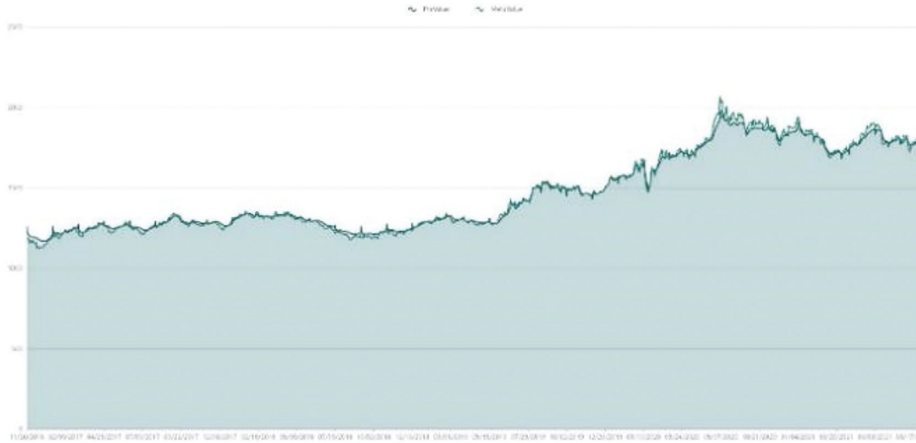


FIGURE 14: Forecast data for gold.

(iii) We backtest some of the time periods with large errors (i.e., time periods with large price swings). These periods tend to have prices that are out of character with previous movements of this investment variety.

Note: we feel that perhaps this is because of the excessive irrational emotional manipulation of investors at this time, causing large ups and downs. It is like a sudden change in a person's personality. These are the times when forecasting models are of little use.

4. Conclusions

The trained LSTM-P model predicts BTC and gold data with a somewhat high degree of accuracy for both BTC and gold. Several times during the training process, we made adjustments to the parameters (epoch, layer, and so on) in order to get a prediction model that was close to ideal. When compared to the general LSTM model, our model has been significantly optimized (explained above). If more optimization is required, the inclusion of an attention layer may be explored.

Data Availability

The labeled datasets 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 regarding this work.

References

- [1] N. Jung, Q. A. Le, B. J. Mafwele, H. M. Lee, S. Y. Chae, and J. W. Lee, "Fractality and multifractality in a stock market's nonstationary financial time series," *Journal of the Korean Physical Society*, vol. 77, no. 3, pp. 186–196, 2020.
- [2] W. B. Shi and P. J. Shang, "The multiscale analysis between stock market time series," *INTERNATIONAL JOURNAL OF MODERN PHYSICS C*, vol. 26, no. 6, 2015.
- [3] Y. Y. Deng, *Design of Stock Trading Strategy Based on WT-ILSTM-ATT Model*, Shanghai Normal University, Shanghai, China, 2021.
- [4] K. Kumar and M. T. U. Haider, "Enhanced prediction of intra-day stock market using metaheuristic optimization on RNN-LSTM network," *New Generation Computing*, vol. 39, no. 1, pp. 231–272, 2021.
- [5] B. X. Yong, M. R. Abdul Rahim, and A. S. Abdullah, "A stock market trading system using deep neural network," in *Proceedings of the Asian Simulation Conference* Singapore, August 2017.
- [6] C. Mallick, S. K. Bhoi, S. K. Panda, and K. K. Jena, "An efficient learning algorithm for periodic perceptron to test XOR function and parity problem," *SN Applied Sciences*, vol. 2, no. 2, p. 160, 2020.
- [7] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.
- [8] R. Zhang, C. Zhang, and M. Yu, "A similar day based short term load forecasting method using wavelet transform and LSTM," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 17, no. 4, pp. 506–513, 2022.
- [9] Y. Söylemez, "Prediction of gold prices using multilayer artificial neural networks method," *Sosyoekonomi*, vol. 228, no. 46, pp. 271–291, 2020.
- [10] D. Saravagi, S. Agrawal, and M. Saravagi, "Indian stock market analysis and prediction using LSTM model during COVID-19," *International Journal of Engineering Systems Modelling and Simulation*, vol. 12, no. 2-3, pp. 139–147, 2021.
- [11] S. Yoo, S. Jeon, S. Jeong et al., "Prediction of the change points in stock markets using DAE-LSTM," *Sustainability*, vol. 13, no. 21, Article ID 11822, 2021.
- [12] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a CNN-LSTM approach," *Expert Systems with Applications*, vol. 157, Article ID 113481, 2020.
- [13] X. Wang, J. Wu, C. Liu, H. Yang, Y. Du, and W. Niu, "Fault time series prediction based on LSTM recurrent neural network," *Beijing Airlines Journal of Aerospace University*, vol. 44, no. 4, pp. 772–784, 2018.
- [14] K. Wu, B. Peng, H. Xie, and S. Zhan, "A coefficient of variation method to measure the extents of decentralization for bitcoin

- and ethereum networks,” *International Journal on Network Security*, vol. 22, no. 2, pp. 191–200, 2020.
- [15] A. Marendra, T. Ramadhani, R. Kim, H. Lee, and E. Ryu, “Bitcoin price forecasting using neural decomposition and deep learning,” *Journal of the Korea Industrial Information Systems Research*, vol. 23, no. 4, pp. 171–180, 2018.
- [16] T. A. Dinh and Y. K. Kwon, “An empirical study on importance of modeling parameters and trading volume-based features in daily stock trading using neural networks,” *Informatcs*, vol. 5, no. 3, pp. 542–547, 2018.
- [17] V. Mnih, N. Heess, A. Graves, and K. Koray, “Recurrent models of visual attention,” *Neural Information Processing Systems*, vol. 18, no. 3, pp. 2204–2212, 2014.
- [18] L. Wang, T. Xue, H. Wang, and Z. Liu, “Research on stock price forecasting based on cyclic neural network,” *Journal of Zhejiang University of Technology*, vol. 47, no. 2, pp. 186–191, 2019.
- [19] S. T. Zheng and F. F. Xu, “Stock price trend prediction based on improved self-attention,” *Computer Technology and Development*, vol. 31, no. 3, pp. 33–38, 201.
- [20] V. D. W. Jos and J. Lasenby, “The Unreasonable Effectiveness of the Forget gate,” *Computer Science*, vol. 2018, pp. 11–49, 2018.
- [21] D. Wu, X. Wang, and S. Wu, “A hybrid method based on extreme learning machine and wavelet transform denoising for stock prediction,” *Entropy*, vol. 23, no. 4, pp. 1–30, 2021.

Research Article

A Novel Sentiment Analysis Model of Museum User Experience Evaluation Data Based on Unbalanced Data Analysis Technology

Xiang Chen ¹, Zhiwei Chen,² Lei Xiao,¹ and Ming Zhou¹

¹School of Design, Jiangnan University, Jiangsu, Wuxi 214122, China

²Graduate School Design, Dongseo University, Jurye-Ro, Busan 47011, Republic of Korea

Correspondence should be addressed to Xiang Chen; chenxiangsj@jiangnan.edu.cn

Received 25 March 2022; Revised 18 April 2022; Accepted 19 April 2022; Published 28 April 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Xiang Chen 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 development of virtual reality and digital reconstruction technology, digital museums have been widely promoted in various cities. Digital museums offer new ways to display and disseminate cultural heritage. It allows remote users to autonomously browse displays in a physical museum environment in a digital space. It is also possible to reproduce the lost heritage through digital reconstruction and restoration, so as to digitally present tangible cultural heritage and intangible cultural heritage to the public. However, the user's experience of using digital museums has not been fully and deeply studied at present. In this study, the user's experience evaluation data of digital museum are classified and processed, so as to analyze the user's emotional trend towards the museum. Considering that the user's evaluation data are unbalanced data, this study uses an unbalanced support vector machine (USVM) in the classification of user evaluation data. The main idea of this method is that the boundary of the support vector is continuously shifted to the majority class by repeatedly oversampling some support vectors until the real support vector samples are found. The experimental results show that the classification obtained by the used USVM has a good practical reference value. Based on the classification results of the evaluation data, the construction of the digital museum can be further guided and maintained, thereby improving the user experience satisfaction of the museum. This research will make an important contribution to the construction of the museum and the inheritance of culture.

1. Introduction

Museums play the role of cultural dissemination and entertainment in people's daily life and are deeply loved by the masses. However, traditional museums are limited in time and space, and their exhibition methods and contents are relatively simple. In addition, in order to watch the intangible cultural heritage exhibition, people need to go to the exhibition location to watch it, which will increase the cost of transportation, accommodation, and time, making people discouraged. In the era of information technology development, the development and application of virtual reality technology provide a new way to solve the above dilemma. The digital museum built with virtual reality technology has impacted the inherent existing form and static display mode of cultural relics. Through digital image rendering and animation technology, the original appearance of the museum can be reproduced. Digital museums allow people to

walk into the museum without leaving their homes, with an immersive feeling. This method greatly reduces the time and capital costs for both the museum and the audience. Digital museums have become a direction that museums have been working on in recent years, with very broad prospects for development. The further development of digital museums in the future not only needs to consider user-centered construction issues but also needs to consider the effectiveness of cultural information dissemination. Therefore, establishing a reasonable and complete user experience evaluation system is an important research direction for digital museums. At present, most of the research on digital museums focus on two aspects. One is the research on the construction and management of digital museums. The other is the research on the dissemination methods, channels, and dissemination effects of cultural information in digital museums. These two research complement each other and influence each other. The quality of digital museum

construction will affect the way and effect of its dissemination. The user's experience of the museum can guide the optimization of the museum construction.

At present, the construction of digital museum mainly has the following problems: first, the advantages of digital technology are not fully utilized, and the content of digital museums remains consistent with traditional museums. Second, the visual design of the user interface is simple and lacks aesthetic and cultural implications. Third, the operation is not smooth due to technical reasons. In order to improve the user experience of virtual exhibition halls, scholars at home and abroad have conducted many studies. Reference [1] evaluated the experience of three groups of children when using virtual exhibition halls to teach children in schools. The research shows that exhibition areas, remote lectures by museum staff, and interaction with children are all integral to virtual exhibitions. Reference [2] develops a Java program of personalized virtual exhibition tour based on World Wide Web thinking, to test whether the communication and cooperation between users can enhance the user's interest in use. Reference [3] explores how to improve interactive behaviors to enhance the user experience by studying user feedback on the "Keys To Rome" interactive virtual exhibition. Reference [4] introduces user experience into the research of virtual exhibition platform and puts forward the relevant theories of virtual exhibition platform research based on user experience through eye-tracking experiments and user survey research. Reference [5] discusses the realization of user experience model and virtual exhibition prototype system based on a computer multimedia environment. Reference [6] proposes to interpret traditional Chinese paintings with commentary dubbing and ambient sound to enhance user experience. When the user watches Chinese paintings, the system can judge the user's position and viewing direction according to the user's actions and visual center points, so as to synthesize stereo sound and play it.

Based on the above research, it can be analyzed that the determination of the user experience evaluation model is very important. The researchers designed questionnaires according to the research content and processed the collected data, using NASA-TLX scale [7], SWAT scale [8], factor analysis method [9], principal component analysis method [10, 11], and analytic hierarchy process [12] to build a quantitative model of user experience. Reference [13] points out that the construction of user experience model is the basis for evaluating user experience and divides the model into two types: structural model and measurement model. Reference [14] proposes a structural model of user engagement. It analyzes six attributes of engagement through validity and exploration factors, namely, perceived usability, aesthetics, attention span, sensory relevance, innovation, and persistence and through the structural equation model to determine the relationship between the various properties. Reference [15] established a total of five evaluation models: the traditional user satisfaction model, the usability research linear evaluation model, the linear model, the S-value model, the joint model, and the disjunctive model. In order to optimize the construction of the

museum and improve the user's somatosensory experience, this study first designs a user questionnaire to collect user evaluation data. Second, considering that in the evaluation data, negative evaluations generally account for a small amount of the total data, so there will be a large gap in the magnitude of the two categories. Traditional classification methods are not competent for the classification of such data. This study introduces an imbalanced data processing method. For the collected and preprocessed data, this study uses the USVM algorithm to apply the sentiment classification of user evaluation data. Finally, based on the evaluation results obtained from the experimental analysis, the construction and optimization of the digital museum are guided.

1.1. Sentiment Classification Theory Model USVM. Various intelligent algorithms are often used in emotion analysis and other fields [16–18]. The core idea of used USVM is as follows: first, the original unbalanced data are divided into three areas: support vector area (SV), majority class nonsupport vector area (MNSV), and minority class nonsupport vector area (FNSV) according to the location. Second, the samples in the MNSV area and the FNSV area are denoised. Third, the oversampling process is repeated for a small number of samples that are misclassified in the SV area and partially correctly classified close to the decision boundary, until the training data set with the best test results is found. Finally, some samples in the MNSV area are randomly deleted.

When using the support vector machine (SVM) to classify unbalanced data, the interference of noise and the lack of true support vectors of minority classes are the main reasons for the poor classification effect of SVM on unbalanced data. Therefore, it is very necessary to remove noise samples and increase a few real support vector samples after classification. At the same time, in order to avoid the influence of sample differences between classes on the classification results as much as possible, it is necessary to selectively delete some majority class samples. The specific operation steps of each link are described as follows:

1.1.1. Denoising. When using SVM to classify unbalanced data, since the decision boundary will shift to the side with fewer samples, the noise samples located in the MNSV area are not necessarily real noise samples. In order to identify all the noisy samples in the MNSV region, it is necessary to obtain k -nearest neighbors for all the minority class samples located in the MNSV region. If none of the k samples belong to the minority class, it is considered as noise. The steps of denoising are as follows:

1.1.2. Oversampling the Samples in the FNSV Area. When classifying unbalanced data, the decision boundary of the classifier will be skewed, so that the obtained support vector samples are not necessarily all real support vector samples. Therefore, in order to obtain the most realistic support vector samples, it is necessary to repeat the

Input: training data X , the number of neighbors k ;
Output: SV area sample set S_{sv} , MNSV area sample set $S_{NSV-maj}$ after denoising, FNSV area sample set $S_{NSV-min}$ after denoising;
Step 1: The training data X is used to construct the SVM classifier;
Step 2: The data set X is divided into the sample set S_{sv} of the SV area after passing through the SVM classifier. The sample set X_{NSV-a} in the MNSV area and the sample set X_{NSV-b} in the FNSV area.
Step 3: Find all the minority class samples a_1 in X_{NSV-a} ; for each sample in a_1 , find the k samples closest to it. If all the k samples belong to the majority class, delete the sample, otherwise do not do it Process to get the processed sample set $X_{NSV-maj}$;
Step 4: Delete the majority class samples in X_{NSV-b} to get the sample set $X_{NSV-min}$.

ALGORITHM 1: A denoising algorithm.

oversampling process for the SV-classified misclassified samples and the minority class samples that are correctly classified and close to the decision boundary. The oversampling steps are as follows:

1.1.3. Undersampling the Samples in the MNSV Area. Since samples in different positions may have different effects on the results of the classifier, it is unreasonable that the difference between samples between classes only refers to the difference in the number of samples. In comparison, it is reasonable that the difference between the accuracy of the minority class and the majority class test sample is the difference between the samples. The steps of undersampling are as follows:

The execution flow of the USVM algorithm is shown in Figure 1

2. User Experience Sentiment Classification Based on USVM

2.1. Experimental Data and Evaluation Indicators. In order to verify the performance of the sentiment classification model used in this study as objectively and accurately as possible, this study first uses public datasets to conduct experiments. This study selects two datasets of SemEval 2014 task 4, both of which are English datasets. SemEval 2014 task 4 includes the Restaurant review dataset and Laptop review dataset, which are divided into the training set and test set. Both training set and test set contain positive, neutral, and negative sentiment polarities, which are respectively 1, 0, and -1. The descriptions of the two datasets are listed in Table 1.

The evaluation indicators used in the experiment are as follows:

$$\begin{aligned}
 \text{accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\
 \text{precision} &= \frac{TP}{TP + FP}, \\
 \text{recall} &= \frac{TP}{TP + FN}, \\
 F1 &= \frac{2 * TP}{N + TP - TN},
 \end{aligned} \tag{1}$$

where TP represents the number of samples whose positive sentiment polarity labels are correctly predicted as positive

by the model. TN represents the number of samples whose negative sentiment polarity labels are correctly predicted by the model to be negative. FP represents the number of samples whose negative class sentiment polarity labels are mispredicted as positive by the model. FN represents the number of samples whose positive sentiment polarity labels are mispredicted as negative by the model. N represents the total sample size.

2.2. Analysis of Experimental Results. In order to analyze the classification performance of the model used in this study on the evaluation data, the selected comparison models are mainly SVM [19], k-nearest neighbor (KNN) [20], logistic regression (LR) [21], and decision tree (DT) [22]. The experimental results of each model on the Restaurant dataset are shown in Table 2 and Figure 2. The experimental results of each model on Laptop review datasets are shown in Table 3 and Figure 3.

It can be seen from the experimental results on the Restaurant dataset that the USVM used in this study has the best classification performance. The reasons are as follows: first, the experimental results on the four indicators are higher than those of other classifiers. Second, the accuracy and precision obtained based on USVM are similar in size, which shows that the algorithm used in this study has good performance for the classification of positive and negative samples. This is very important for imbalanced data.

From the experimental results on the Laptop dataset, it can be seen that the classification performance of the USVM used in this study is significantly better than other classifiers. The performance of traditional SVM is relatively poor, because the imbalanced data are constructed in this study, so the traditional SVM algorithm has a poor classification effect on this kind of data set. Among the other classifiers, the performance of DT is relatively stable, but it is slightly lower than USVM as a whole, so this study finally chooses USVM to classify the collected museum evaluation data.

3. Sentiment Analysis of Museum User Experience Evaluation Data

First, we used a *Python* crawler to get the museum user review text and store it as a CSV file. Second, data preprocessing is performed on the comment text, such as mechanical deduplication, cleaning, word segmentation, and word vector transformation. Third, the USVM is used to

Input: SV area sample set X_{sv} , denoised MNSV area sample set XNSV-maj, denoised FNSV area sample set XNSV-min, test data D ;
 Output: training samples X_{tr} , real SVM sample set X_{sv} .
 Step 1: Initialize the sample, let $X4 = X_{sv}$.
 Step 2: Find all the minority class samples in the sample set X_{sv} and divide them into correct samples $X1$ and wrong samples $X2$.
 Step 3: The first δ samples in $X1$ close to the decision boundary and the samples in $X2$ form a new sample set X_j ; for each sample in X_j , use the SMOTE algorithm to oversample between the same class, generate $X3$, and add $X3$ into $X4$.
 Step 4: $X4$, XNSV-maj, XNSV-min form a new training set X'_{tr} , and the training set trains the SVM classifier to find the corresponding SSV.
 Step 5: Input the test data D , obtain the accuracy rate ACC1 of the minority class and the accuracy rate ACC2 of the majority class after classification, and calculate the corresponding G-mean at this time.
 Step 6: Compare the size of ACC1 and ACC2, if ACC1 is greater than or equal to ACC2, then terminate the operation; otherwise, continue the following operations;
 Step 7: Repeat steps 2 to 6 until ACC1 is greater than or equal to ACC2, select the training sample X'_{tr} corresponding to the largest G-mean during the period, then X'_{tr} is the required optimal data. Correspondingly, the support vector sample set X'_{sv} corresponding to X'_{tr} is the closest support vector set.

ALGORITHM 2: An oversampling method.

Input: training dataset X'_{tr} . Denoised MNSV area sample set XNSV-maj, denoised FNSV area sample set XNSV-min support vector sample set X'_{sv} , test data D ;
 Output: The final sampling result G-mean.
 Step 1: Train a new classifier with dataset X'_{tr} ;
 Step 2: Input the test data D , and get the accuracy rate ACC1 of the minority class and the accuracy rate ACC2 of the majority class after the test.
 Step 3: If ACC1 is greater than or equal to ACC2, end the operation; otherwise, go to step 4.
 Step 4: Calculate the number difference h between different types of samples in X'_{tr} ;
 Step 5: Randomly delete h majority class samples in XNSV-maj to get the deleted dataset X1NSV-maj.
 Step 6: X'_{sv} , X1NSV-maj, XNSV-min form a new SVM classifier. Input the test set into the classifier to get the final sampling result G-mean.

ALGORITHM 3: An undersampling method.

classify the data and obtain positive and negative text data. Finally, the semantic user satisfaction analysis of the comment text is carried out, and relevant conclusions and museum construction and maintenance strategies are drawn. The research framework is shown in Figure 4.

3.1. Data Acquisition and Preprocessing. The network questionnaire is mainly expanded from the aspects, as listed in Table 4. A total of 8,280 pieces of experience evaluation data of museum users were obtained through online surveys. Since there is a certain degree of noise in the texts crawled on the Internet, it is necessary to clean the collected comment texts to ensure more accurate analysis results. The comment text with less than 2 characters and duplicate content are deleted to ensure the availability and uniqueness of the comment text. After cleaning the comment text, a total of 8,000 valid comment texts were retained.

3.2. Sentiment Classification of Museum User Experience Data. Using the 10-fold cross-validation method, 6000 items in the dataset are used as the training set, and the remaining 2000 items are used as the test set to test the

sentiment classification performance of the museum evaluation data. The number of categories is 3, which represent three emotions: satisfaction, neutrality, and dissatisfaction. Table 5 lists examples of sentiment classification rule words. Table 6 lists examples of manual labeling.

Considering that in the user experience evaluation data of the museum, the data of negative evaluation are much less than the number of positive and neutral evaluations. In view of this situation, the USVM used in this study is suitable for classifying data. The classification results based on USVM are listed in Table 7.

It can be seen from Table 7 that the accuracy of the training set and test set of the algorithm is greater than 86%, and the classification effect is ideal.

3.3. Interactive Service Comment Text Extraction and Analysis

3.3.1. Interactive Comment Text Extraction. On the basis of the above classification, this part extracts comments under different sentiment classifications and sets the comments extracted from the “satisfied” text set as positive comments and the ones extracted from the “neutral” text set as neutral comments, and negative comments are extracted from the

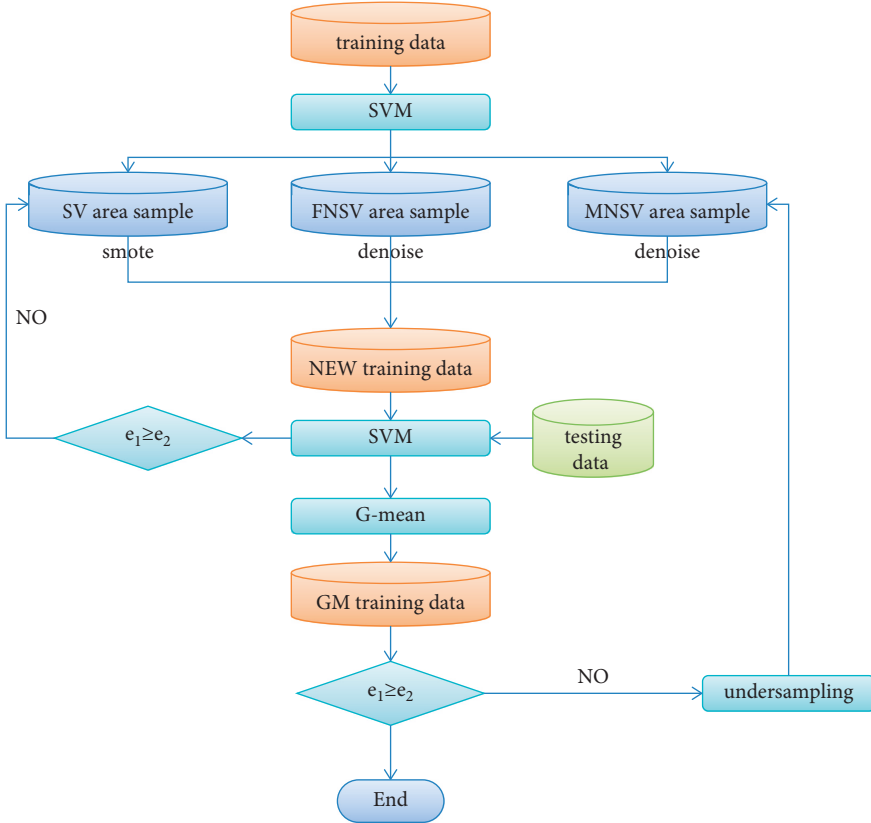


FIGURE 1: USVM algorithm flow chart.

TABLE 1: Experimental dataset.

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	637	196	807	196
Laptop	994	341	464	169	870	128

TABLE 2: Experimental results on the Restaurant dataset.

Model	Accuracy	Precision	Recall	F1
SVM	0.7745	0.8215	0.8477	0.8344
KNN	0.8098	0.8531	0.8129	0.8325
LR	0.8956	0.8606	0.8533	0.8569
DT	0.8258	0.8366	0.8050	0.8205
USVM	0.8757	0.8664	0.8566	0.8615

“unsatisfactory” text set. We analyze and summarize the feature words related to interactive reviews, as listed in Table 8

According to the constructed interactive comment vocabulary dictionary, each comment in the “satisfactory” comment text set, “neutral” comment text set, and “unsatisfactory” comment text set is traversed, respectively. This comment text sentence is then extracted. A total of 3200 related interactive evaluations were extracted. It can be seen from Table 9 that the positive description interaction accounts for 61% of the total comment text, which is 1952.

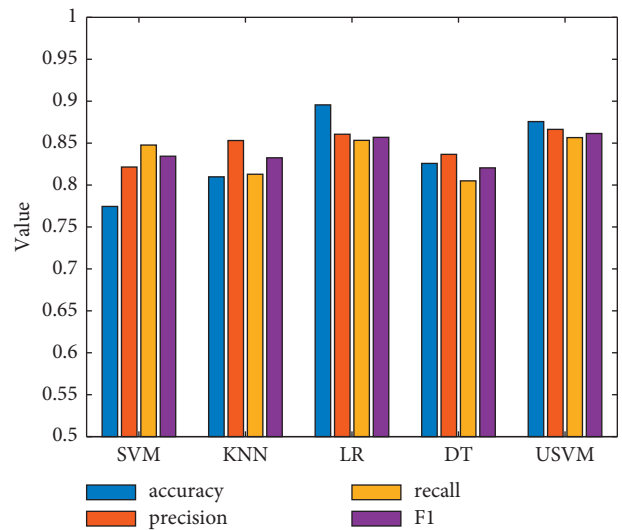


FIGURE 2: Comparison of classification results on the Restaurant dataset.

Negative texts accounted for 21% or 672. Neutral texts accounted for 18% or 576.

3.3.2. Recognition of Interactive Elements of User Attention Based on Word Frequency Method. This part uses the word frequency analysis method to analyze the word frequency of positive and negative logistics reviews. The “stutter” word

TABLE 3: Experimental results on Laptop.

Model	Accuracy	Precision	Recall	F1
SVM	0.6716	0.7521	0.6988	0.7245
KNN	0.7342	0.7652	0.8012	0.7828
LR	0.7365	0.7788	0.7569	0.7677
DT	0.7869	0.7952	0.7763	0.7856
USVM	0.8038	0.8335	0.8532	0.8432

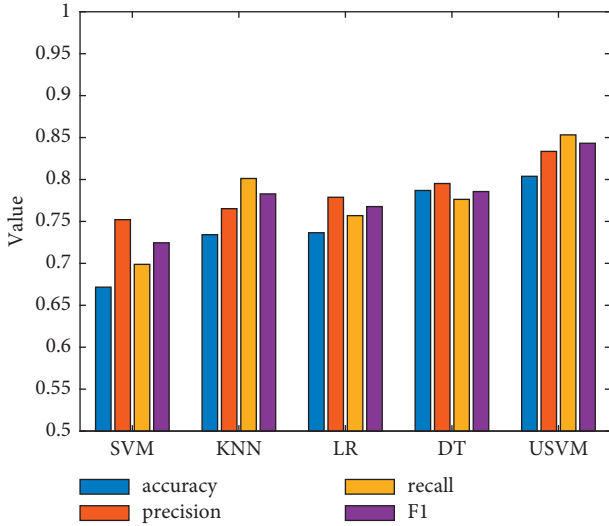


FIGURE 3: Comparison of classification results on the Laptop dataset.

segmentation tool of the *Python* language is used to segment the positive comment text and make word frequency statistics. The word frequency statistics are shown in Figure 5. In the top 10 word frequencies in Figure 5, 1 is convenient, 2 is clear, 3 is easy to use, 4 is interesting, 5 is smooth, 6 is rich, 7 is wonderful, 8 is novel, 9 is fast, and 10 is simple.

Based on the results shown in Figure 5, these are the factors that users pay more attention to in the museum experience process. Through word frequency statistics, it can be analyzed whether users pay more attention to whether the interaction is smooth and the operability is simple and clear.

Figure 6 shows a graph of the top 10 word frequency statistics of negative comment texts. Based on the results shown in Figure 6, it can be analyzed that users will generate negative comments due to the complex interactive interface and unclear feedback information. In the top 10 word frequencies in Figure 6, 1 is inconvenient, 2 is fuzzy, 3 is not easy to use, 4 is boring, 5 is laggy, 6 is monotonous, 7 is not exciting, 8 is not novel, 9 is too slow, and 10 is not fun.

3.3.3. Apriori-Based Analysis of the Impact of Interactivity on User Experience Emotion. In order to further mine the relationship between the interaction-related elements in the comment text and the user experience emotion, this part uses the Apriori algorithm based on association rules to mine the relationship between the interactive comment words in the user comment text and the user experience

emotion. Our purpose is to try to discover the interactivity-related vocabulary that affects user satisfaction and to further analyze the impact of interactivity on user experience emotions. Among them, the text data come from the “satisfied” and “dissatisfied” interactive comment texts extracted above. The words of user satisfaction and dissatisfaction come from the sentiment dictionary, as listed in Table 6.

Association rules can reflect the frequency of two itemsets appearing at the same time. The Apriori algorithm mainly generates association rules through the discovery of frequent itemsets [23, 24]. This part applies association rules to mine the co-occurrence of interaction-related words and satisfaction words in positive and negative comments. The metrics describing association rules include support (support) and confidence (confidence). When the support and confidence between itemsets are greater than the threshold, it is considered to have a strong association relationship. In the study, the support ($A \rightarrow B$) represents the probability of word A and word B appearing at the same time, the confidence ($A \rightarrow B$) indicates the probability that word A appears and word B also appears at the same time, and their expressions are as follows:

$$\text{support}(A \Rightarrow B) = P(A \cup B), \quad (2)$$

$$\text{confidence}(A \Rightarrow B) = P\left(\frac{B}{A}\right) \frac{\text{support}(A \cup B)}{\text{support}(A)}. \quad (3)$$

Lift ($A \rightarrow B$) is used to reflect the correlation between vocabulary A and vocabulary B. If lift > 1 , then the occurrence of A and B is positively correlated. If lift < 1 , then the occurrence of A and B is negatively correlated. If lift = 1, then A and B are independent of each other, and the expression is as follows:

$$\text{lift}(A \Rightarrow B) = \frac{\text{confidence}(A \cup B)}{P(A)P(B)}. \quad (4)$$

The study sets lift > 1 , support > 0.01 , and confidence $> \text{average}$ as the threshold to mine the association between words.

In this section, we apply *R* language to segment and extract the comment text. The extracted words are replaced by synonyms, and words such as “extremely satisfied,” “very satisfied,” and “satisfied” are replaced with “very satisfied” and “ok,” “okay,” and “good” are replaced with “good.” The related words such as smooth interaction and free operation are replaced with “good interaction” and converted into a transaction type data set for Apriori correlation analysis. The data processing results and visualization are listed in Table 9:

It can be seen from Table 9 that the words “easy to use,” “smooth,” “beautiful,” and “interesting” in the positive comments all have an effective correlation with the words “very satisfied” and “very good” that express satisfaction (lift > 1). It can be seen that the interactive function of the birth digital museum positively affects the user’s experience emotion, especially the interactive function represented by the above words has a high correlation with satisfaction.

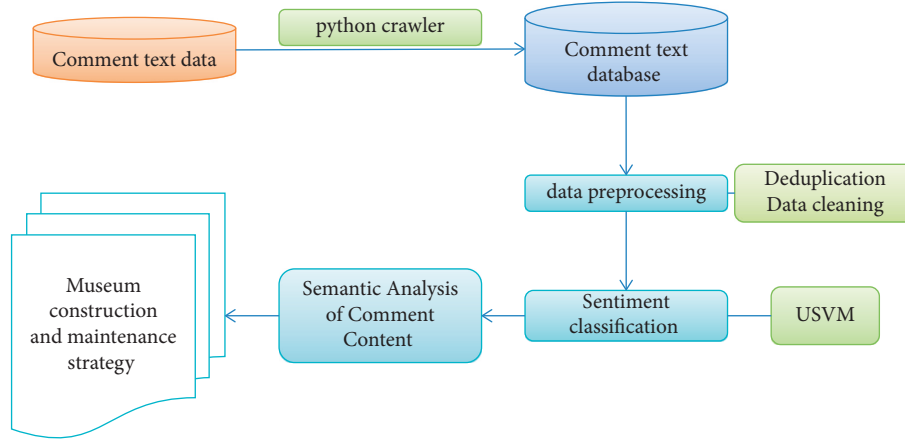


FIGURE 4: Sentiment analysis process.

TABLE 4: Questionnaire survey evaluation indicators.

First indicator	Secondary indicators	First indicator	Secondary indicators
Evaluation of visual	Color matching	Evaluation of interactivity	Ease of use
	Typesetting		Clear feedback
	Illustration design		Interactive entertainment
	Navigation design		Freedom of operation
	Page layout		Interactive visual consistency
	Button shape		Simple and clear operation
Evaluation of the function	Display function	Evaluation of information content	Accuracy
	Comment function		Content validity
	Interactive function		Amount of information
	Research function		Emotional support
	Entertainment function		Relevance
	Education function		Value
	Collection function		

According to the confidence value (confidence) value, the correlation analysis between words is mined as follows: {easy to use}→{satisfied} confidence is 0.7895, {convenient}→{easy to use} confidence is 0.6542, indicating that when the words “easy to use” and “convenient” appear, the probability of the word “easy to use” is more than 65%; {good ease of use}→{satisfied} confidence level is 0.6029, indicating that it is expected that the system is convenient to use and conforms to the usage habits of the system. The confidence level of {clear}→{very good} is 0.5321, and the confidence level of {good feedback}→{satisfied} is 0.5352, indicating that users have a high demand for picture clarity during the use of the entire museum system. The confidence level of {interesting}→{good entertainment} is 0.4434, the confidence level of {fun}→{good entertainment} is 0.3848, and the confidence level of {good entertainment}→{satisfied} is 0.4658, which indicates that entertainment is also very concerned. The confidence level of {fluency}→{good} is 0.4126, which indicates that users have certain requirements for the fluency and operability of the system. The confidence level of {wonderful}→{good visuality} is 0.2728, and the confidence level of {good visuality}→{satisfaction} is 0.5875, which indicates that users are very concerned about the visual experience presented by the system. The confidence level of {easy to operation}→{satisfied} is

0.4659, which indicates that users want the system to be simple and not too complicated.

The top 5 items sorted in descending order according to the lexical support value are as follows: {not stuck}→{good operability}, {fluency}→{good}, {clear feedback}→{good feedback}, {interesting}→{good entertainment}, and {wonderful}→{good visuality}. The support degree indicates the probability of word combinations appearing in all comments; at the same time, it also reflects the ranking of users’ attention to a certain extent. Among them, similar words represented by “fluency” and “interesting” have the highest probability of appearing in positive comments.

After tokenizing the negative comment text, the arules package of the R language is used for association analysis. First, the extracted words are replaced with synonyms, words such as “click,” “input,” “check,” and “switch” are replaced with “interaction,” “not very good” and “too bad” with “dissatisfied,” and the “fuzzy” and “not clear” with “unclear”, the related words indicating inconvenient operation are replaced with “poor operability,” the comment words are converted into a transaction type data set, and Apriori correlation analysis is performed. The data processing results and visualization are listed in Table 10.

It can be seen from Table 10 that in the negative comments, the words that have an effective correlation with

TABLE 5: Example of sentiment classification rule words.

Sentiment classification	Typical vocabulary examples
Satisfied	Very good, very satisfied, very satisfied, not bad, ok, very beautiful, very realistic, and magnificent
Neutral	a little, barely, not very, average, and mediocre
Dissatisfied	Not smooth, disappointed, dissatisfied, disliked, not flashy enough, inconvenient, lacking, inconsistent, and uninteresting

TABLE 6: Examples of manual marking.

Sentiment classification	Label	Text example
Satisfied	1	1. Wow, I really like it, it feels like being there. 2. The online high-definition big picture exhibition is great, and you can even see the clothes of the characters.
Neutral	0	1. Okay, at least you can visit the museums without going out. 2. Such a museum is generally interactive.
Dissatisfied	-1	1. Lack of interesting and vivid explanations, it is boring. 2. On the whole, I am a little disappointed, the 3D effect is not dazzling enough.

TABLE 7: Sentiment classification results of museum user experience evaluation.

Data set	Accuracy	Precision	Recall	F1
Training set	0.8973	0.8645	0.8965	0.8802
Test set	0.8606	0.8176	0.8256	0.8216

TABLE 8: Examples of interactive comment feature words.

Index	Feature word example
Ease of use	Convenient, troublesome, simple, difficult to use, and easy to use
Clear feedback	Clear, clear, blurry, too dark, and too bright
Interactive entertainment	Funny, funny, interesting, cute, boring, boring, and so funny
Freedom of operation	Smooth, not stuck at all, not stuck, very smooth, a little stuck, and not moving
Interactive visual consistency	Rich, wonderful, classic, novel, and good looking
Simple and clear operation	Simple, fast, and clear at a glance

TABLE 9: Analysis results of the correlation between positive reviews and satisfaction.

Rules	Lift	Support	Confidence
{Easy to use}→{satisfied}	6.5323	0.0112	0.7895
{convenient}→{easy to use}	5.2546	0.0108	0.6542
{Good ease of use}→{satisfied}	7.6440	0.0132	0.6029
{Clear}→{very good}	3.3865	0.0228	0.5321
{Clear feedback}→{good feedback}	4.5123	0.0420	0.4652
{Good feedback}→{satisfied}	1.9758	0.0136	0.5352
{interesting}→{good entertainment}	3.7549	0.0349	0.4434
{fun}→{good entertainment}	3.1718	0.0256	0.3848
{Good entertainment}→{satisfied}	2.2938	0.0243	0.4658
{fluency}→{good}	1.6284	0.0466	0.4126
{Not stuck}→{good operability}	2.0236	0.0501	0.3102
{wonderful}→{good visuality}	5.7431	0.0332	0.2728
{novel}→{satisfaction}	4.8313	0.0284	0.3006
{Good visuality}→{satisfaction}	5.7652	0.0119	0.5875
{easy}→{good}	1.3526	0.0132	0.7643
{Clear at a glance}→{easy to operate}	3.5435	0.0157	0.5456
{Easy to operate}→{satisfied}	6.0348	0.0265	0.4659

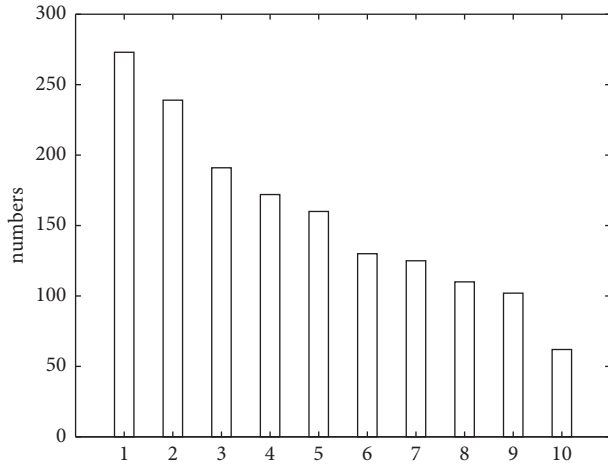


FIGURE 5: Top 10 word frequency statistics of positive comment text.

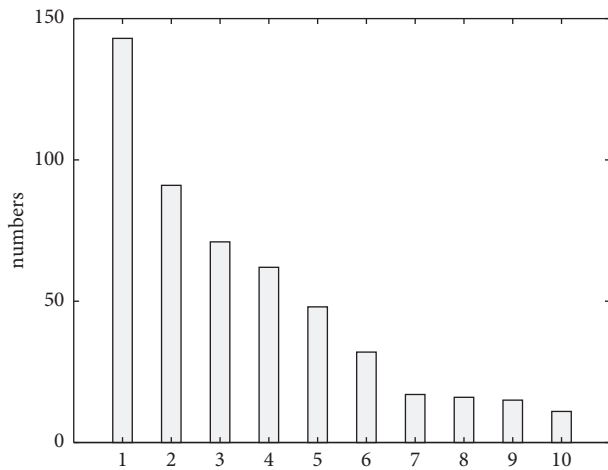


FIGURE 6: Top 10 word frequency statistics of negative comment text.

the satisfaction words are {not easy to use}→{too bad}, {poor picture}→{uninteresting}, and {monotonous}→{dislike}. It can be seen that the main factors causing the negative emotions of consumers are the fluency of the system and the unsatisfactory picture sense. The vocabulary combinations that are effectively related to words with poor operability are {stuck}→{dissatisfied}, {slow feedback}→{poor interaction}, and {cumbersome}→{poor operability}, and this result is similar to the correlation analysis result of positive comment text, which further indicates that the system interactivity elements are highly correlated with user satisfaction.

According to the confidence value, the correlation analysis results of words can be obtained as follows: (1) The confidence level of {not easy to use}→{too bad} is 0.5228; (2) The confidence level of {inconvenient}→{dissatisfied} is 0.4225; (3) The confidence level of {cumbersome}→{poor operability} is 0.4048. When the words “not easy to use” and “inconvenient” appear in the comments, there will be a 40% to 55% probability that words similar to “poor operability”

TABLE 10: Analysis results of the correlation between negative reviews and satisfaction.

Rules	Lift	Support	Confidence
{Not easy to use}→{too bad}	3.7821	0.0869	0.5228
{inconvenient}→{dissatisfied}	5.4655	0.0466	0.4225
{cumbersome}→{Poor operability}	2.8632	0.0298	0.4048
{Poor picture}→{uninteresting}	1.9874	0.0373	0.4866
{monotonous}→{dislike}	4.3245	0.0198	0.3738
{stuck}→{dissatisfied}	2.4315	0.0731	0.3994
{Will not use again}→{dissatisfied}	5.2832	0.0698	0.4687
{Slow feedback}→{poor interaction}	2.0437	0.0385	0.5961
{disappointed}→{dissatisfied}	4.7633	0.0279	0.7325
{boring}→{dissatisfied}	3.7656	0.0817	0.7193
{missing}→{dislike}	1.7855	0.0693	0.5532

will be mentioned. The confidence levels of {poor picture}→{uninteresting} and {monotonous}→{dislike} are 0.4866 and 0.3738, respectively, which indicates that users believe that the monotony of the picture will affect the user experience and thus reduce their satisfaction. The confidence levels of {stuck}→{dissatisfied} and {slow feedback}→{poor interaction} are 0.3994 and 0.5961, respectively, which indicates that users believe that interactivity is a major factor affecting their satisfaction.

According to the descending order of the vocabulary support value, the top 3 are {not easy to use}→{too bad}, {boring}→{dissatisfied}, and {stuck}→{dissatisfied}, and the corresponding support values are 0.0869, 0.0817, and 0.0731, respectively. Support indicates the proportion of word combinations appearing in all comments and also reflects the ranking of user attention to a certain extent. Among them, similar words represented by “not easy to use” and “stuck” have the highest probability of co-occurring in negative comments.

4. Conclusion

This study takes the user experience review texts of digital museums as the research object and proposes a USVM model that can deal with imbalanced data to classify related text sentiments. Based on the classification results, the top 10 word frequencies of different categories are extracted. The Apriori algorithm based on association rules is used to analyze the correlation between the different emotions of users and the interactive experience of digital museums. By analyzing the experimental results, we can find some problems existing in the construction and use of the existing digital museum and put forward corresponding improvement strategies accordingly. Conclusions are drawn from the above research: (1) the ease of use of digital museums is positively related to user satisfaction, that is, the simple and clear operation will help users to feel satisfied, while the cumbersome operation will easily lead to negative comments from users; (2) the interactive performance factors that museum users pay attention to when using the system are clear feedback, interactive entertainment, freedom of

operation, consistent interactive vision, and easy and clear operation; (3) users' dissatisfaction with museum interactivity mainly focuses on unclear feedback, lack of interactivity, cumbersome operation, and inconsistent interaction. In this study, three-level classification is adopted for the processing of comment text sentiment, and it can be upgraded to five-level sentiment classification in the future, so as to dig deeper into the internal perception elements of users. At the same time, future research can enrich the data collection channels and indicators of concern and carry out comparative research on the mining of comment texts on different platforms or indicators.

Data Availability

The datasets used in this paper is available at <https://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Jiangsu Social Science Foundation (study on cultural and creative transformation of Jiangsu museum cultural resource based on data driven) under grant no. 21YSD018.

References

- [1] V. B. Naya and L. A. H. Ibanez, "Evaluating user experience in joint activities between schools and museums in virtual worlds," *Universal Access in the Information Society*, vol. 14, no. 3, pp. 1–10, 2014.
- [2] D. I. Ko, Y. Sumi, and Y. Choi, "Personalized virtual exhibition tour (PVET): an experiment for Internet collaboration," in *Proceedings of the IEEE International Conference on Systems*, Tokyo, Japan, October 1999.
- [3] A. Pagano, G. Arnone, and E. D. Sanctis, "Virtual museums and audience studies: the case of "Keys to Rome exhibition," in *Proceedings of the Digital Heritage International Congress*, Granada, SPAIN, November 2016.
- [4] M. Moussaïd, V. R. Schinazi, M. Kapadia, and T. Thrash, "Virtual sensing and virtual reality: how new technologies can boost research on crowd dynamics," *Frontiers in Robotics and Ai*, vol. 5, p. 82, 2018.
- [5] S. Yu Song and S. Shouqian Sun, "Research on virtual exhibition interaction and user experience technology," in *Proceedings of the 2008 9th International Conference on Computer-aided Industrial Design & Conceptual Design*, Beijing, China, November 2008.
- [6] W. Ma, Y. Wang, Y. Q. Xu, Q. Li, X. Ma, and W. Gao, "Annotating traditional Chinese paintings for immersive virtual exhibition," *Journal on Computing and Cultural Heritage*, vol. 5, no. 2, pp. 1–12, 2012.
- [7] M. Lindvall, D. Ganesan, and R. Ardal, "Metamorphic model based testing applied on NASA DAT -an experience report," *International Conference on Software Engineering*, vol. 129, 2015.
- [8] S. Rubio, E. Diaz, J. Martin, and J. M. Puente, "Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods," *Applied Psychology*, vol. 53, no. 1, pp. 61–86, 2004.
- [9] Q. Li, D. Xiong, and M. Shang, "Adjusted stochastic gradient descent for latent factor analysis," *Information Sciences*, vol. 588, pp. 196–213, 2022.
- [10] R. Pervez, "Shakespeare and principal components analysis," *Digital Scholarship in the Humanities*, vol. 36, no. 4, pp. 1030–1041, 2021.
- [11] S. Keishi and H. Hideitsu, "Modal principal component analysis," *Neural Computation*, vol. 32, no. 10, pp. 1901–1935, 2020.
- [12] A. Bafahm and M. Sun, "Some conflicting results in the analytic Hierarchy process," *International Journal of Information Technology and Decision Making*, vol. 18, no. 2, pp. 465–486, 2019.
- [13] E. L. C. Law and P. V. Schaik, "Modelling user experience - an agenda for research and practice," *Interacting with Computers*, vol. 22, no. 5, pp. 313–322, 2010.
- [14] H. L. O'Brien and E. G. Toms, "The development and evaluation of a survey to measure user engagement," *Journal of the Association for Information Science & Technology*, vol. 61, no. 1, pp. 50–69, 2014.
- [15] J. Park, S. H. Han, H. K. Kim, S. Oh, and H. Moon, "Modeling user experience: a case study on a mobile device," *International Journal of Industrial Ergonomics*, vol. 43, no. 2, pp. 187–196, 2013.
- [16] X. Chen, R. Huang, X. Li, L. Xiao, M. Zhou, and L. Zhang, "A novel user emotional interaction design model using long and short-term memory networks and deep learning," *Frontiers in Psychology*, vol. 16, no. 12, pp. 1664–1078, 2021.
- [17] X. Chen, M. Cao, H. Wei, Z. Shang, and L. Zhang, "Patient emotion recognition in human computer interaction system based on machine learning method and interactive design theory," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 2, pp. 307–312, 2021.
- [18] X. Chen, "Lijun Xu, etc. Emotion interaction recognition based on deep adversarial network in interactive design for intelligent robot," *IEEE Access*, vol. 7, pp. P166860–P166868, 2019.
- [19] K. Priyadarshini Manisha and B. Sathya Bama, "Hyper-spectral image classification with support vector machine," *Lecture Notes in Electrical Engineering*, vol. 700, pp. 587–593, 2021.
- [20] J. Li, S. Lin, K. Yu, and G. Guo, "Quantum K-nearest neighbor classification algorithm based on Hamming distance," *Quantum Information Processing*, vol. 21, no. 1, pp. 1–17, 2022.
- [21] F. Abramovich, V. Grinshtein, and T. Levy, "Multiclass classification by sparse multinomial logistic regression," *IEEE Transactions on Information Theory*, vol. 67, no. 7, pp. 4637–4646, 2021.
- [22] G. R. Kumar, K. V. Sheshanna, S. R. Basha, and P. K. K. Reddy, "An improved decision tree classification approach for expectation of cardiocogram," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 62, pp. 327–333, 2021.
- [23] A. Sharma and K. Tripathi, "Hybrid version of apriori using mapreduce," *Mobile Radio Communications and 5G Networks*, vol. 140, pp. 585–592, 2021.
- [24] K. S. Pradeep, O. Esam, A. Rafeeq, M. Awais, D. Habib, and C. Prasenjit, "Optimized recommendations by user profiling using apriori algorithm," *Applied Soft Computing*, vol. 106, Article ID 107272, 2021.

Research Article

A Dynamic Model of Evolutionary Knowledge and Capabilities Based on Human-Machine Interaction in Smart Manufactures

Shuxian Chen, Zongqiang Ren, Xikai Yu , and Ao Huang 

School of Business, Wenzhou University, Wenzhou 325035, China

Correspondence should be addressed to Xikai Yu; yuxikai@stu.wzu.edu.cn and Ao Huang; huangao326@163.com

Received 17 March 2022; Revised 30 March 2022; Accepted 5 April 2022; Published 26 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Shuxian Chen 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 increasing use of smart machines and devices is not only changing production principles but also reshaping the value of cocreation logic. The interaction between human and smart machine is the enabler of generating augmented intelligence. A system dynamics model is abstracted from smart manufacturing practices to represent the evolutionary processes of inertia, capability, and reliability induced by human-machine interaction. Human-machine interaction is conceptualized into two dimensions: technical and cognitive interaction. Simulation experiments illustrate how the improvement of human-machine interaction can leverage the dynamic capability and reduce the inertia in enterprises through multiple nonlinear feedbacks. There are two pathways to improve reliability and performance in enterprises by human-machine interaction: (1) to promote initiative innovation (change) from endogenous enabler by improving dynamic capability and (2) to promote transformation of knowledge and variation triggered by exogenous environmental changes to improve the dynamic capability for the flexibility and reliability.

1. Introduction

Modern technologies are redefining and reshaping lifestyles and social and economic practices, especially machines embedded with intelligence can do complex tasks and even generate art and new knowledge, not just simple, routinized tasks. Smart manufacturing is not only a production mode in Industry 4.0 but also a new value creation logic and innovation paradigm for the new industrial revolution [1, 2], for example, it reshaped the command and control paradigm in traditional management theory [3]. Organizational innovation in this digital transformation landscape has become a fast-growing and highly important arena for research by understanding the implications of these trends to inform designing for social and economic change.

Given the soaring numbers of smart machines used in workplaces, they are becoming our new colleagues to co-create value with and for their users [4, 5]. Intelligent organizations in future need to blend technology-enabled insights of smart machines with a sophisticated

understanding of human emotion-cognition. Accordingly, cooperation and interaction of humans with smart machines for augmented intelligence are becoming more important than ever. It is time to consider the strategic role of smart machines, new colleagues of human beings, for organizational change. Numerous literatures have investigated the complex relationships, between human and machine, such as technical frameworks, social norms, and networking principles, even emerging a new discipline of ethorobotics [6]. Nonetheless, most of them focused on socio-technical, psychological, and behavioral problems from micro (individual) and macro (social) levels. These conclusions have been exercised at a meso level (organizational) and are little examined. Theoretically, it remains unclear how human-machine interactions shape or enable the change of an organization, in instances such as structural evolution, capability transformation, or production innovation.

This paper explores the mechanism capability transformation enabled by human-machine interaction in the workplace by system dynamic model, which are

- (1) Identifying and representing dynamic elements related to dynamic capability for the smart manufacturing.
- (2) Investigating the evolution of knowledge, inertia, capability, and variation in the human-machine interaction to promote the performance.

2. The Ories and Models

2.1. Human-Machine Cooperation: Technical and Cognitive Interaction. Human-machine interaction is defined as the interaction and communication between human users and machines through multiple interface channels in a dynamic environment [1]. Scholars summarized the evolution of human-machine interaction based on the developing functions of machines [7, 8]. In early work, machines were operated as tools to improve productivity and quality. With the development of electrification and automation, machines can facilitate taskwork by automating production. After the 1970s, with the introduction of information and communication technology, machines could leverage people's physical and cognitive (e.g., data retrieval and processing) capabilities on the centrality of humans [9]. In the latest industrial revolution (Industry 4.0), the extensive application of digital technologies is promoting the interaction between workers and machines to a new level of interaction of heterogenic agents [10], network communication, cognitive processes, and emotional bounding which are emerging in human-machine interaction systematically.

From the evolution of the relationship between man and machine, we can grasp that the interaction between man and machine evolves along technical and cognitive dimensions. Technical dimension refers to the changes in the way and content of human-machine interaction brought by technology [11]. For example, technologies like the Internet of things (IOT), artificial intelligence (AI), and cloud computing promote multichannel interaction between humans and machines throughout a highly networked environment [2, 12]. Schoemaker and Tetlock [13] argued that the most intelligent organizations will need to blend technology-enabled insights with a sophisticated understanding of human reasoning and creativity in the coming years. Today's robotics become more powerful, more flexible, and smarter. Ultimately, these robots will not only be able to communicate with each other, but it is more safe and smarter to work with and even learn new skills from humans.

The cognitive dimension refers to the changes in the social and psychological relations in human-machine interacting as peers or companions, such as trust, social recognition, and emotional-cognitive creativity [11, 14, 15]. Ethological approach shows that social robotics can provide a more plausible functional human-robot interaction [16]. For example, humans are willing to accept robots as trustable partners if they can ascribe some form of awareness and true understanding to them [16, 17]. Cognitive dimension of human-machine interaction will define the partnerships as the centrality of human-machine in the future, which means

that the coexistence of humans and machines can make both sides smarter over time [6].

2.2. Human-Machine Interaction: Enabler for Development. Organizations are increasingly engaging in digital competition that is enabled or induced by information technology [18]. To create a more intelligent enterprise, executives need to leverage the strengths of both humans and computers in order to produce superior judgments [13]. Intelligent machines as coworkers, rather than technological equipment, may result in a profound transformation to organization, from the strategy structure to the behaviors and feelings of the workers [17, 19]. Furthermore, the cooperation between smart machines and humans generates augmented intelligence, allowing them to become more attuned to the competitive dynamics to integrate, establish interaction with internal and external resources, and deal with environmental changes [13, 20]. These distinctive processes including ways of coordinating and combining are the base of capability generation in smart manufacturing.

Firstly, inertia is institutionalized into a regulatory structure to shape (be shaped by) people's behavior in an organization [21]. Intelligent machine changes the formation and existence of inertia plus routine of organization around which organizations are constructed also through which they operate. For example, ERP can replace the traditional manual procurement and logistics management processes with procedures. Knowledge is the basic material to produce capabilities; at the same time, some capabilities will also be solidified as inertia in the organization by means of stocked knowledge. However, in today's turbulent market environments, firms are increasingly facing challenges to keep their knowledge base up-to-date and upgrade their dynamic capabilities [22].

Secondly, human-machine interaction not only improves the ability of organization but also expands the sourcing and way for ability and knowledge generation [23]. Organizational capability is no longer depended on what humans do well or poorly [5]. Organizational capabilities are defined as a body of knowledge about how to do things [24]. On the other hand, the smart machine can also produce knowledge by itself and complement human capabilities to form augmented intelligence. Additionally, smart machines just like people become the dualistic carriers of capabilities and strategic advantages will increasingly depend on a shared capacity of human and machine [13, 25].

Thirdly, according to evolutionary theories, the innovative agents are evolving into the dualistic actors of human and machine which would trigger more variation. For example, networking machines can expand the scope of knowledge search and speed up information processing. Successful variations are retained and turned into knowledge or capabilities that can be invoked when needed to improve the performance. Performance is measured in the desired levels of reliability to respond flexibly to changes.

In summary, we integrate the key dynamic elements related to dynamic capability based on the human-machine interaction to investigate the inertia accumulation, capability

formation, and the reliability growth in the environment turbulence (Figure 1). This framework includes internal and external changes across the boundary of organization: external changes mainly reflect the triggers of environmental changes on organizational adaptability, internal changes demonstrate the interdependence and feedbacks of variables within organizations to generate capabilities for adaptability.

3. Model Structure and Methods

Drawing on organizational inertia and capabilities dynamic accumulation model proposed by Larsen and Lomi [24], we construct a dynamic model to capture the complex relationship among the capability, inertia, knowledge, and variation based on human-machine interaction. System dynamic modelling has specific advantages to investigate complex dynamic relationships between variables in multiple feedbacks by the observation-generating mechanism [24, 25, 26]. For example, the real-time data tracking and sensemaking capabilities of intelligent technology provide context-aware and curated information to users, enabling them to generate dictate precisely decisions and control the directions of variations. Our strategy is to keep the notation as much intuitive as possible with the practice of smart manufacturing in enterprise. Accordingly, we hypothesize the microstructural relations of components with each model equation from empirical studies and field practice.

3.1. Organizational Inertia. Enterprises' inertia generates from organizational routines and experience as well as culture in business contexts, which is social construal process based on knowledge production, accumulation, and usage. Knowledge can be measured by a stock-and-flow approach [27]. Structural inertia (I) in an enterprise is defined as a stock (or accumulator) variable that integrates the corresponding net flow between an increase in inertia I (+) and a decrease in inertia I (-).

$$I_t = \int_{t_0}^t [I^{(+)}(s) - I^{(-)}(s)] ds + I(t_0). \quad (1)$$

Inertia accumulates over time and is affected by enterprise size (S), accumulation (inertia), and experience (capability from practice). To represent these concepts, we specify logistic functional relations between increase in inertia, size, and capabilities. Furthermore, we assume that inertia will increase by a small amount of capability every year at the rate of μ ,

$$\frac{d(I^+)}{dt} = s * \text{Ln}(I_t - C_t) * \text{Ln}(C_t) + \mu * \text{Ln}(C_t). \quad (2)$$

We assume that change attempts (CA) are intendedly adaptive and have the basic objective of decreasing structural inertia, which means the innovation will decrease the inertia in the environment turbulence. So, we model the decrease in inertia as a random variable which represents how much inertia decreases when a change attempt is initiated in the environment turbulence (E). E is the stochastic variable determining the actual level of environmental turbulence.

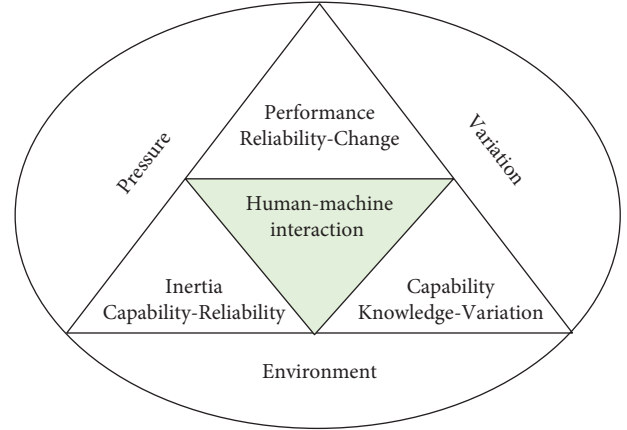


FIGURE 1: Research framework.

$$I_t^{(-)} = \begin{cases} 0, & \text{if } CA_t \leq 0, \\ 0.1 * I_t * E, & \text{if } CA_t > 0. \end{cases} \quad (3)$$

3.2. Change Pressure and Change Attempts. Change attempts (CA) are triggered by the accumulation of the pressure for change (PC) in environmental turbulence. When the pressure for change becomes bigger than the actual threshold (AT), change attempts emerge (i.e., CA takes on a value of 1). The actual threshold for change (AT) is a function of a baseline threshold (BT) and environmental turbulence (E) at the level of inertia in the enterprise. Therefore,

$$CA_t = \text{IF THEN ELSE } (CA_t \geq 0.1, PC_t \times \zeta, 0), \quad (4)$$

$$AT_t = \begin{cases} 0, & \text{if } E < BT, \\ e * \text{Ln}(I_t), & \text{if } E \geq BT, \end{cases}$$

where the BT can be interpreted as the minimum level that pressure for change (PC) must reach in order to trigger change attempts, and e is the average value in the random function of environmental turbulence.

Pressure for change cumulates over time as the actual level of performance diverges from the expected level of performance expressed in terms of reliability. Change attempts then diminish the pressure for change that is represented as

$$PC_t = \int_{t_0}^t [PC^{(+)}(s) - PC^{(-)}(s)] ds + PC(t_0). \quad (5)$$

The pressure for change increases is defined as a function of the gap between expected reliability (ER) and reliability (R) which will accumulate into additional units of pressure for change. However, if actual reliability is better than expected, then no additional pressure for change is recorded. The Max operator is used for this function as follows:

$$PC_t^{(+)} = \text{Max}(0, ER_t - R_t). \quad (6)$$

If change attempts are made, the pressure for change will decrease. The actual effect of this decrease is based on how

successful the change attempt is, so the random component ζ is included in the equation that regulates the decrease in pressure for change.

$$PC_t^{(-)} = \begin{cases} 0, & \text{if } CA_t \leq 0, \\ PC_t * \zeta, & \text{if } CA_t > 0. \end{cases} \quad (7)$$

3.3. Performance and Reliability. Reliability as an indicator of performance is the joint consequence of routinization, formalization, and institutionalization in an organization. So, the increase in reliability is based on the capability, which is enhanced by man-machine interaction (coHM). Increase in reliability in the baseline variability is given as

$$\begin{aligned} R_t^{(+)} &= v_t - R_t \\ &= BV * C_t * coHM - R_t. \end{aligned} \quad (8)$$

The decrease in reliability is defined as a threshold function in the same spirit of above I (-) and PC (-), which means any change attempt would decrease the reliability. The expected reliability (ER) will depend on a trend observed from previous periods and on an explicit managerial goal at the specific level of reliability. TR is the trend in reliability, and SR is the “stretch” in reliability (strategic goal). The trend in reliability is formulated as a first-order exponential smoothing of reliability pushed by the development of human-machine technology. The stretch of reliability as a proxy for the expectations of a steady increase in performance from period to period may be influenced by several managerial factors and defined as a fractional improvement over the level of reliability.

$$ER_t = R_t + TR_t + SR_t,$$

$$TR = \text{Smooth} \left(R_t \cdot \frac{R_t - AR_t}{AR_t}, \tau \right) * Ln(teHM + 1). \quad (9)$$

3.4. Capability, New Knowledge, and Variation. The cognitive interaction of human-machine interaction moderates the increase in capabilities generated by new knowledge and the transformation process. The decrease of capabilities comes from change attempts, which means the existing capability is outdated. This function can be formalized in the following way:

$$C_t = \int_{t_0}^t [LR(s) - EC(s)] ds + C(t_0), \quad (10)$$

where the LR is the learning rate which is affected by the cognitive interaction of human and machines. A high level of cognitive interaction can enhance the quality of transformation from knowledge to capability.

$$LR_t = (1 + coHM)^2 * Ln(NK_t + 1). \quad (11)$$

New knowledge (NK) mainly generates from two sources: variations and cocreation of human-machine. According to evolution theory, some variations will be

selected and converted into new knowledge retained in organizations, which can be leveraged by the synergy of human-machine interaction. At the same time, the synergy of human-machine will create new knowledge. For example, people can get new insights from machine operation by machine-learning.

$$NK_t = \int_{t_0}^t [RR(s) - LR(s)] ds + NR(t_0). \quad (12)$$

The retention rate (RR) is in two parts: the selected fraction of variation that is leveraged by the synergy of human-machine and the new knowledge that is cocreated by human-machine.

$$RR_t = Ln(Vs * V_t + 1) + (1 + coHM)^2, \quad (13)$$

where VS is the selected variation to be retained in the enterprise, which is defined by the technical interaction and cognitive interaction of human-machine. The technology interaction of human-machine can diminish the variability in management, thus decreasing the number of variations to be selected for retention, such as precise decision by big data so that

$$VS_t = \frac{\cdot (coHM + teHM)}{1 + teHM * soHM} \quad (14)$$

Variation (V) originated from the change attempts and innovations in the environment turbulence based on previous reliability. So, the formulation is described as

$$V_t = \int_{t_0}^t [EP(s) - VSO(s)] ds + V(t_0), \quad (15)$$

where EP is the amount of exploration going on in one time period and VSO is the variation selected out, i.e., not adopted by the enterprise, which is a fraction of the total variation minus the selected variation. The exploration is based on the average reliability of enterprises to meet the change attempts of environmental turbulence and is described as

$$EP_t = CA_t * EN_t + Ln(AR_t + 1) * (1 + teHM). \quad (16)$$

The holistic model is presented in Figure 2. The quality and validity of second-order models are tested by internal validity and correspondence between a theoretical narrative and its reconstruction [24].

4. Simulation Result and Analysis

The key enablers in this model are the environments turbulence, technological interaction of human and machines, and cognitive interaction of human and machines. Environment turbulence is defined as a stochastic variable which obeys the random normal distribution (mean = 5, S.D. = e), and e is an indicator to reflect the average degree of environmental change. Technological interaction (teHM) and cognitive interaction (coHM) are evaluated based on human-machine evaluation criteria ranging from 0 to 1. Cognitive interaction includes emotional and cognitive

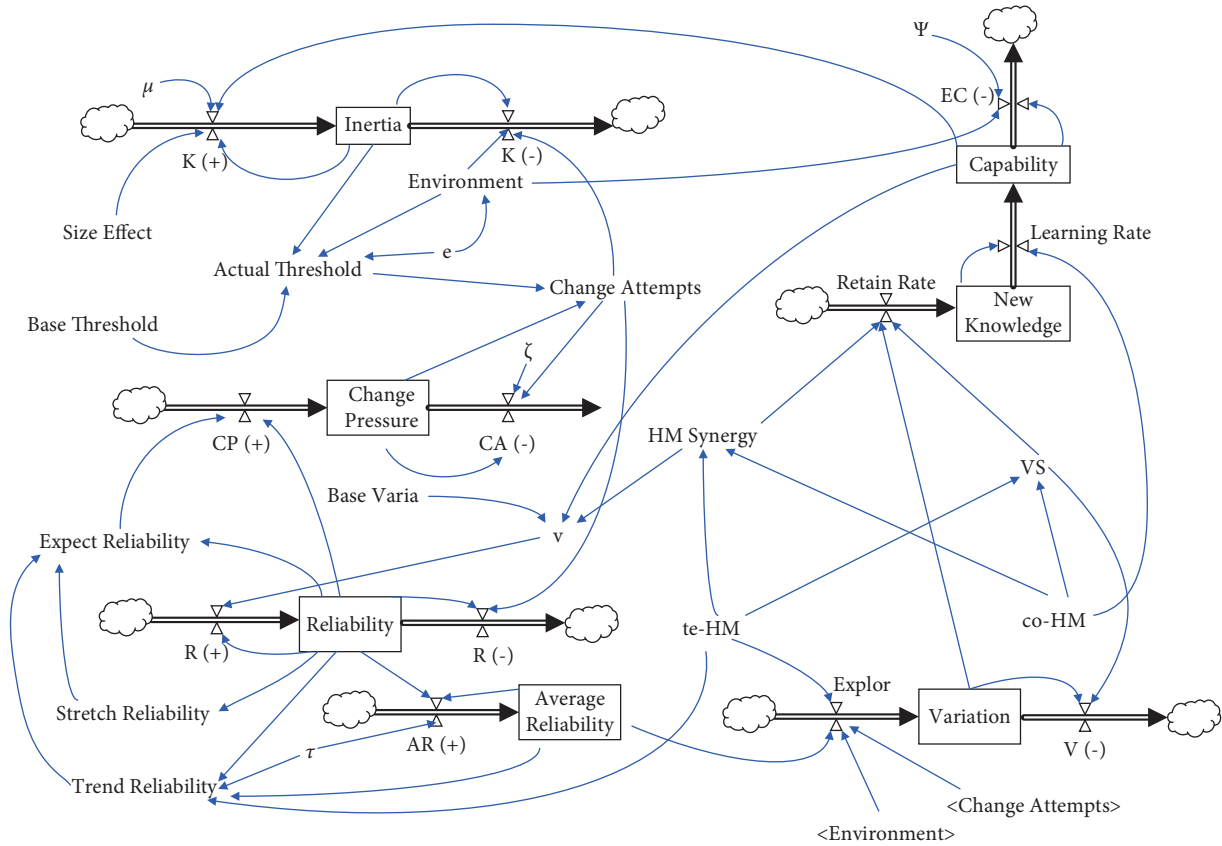


FIGURE 2: Multiple feedback for inertia, capability, variation, and reliability.

aspects, such as friendly interfaces, comfortable feelings, and situated learning. Technological interaction includes function, flexibility, and efficiency, such as visibility of system status, timely response, and smart control.

The simulation results are modeled by the software Vensim•PLE.8.1.0 with numerical integration using the fourth order Runge–Kutta method in a fixed step. This software for SD simulation can be used to define the stochastic components of the models via internal embedded random-number generation in this model, such as environmental turbulence or pressure regulation. Below, we present the simulation results under different scenarios.

4.1. Scenario 1 (Baseline Model). Figure 2 shows the simulation results of the baseline model, which operates at a low level of technological interaction ($teHM = 0.2$) and cognitive interaction ($coHM = 0.2$) of human-machine in the low level environment turbulence ($e = 2$). Figure 3 presents the comparison of operation results of environmental changes (from $e = 2$ to $e = 4$). In Figure 4, we can observe that inertia is building up over time which will accumulate and convert into organizational rigidity. However, the dynamic capability of enterprises only fluctuates at a low level based on the associated new knowledge and variation.

When the environment turbulence increases (e from 2 to 4) resulted in the bigger pressure for change (Figure 4), reliability has not increased but even decreased

accordingly, which means that enterprises who just rely on their inertia without dynamic capability cannot respond to environmental changes timely. In this condition, according to these multiple causal cycles, the dynamic capability cannot provide strong support for enterprises to meet the demand of change attempts in the environment of turbulence. Thus, it is impossible to produce enough variation and new knowledge in the main drivers of capabilities.

4.2. Scenario 2 ($teHM = 0.2, coHM = 0.6$). To probe further the enabler of human-machine interaction for inertia and capabilities, in scenario 2, we improve the level of human-machine cognitive interaction step by step (from 0.2 to 0.6) and set the level of human-machine technological interaction as constant ($teHM = 0.2$). We find that the inertia will increase with the incremental level of cognitive interaction, but as the level of cognitive interaction exceeds the average value (0.5), the accumulation of inertia will decrease after experiencing rapid growth in the initial stage (Figure 5). In this case, the dynamic capability has been growing steadily at a high level. The most fascinating finding is that the change pressure increases dramatically at the constant indicator of the environment ($e = 2$), which is triggered by the endogenous enabler (capability) (Figure 6). Furthermore, the enterprise reliability is still growing greatly, and this growth mainly comes from dynamic capability rather than inertia.

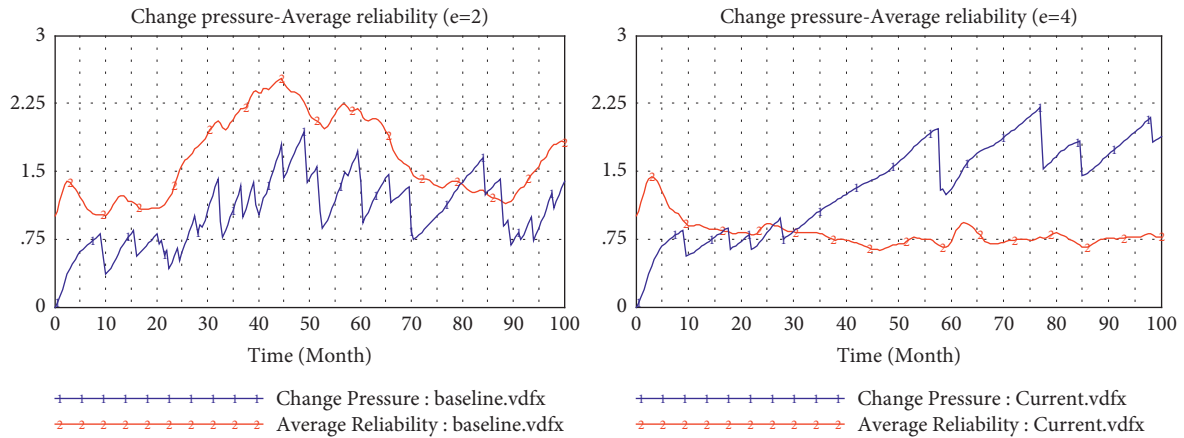


FIGURE 3: Comparison of environmental changes.

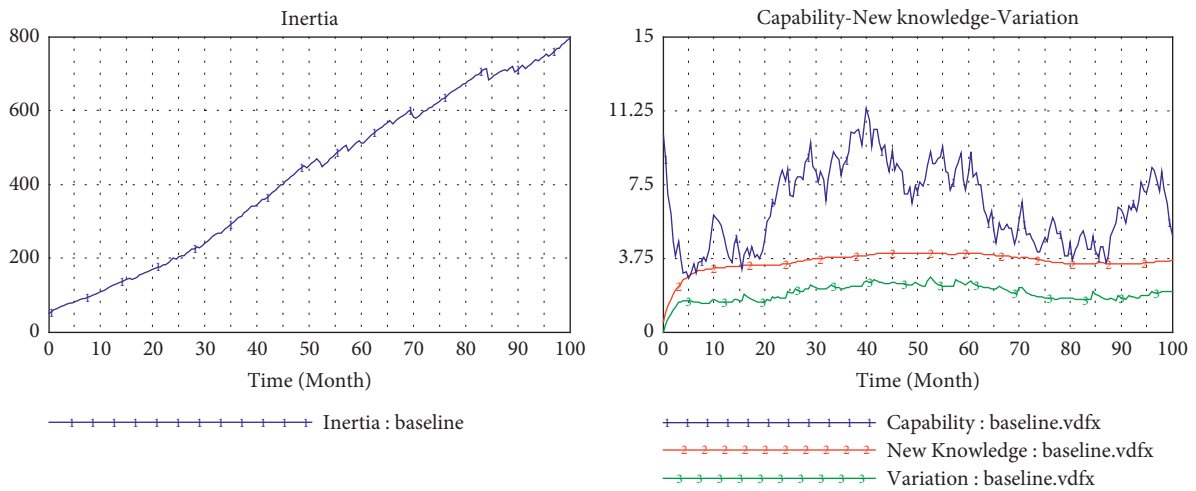


FIGURE 4: Baseline model.

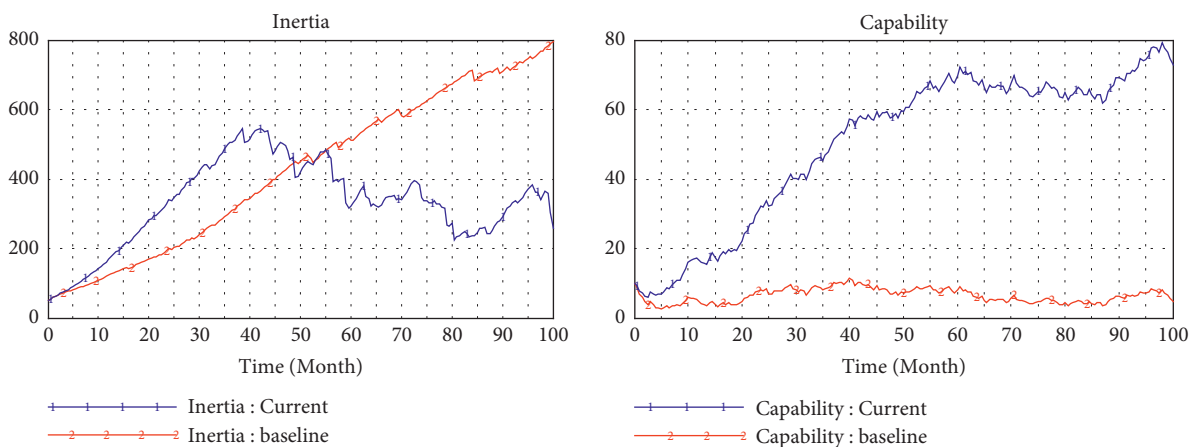


FIGURE 5: Simulation in high level cognitive interaction (0.6).

So, cognitive interaction of humans and machines can improve the dynamic capability to generate growth and reduce organizational rigidity.

4.3. Scenario 3 ($teHM = 0.6, coHM = 0.2$). By the same reasoning above, we improve the level of technical interaction step by step (from 0.2 to 0.6) to examine the effects of

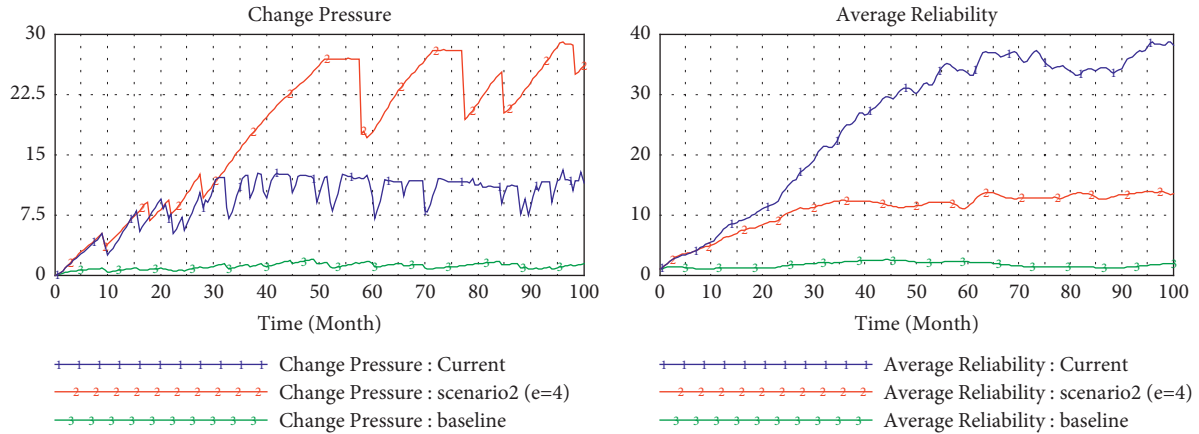


FIGURE 6: Comparison of environmental changes.

human-machine technological interaction. The evolutionary trajectory of inertia is the same as that in scenario 2, but it has more lag in time (Figure 7). Although the amount of variation and new knowledge induced by innovation is greater than that of scenario 2, the dynamic capability is lower than that of scenario 2, which may be due to the low support of human-machine cognitive interaction for knowledge transformation and learning efficiency.

Compared with scenario 2, under the same conditions, the effect of improving technical interaction of human-machine on capability is approximately the same as that of improving cognitive interaction, but it can effectively reduce the organizational inertia than latter.

4.4. Scenario 4 ($teHM = 0.5$, $coHM = 0.5$, $e = 2$). Both cognitive interactions and technical interactions between humans and machines are improved to a medium level of 0.5 in scenario 4. In Figure 8, we observe the level of inertia accumulation is lower than that in scenario 2, and the change pressure is almost the same level with that of scenario 2 and scenario 3 at the same environment turbulence ($e = 2$). However, in Figure 9, the evolutionary trajectory of capability and reliability is almost the same outline with that in scenario 2 and scenario 3, but whose value is significantly greater than the latter two. This is an interesting revelation because it means we can achieve the same result by improving one dimension of human-machine interaction significantly or balancing two dimensions to a medium level.

4.5. Scenario 5 ($teHM = 0.9$, $coHM = 0.9$, $e = 4$). In scenario 5, we simulate the relationship between inertia and dynamic capability when human-machine interaction reaches a perfect level in violent environmental turbulence ($e = 4$). The result shows that the accumulation of inertia remains at a very low level and almost disappears (Figure 10). At this stage, enterprises will become capability-driven organizations where capability is embedded in the human-computer interactive systems of organizations which inspire more attempts for change (Figure 11). Accordingly, enterprises can smartly respond to environmental changes through

ingenious human-machine cooperation without the constraints of inertia and rigidity.

From the simulating results above, we can find the cognitive and technical interaction of human-machine can affect the evolution trajectory of inertia, capability, and knowledge, even the variation in organizations. If organizational inertia are solidified into organizational systems and process specifications, they will be transformed into organizational rigidity, which can not only maintain the stability of the organization but also hinder the adaptability of the organization to the environment. So, dynamic capability is taken as knowledge reconfiguration capability [20]. For example, the cognitive interaction of human-machine can promote the new knowledge convert to the internal capital as form of capability, and the technical interaction of human-machine can promote variation by the dynamically exchanged with external players.

5. Discussion and Implications

Based on smart manufacturing in enterprises, we present an abstract model of operations systems as representing the evolution of inertia, capability, and innovation induced by human-machine interaction in the environment of turbulence and some of the assumptions and propositions in system dynamics terms. The simulations show that the improvement of human-machine interaction can leverage dynamic capability and reduce the inertia in enterprises through multiple nonlinear feedback circles. There are two paths to improving reliability and performance in enterprises: (1) human-machine interaction can promote initiative innovation (change) from endogenous power by improving dynamic capability, and (2) it can improve the transformation efficiency of knowledge and variation induced by exogenous environmental changes to improve the dynamic capability, thus to improve the flexibility and reliability.

Various practical implications are proposed in this research. There are two policies of digital transformation for managers to select to achieve the same effect, which include (1) focusing on one dimension of human-machine interaction (technical or cognitive interaction) significantly; (2)

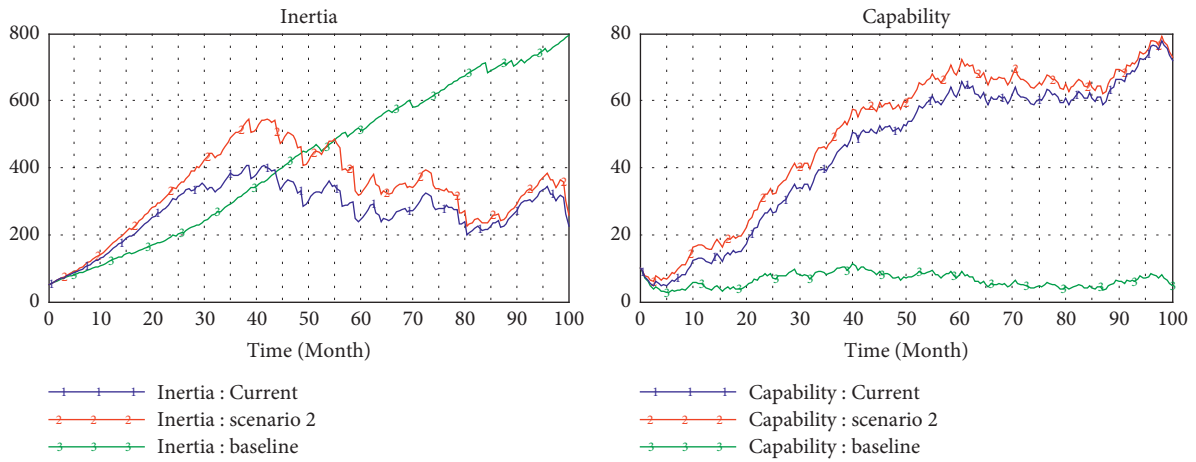


FIGURE 7: Simulation in high level of technical interaction (0.6).

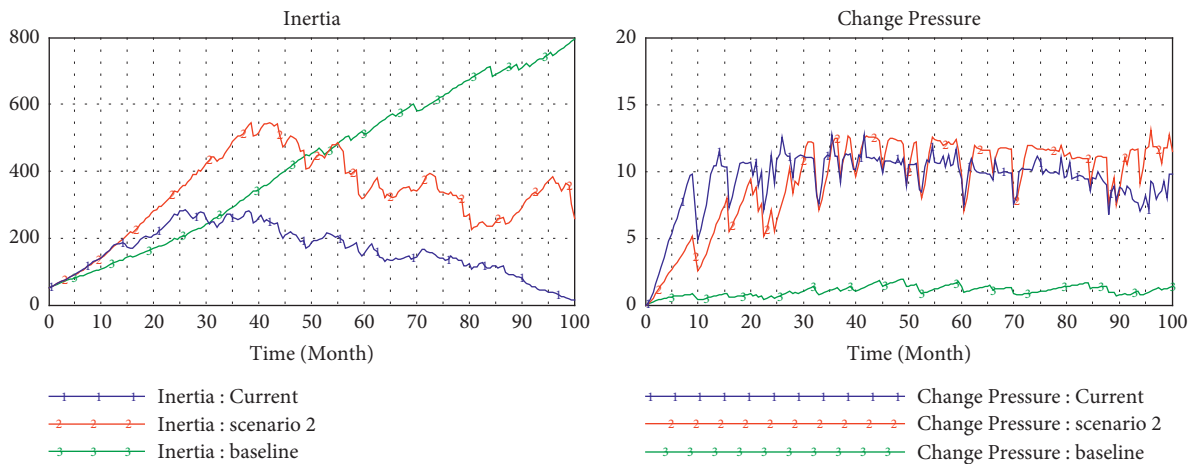


FIGURE 8: Simulation in medium level cognitive and technical interaction (0.5).

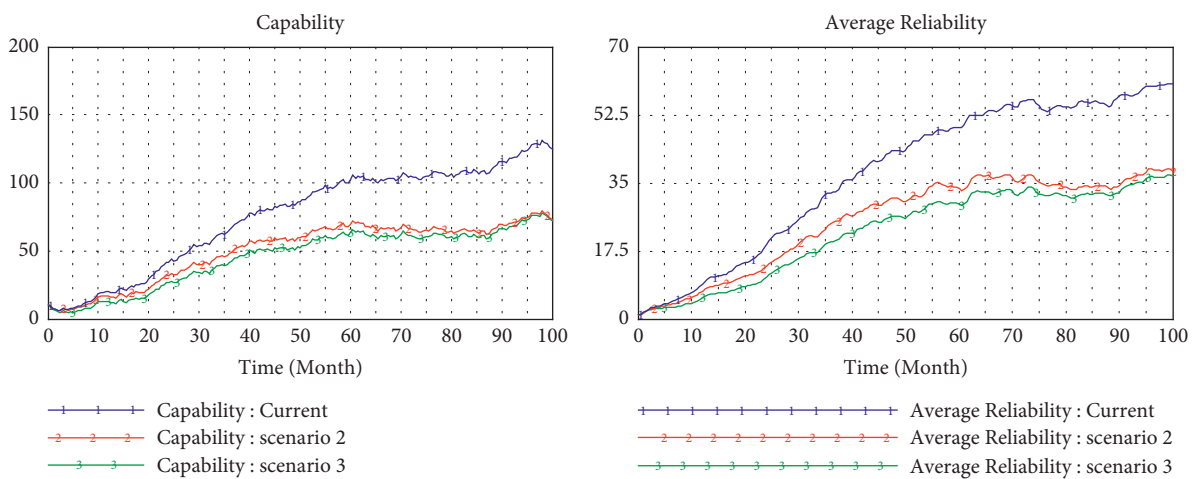


FIGURE 9: Comparison in different scenarios.

promoting both cognitive interaction and technical interaction to a medium level simultaneously; and (3) promoting users' feelings of social presence. This finding suggests that designing smart-manufacture should take into account

anthropomorphic cues such as voices, characters, or feeling and body gestures may foster users' feeling of having an authentic social interaction with a smart object, especially when the technology mostly functions autonomously. In the

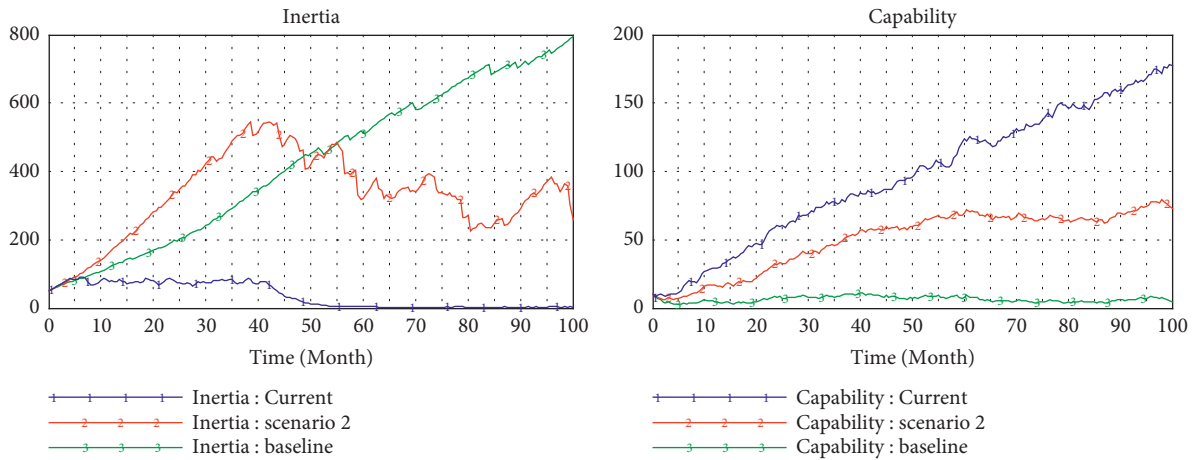


FIGURE 10: Simulation in high level of cognitive and technical interaction (0.9).

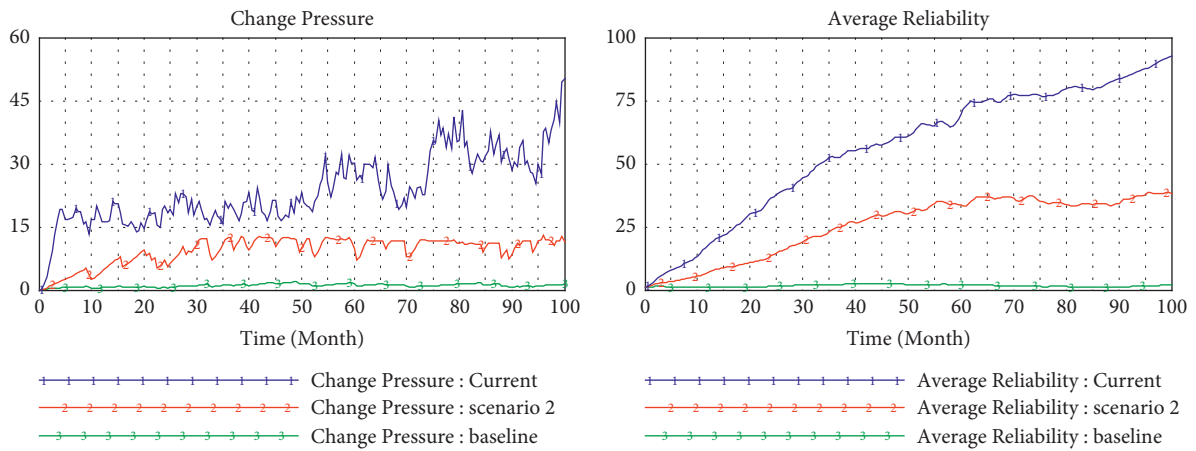


FIGURE 11: Comparison in different scenarios.

direction of creating more intelligent enterprises, managers need to leverage the strengths of both humans and robots to produce superior judgments and productivity, so as to integrate both human creativity and digitally enabled capabilities ingeniously. For example, some small enterprises can prompt the knowledge transformation through the cognitive interaction of human-machine pairs to exploit the potential of existing digital equipment. Big enterprises can improve the technical interaction by digital investment to explore the potential of variation induced by the environmental changes.

Our current modelling efforts suffer from two main sets of limitations. Firstly, just as Larsen and Lomi’s [24] works, we presented a “model of a model” as a “second-order” model which accentuates the demonstration of the theoretical narrative and assumptions, rather than a copy model of a specific enterprise. Secondly, the two dimensions of human-machine interaction are set as exogenous variables in this system. However, they are also the embedded components of this system which are affected by factors such

as inertia accumulation, organizational structure, and investment strategies.

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 authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

Funding from the National Philosophy and Social Science Foundation is gratefully acknowledged. This work was supported by National Office for Philosophy and Social Sciences, China, for the research on the “Mechanism and

Path to Leverage Innovation Capability of Traditional Enterprises for Smart Manufacturing (17BGL044).”

References

- [1] M. Nardo, D. Forino, and T. Murino, “The evolution of man-machine interaction: the role of human in Industry 4.0 paradigm,” *Production & Manufacturing Research*, vol. 8, no. 1, pp. 20–34, 2020.
- [2] A. G. Frank, L. S. Dalenogare, and N. F. Ayala, “Industry 4.0 technologies: implementation patterns in manufacturing companies,” *International Journal of Production Economics*, vol. 210, pp. 15–26, 2019.
- [3] G. Rzevski, “A framework for designing intelligent manufacturing systems,” *Computers in Industry*, vol. 34, no. 2, pp. 211–219, 1997.
- [4] W. Frick, “When your boss wears metal pants,” *Harvard Business Review*, vol. 6, pp. 84–89, 2015.
- [5] J. S. Wesche and A. Sonderegger, “When computers take the lead: the automation of leadership,” *Computers in Human Behavior*, vol. 101, pp. 197–209, 2019.
- [6] Á Miklósi, P. Korondi, V. Matellán, and M. Gácsi, “Ethorobotics: a new approach to human-robot relationship,” *Frontiers in Psychology*, vol. 8, no. 958, pp. 958–8, 2017.
- [7] M. Mori, “The uncanny valley,” *Energy*, vol. 7, no. 4, pp. 33–35, 1970, (in Japanese).
- [8] T. Fong, I. Nourbakhsh, and K. Dautenhahn, “A survey of socially interactive robots,” *Robotics and Autonomous Systems*, vol. 42, no. 3–4, pp. 143–166, 2003.
- [9] C. Breazeal, “Emotion and sociable humanoid robots,” *International Journal of Human Computer Interaction*, vol. 59, pp. 115–119, 2003.
- [10] M. H. Jarrahi, “Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making,” *Business Horizons*, vol. 61, no. 4, pp. 577–586, 2018.
- [11] G. R. Collins, “Improving human-robot interactions in hospitality settings,” *International Hospitality Review*, vol. 34, no. 1, pp. 61–79, 2020.
- [12] P. Kulms and S. Kopp, “A social cognition perspective on human-computer trust: the effect of perceived warmth and competence on trust in decision-making with computers,” *Frontiers in Digital Humanities*, vol. 5, p. 14, 2018.
- [13] P. Schoemaker and P. E. Tetlock, “Building a more intelligent enterprise,” *MIT Sloan Management Review*, vol. 58, no. 3, pp. 28–37, 2017.
- [14] S. Card, T. P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1983.
- [15] M. A. Goodrich and A. C. Schultz, “Human-robot interaction: a survey,” *Foundations and Trends® in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2008.
- [16] P. Morasso, “Gesture formation: a crucial building block for cognitive-based Human-Robot Partnership,” *Cognitive Robotics*, vol. 1, pp. 92–110, 2021.
- [17] S. Brondi, M. Pivetti, S. D. Battista, and M. Sarrica, “What do we expect from robots? Social representations, attitudes and evaluations of robots in daily life,” *Technology in Society*, vol. 66, Article ID 101663, 2021.
- [18] P. A. Pavlou and O. A. El Sawy, “The “third hand”: IT-enabled competitive advantage in turbulence through improvisational capabilities,” *Information Systems Research*, vol. 21, no. 3, pp. 443–471, 2010.
- [19] M.-P. Pacaux-Lemoine, D. Trentesaux, G. R. Rey, and M. Pattrik, “Designing intelligent manufacturing systems through Human-Machine cooperation principles: a human-centered approach,” *Computers & Industrial Engineering*, vol. 111, pp. 581–595, 2017.
- [20] D. J. Teece, G. Pisano, and A. Shuen, “Dynamic capabilities and strategic management,” *Strategic Management Journal*, vol. 18, no. 7, pp. 509–533, 1997.
- [21] B. Levitt and J. G. March, “Organizational learning,” *Annual Review of Sociology*, vol. 14, pp. 319–340, 1988.
- [22] V. Wohlgemuth and M. Wenzel, “Dynamic capabilities and routinization,” *Journal of Business Research*, vol. 69, pp. 1944–1948, 2016.
- [23] F. Gobet and G. Sala, “How artificial intelligence can help us understand human creativity,” *Frontiers in Psychology*, vol. 10, no. 1401, pp. 1–6, 2019.
- [24] E. Larsen and A. Lomi, “Representing change: a system model of organizational inertia and capabilities as dynamic accumulation processes,” *Simulation Modelling Practice and Theory*, vol. 10, pp. 271–296, 2002.
- [25] A. Kusiak, “Smart manufacturing,” *International Journal of Production Research*, vol. 56, no. 1-2, pp. 508–517, 2018.
- [26] J. D. Sterman, *Business Dynamics*, McGraw-Hill, Boston, MA, USA, 2000.
- [27] E. Bolisani and A. Oltramari, “Knowledge as a measurable object in business contexts: a stock-and-flow approach,” *Knowledge Management Research and Practice*, vol. 10, pp. 275–286, 2012.

Research Article

Research on Rice Yield Prediction Model Based on Deep Learning

Xiao Han ¹, Fangbiao Liu,¹ Xiaoliang He,² and Fenglou Ling¹

¹College of Agriculture, Jilin Agricultural University, Changchun 130000, Jilin, China

²Jilin Danong Seed Co, Ltd, Changchun 130000, Jilin, China

Correspondence should be addressed to Xiao Han; hanxiaoyy@jlau.edu.cn

Received 24 February 2022; Revised 18 March 2022; Accepted 24 March 2022; Published 26 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Xiao 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.

Food is the paramount necessity of the people. With the progress of society and the improvement of social welfare system, the living standards of people all over the world are constantly improving. The development of medical industry improves people's health level constantly, and the world population is constantly climbing to a new peak. With the continuous development of deep learning in recent years, its advantages are constantly displayed, especially in the aspect of image recognition and processing, it drives into the distance. Thanks to the superiority of deep learning in image processing, the combination of remote sensing images and deep learning has attracted more attention. To simulate the four key factors of rice yield, this article tries a regression model with a combination of various characteristic independent variables. In this article, the selection of the best linear and nonlinear regression models is discussed, the prediction performance and significance of each regression model are analyzed, and some thoughts are given on estimation of actual rice yield.

1. Introduction

Agriculture is a primary industry, and it is also a security industry [1] for China's economic construction and social development. As one of the most important kinds of grain crops, the research of rice plays an important role in agricultural production and practice [2, 3]. Especially in China, the average planting area, yield per unit area, and total yield of rice rank second in the national grain crop. As the largest rice producer and consumer in the world, it is particularly important to ensure the high yield of rice in China.

Constantly breeding new rice varieties with high yield, good stress resistance, and high nutrient utilization rate, increasing rice yield per unit area, and developing the genetic potential of rice yield as much as possible have become important goals in the field of rice breeding and cultivation in the new era [4]. Studying the characteristics of rice yield is very important for promoting land scale management, ensuring national food security, increasing farmers' income, which is of great significance to effectively alleviate the problem of food shortage [5, 6].

The traditional method of measuring rice yield in the field is destructive, that is, according to the principle of equal area or average sampling in groups, select some small fields, thresh, dry, clean, and weigh the rice after harvest, then measure the water content with moisture meter, and calculate the final rice yield according to the proportion of indica rice and japonica rice of 13.5% and 14.5% [7]. This method is not only cumbersome to operate but also needs to consume a lot of manpower and material resources. Moreover, these steps will lead to larger measurement error. In recent years, the demand for workers in China has been rising, but the cost of employing people has not decreased, but greatly increased. In some regions, the phenomenon of "cannot afford to hire" has even appeared [8].; therefore, there is an urgent need to study a new method of accurate rice yield estimation in the field. At present, governments of various countries are very concerned about food security and food shortage. Accurate estimation of crop yield is an important basis for agricultural departments at all levels to carry out cultivation management and scientific production regulation which is an important reference for countries to formulate corresponding schemes of crop management.

2. Deep Learning Algorithm and Its Application in Crop Yield Prediction

2.1. Research of Deep Learning Based Image Segmentation Algorithm. In recent years, artificial intelligence and image segmentation algorithms are developing rapidly and gradually replace the traditional method with its flexibility of self-adaptive learning from numerous samples [9]. The landmark network structure in the field of deep learning is convolutional neural network, which reduces the number of parameters and improves the generalization ability through local perception and weight-sharing [10]. The main operation of the classical convolutional neural network is to obtain the classification feature vector [11–13] by using the fully connected layer and softmax output after many convolutions. Among them, the full convolution neural network uses deconvolution to restore the size of feature map, which can not only retain the input spatial information but also obtain the output with the same size. This operation can realize the pixel-level segmentation of the image, thereby solving the problem of segmentation [14].

The differences between neural network structures become larger with the increase of network layers. Related researchers have explored different network structures [15–17] and put forward a variety of networks for image segmentation after the appearance of fully convolutional neural networks. It is mainly divided into encoding and decoding structure and expansion of convolution structure; The network representatives of encoding and decoding structure include U-net [18], Seg Net [19], Refinet [20], etc., where an encoder is used to extract image features and dimension reduction, and a decoder is used to recover image dimension and spatial information. The representative networks of expansive convolution are Deep Labv1 [21], V2 [22], V3 [23], V3+ [24], and PSPNet [25] which can increase the size of the input image even if no pooling layer is used so that each convolution can contain more information when outputting. In addition, the networks with good effect in the field of target detection have also been applied to the field of instance segmentation, and achieved good segmentation results, such as regional convolution network (R-CNN) [26], FAST R-CNN [27], Faster R-CNN [28], Maskr-CNN [29], and so on. On the basis of R-CNN, Hybrid Task Cascade (HTC) framework was proposed, which broke through the previous segmentation effect once again. In addition, many researchers have also proposed attention mechanism and applied it to segmentation networks. On this basis, some scholars put forward the DANet, which attached two attention modules to FCN and achieved the latest achievements [30]. In addition to the abovementioned networks, among the image segmentation networks, the networks for feature segmentation are constantly developing which achieved promising results in image classification and target detection. In this process, some classic network structures emerged, such as Lenet in 1998, Alex Net in 2012, Google Net and VGG in 2014, and Res Net in 2015 [31]. With the development of technology, the complexity of the model increases, and the application fields are more ex-

tensive. Aiming at a deep learning algorithm in the field of image segmentation, Minaee and others systematically summarized and introduced all details of it, which is helpful for us to better understand and use it.

2.2. Production Forecast Method. Deep learning is a new technology of image processing and data analysis, which has a good effect and a great potential. With the successful application of deep learning in various fields, the prospect of smart agriculture supported by deep learning is very clear. At present, more than 40 studies in the agricultural field have adopted deep learning technology. These studies show that deep learning provides high precision, which is superior to the existing common image processing technology. In recent years, the exponential growth of remote sensing data has also provided a large number of data sources for geoscience tasks, which give full play to the role of the combination of remote sensing and deep learning in practical applications.

Convolutional neural network (CNN) is one of the most successful deep learning frameworks, which greatly reduces the training parameters and improves the computational efficiency and generalization ability. Recently, many scholars have made a lot of attempts based on CNN structure and applied them to their respective research fields to identify different types of targets in satellite and aerial images through innovative algorithms. Landsat series satellites are widely used data sources, with a spatial resolution of 30 m and a temporal resolution of 16 days. In 2013, China launched the GF-1 satellite, which is equipped with two full-color cameras with a resolution of 2 m and a multispectral camera with a resolution of 16 m. The revisit time of GF-1 satellite is about 4 days. Considering its spatial and temporal resolution, it has obvious advantages. GF-1 is a high-resolution remote sensing image, which contains more spatial information than the medium-resolution remote sensing image. According to this feature, more detailed field crop feature information can be extracted to achieve the purpose of precision in agriculture. Thus far, there is little research on the application of GF-1 satellite images to farmland extraction, especially the advanced deep learning technology.

Prediction model based on image feature is one of the important methods of deep learning in yield prediction of agricultural product. The output can be divided into two types, loss and lossless in vivo prediction. The loss prediction is to measure the length, width, and weight of the ear based on image processing technology after harvesting the mature rice ear, and the prediction model can be established by extracting grain yield characters such as grain length, grain width, total number of grains, aspect ratio, standard deviation, and 1000-grain weight. Some researchers have designed a set of automatic threshing, image acquisition, extraction of grain length, grain width and other data, and automatic bagging. By this device, grain characteristics can be automatically extracted, which is highly correlated with yield and can be used for damage prediction [32]. Lossless yield prediction is mainly based on panicle cutting, color feature extraction, and regression model with yield through

RGB images. A small area of wheat is extracted from the field, RGB images of wheat are taken with the background board, the panicles are segmented by color space conversion and image processing technology, the number of ears is identified, and the number of grains is predicted. Establishing a model to predict the wheat yield per unit area means by employing the image of a single ear and MATLAB image processing, the texture features of some ear images are extracted, and the parameters significantly related to panicle yield are selected, and the prediction model is established by multiple linear regression. UAV takes RGB pictures of rice canopy, uses K-means clustering to segment the pictures, extracts rice ears, and obtains the number of ears which forecast the output.

3. Experimental Analysis of Rice Yield Prediction Based on Deep Learning

3.1. Overall Design. In this experiment, a total of 207 paddy plots in the field were measured. Each plot planted 20 rice plants of the same variety, but the rice varieties in different plots were different. Each rice plot separately extracts the image features of the rice ear plot, the detailed image features of a single rice ear and the seed test features of a single grain. Then, the regression equation between image features and plot total output is constructed, so as to realize the purpose of rice plot yield prediction.

Regression analysis is a method to build a complex regression equation based on the analysis of the correlation between the independent variables of image features and the dependent variables of plot output. According to the difference in the number of independent variables, regression analysis can be divided into univariate regression and multivariate regression. The so-called “univariate correlation regression prediction” is to construct correlation analysis between an independent variable and a dependent variable. Multivariate regression prediction means that multiple independent variables are integrated to predict dependent variables.

Usually, the construction of regression model includes the following steps:

- (1) Make a scatter chart for each independent variable, observe the change trend of independent variable, and analyze whether the dependent variable conforms to the normal distribution, etc., and investigate whether it can be constructed by linear model.
- (2) Selection of characteristic variables and construction of regression model.
- (3) Make a scatter plot between the predicted regression value and the true value, and test the regression equation. The main test contents include goodness of fit test, *F*-test for the significance of regression equation, and *T*-test for the significance of regression parameters.
- (4) Residual analysis and collinearity diagnosis of regression results, including whether the residual distribution of the main accords with the normal

distribution, the judgment of multicollinearity of the regression equation, etc.

- (5) Analysis and diagnosis of regression characteristics in regression model, explanation and discussion of prediction model.

3.2. The Normality Test of the Total Output of the Plot. In the actual research on the estimation of total yield of rice plot, it is necessary to measure the normal distribution characteristics of dependent variables (total yield of plot) at first. The common measurement method is to calculate the skewness according to the histogram of dependent variable to judge the normality of the histogram. Skewness is a measure of skewness of histogram distribution of target data. For a set of data, the histogram distribution is not necessarily symmetrical, and it may be skewed from left to right. Usually, the mode of the histogram is located on the left side of arithmetic mean, which is called left deviation. At this time, the calculated skewness is positive. When the mode of histogram is on the right side of arithmetic mean, the whole distribution is in a state of right deviation, and the skewness is negative. Therefore, when the skewness is closer to 0, it means that the histogram is closer to normal distribution.

The histogram and Q-Q diagram of the total yield (dependent variable) of rice plot are shown in Figure 1.

From the statistical description of the histogram in Figure 1(a), the skewness of the histogram is -0.037 , which is close to 0, indicating that the normality of the dependent variable is better. The Q-Q diagram of the dependent variable (as shown in Figure 1(b)) further verifies the normal property of the distribution. The closer the distribution of observation points in Figure 1(b) is to a straight line, the better the normal property of its distribution.

The normality of dependent variables is tested by using two testing methods in SPSS as shown in Table 1. In this article, the number of effective rice plots is 179 that is not large, so the Kolmogorov–Smirnov normal test results shall prevail. It can be seen from Table 1 that the significant *P* of the dependent variable is 0.200, which is higher than the threshold of 0.05. The above inspection results show that it is significant that the total yield of rice plot obeys the normal distribution, so the linear model can be used to estimate the yield.

3.3. Total Output Estimation of Four Factors Related to Simulated Rice Yield. Four key factors are related to rice yield, including number of ears per unit area, number of grains per ear, seed setting rate, and 1000-grain weight. In this article, based on the simulation of four factors related to rice yield, the characteristics with the ability are selected to represent the yield to establish a regression model and estimate the yield of rice plot.

The corrected area of rice ears in different angles can reflect the number of ears per unit area of rice to a certain extent. However, the area and ear length of detail image about a single rice ear are related to the number of grains per ear. For each plot, because the varieties of rice in the plot are

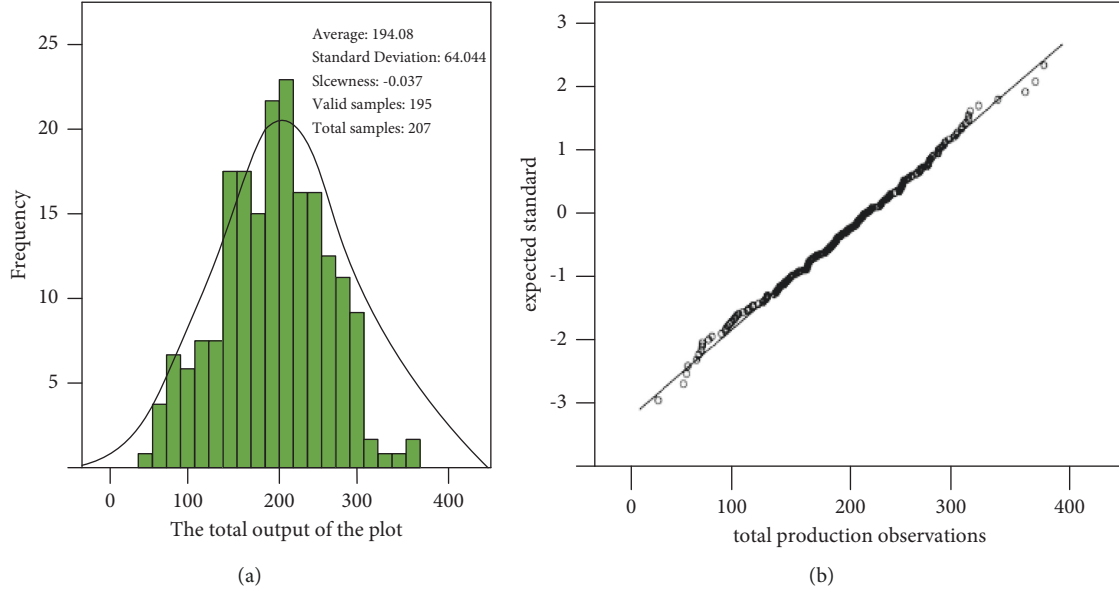


FIGURE 1: Histogram and Q-Q diagram of total production in the field.

TABLE 1: Results of the normality test for rice yield in field plot.

	Kolmogorov–Smirnov			Shapiro–Wilk		
	Statistics	Df	Salience	Statis	Df	Salience
Total plot yield (dependent variable)	0.30	195	0.200b	0.995	195	0.715

consistent, the characteristics of grain seed test of single rice, for example, the seed setting rate and 1000-grain weight can represent the seed setting rate and 1000-grain weight of the whole plot to a certain extent. Therefore, the key point of this section is to build a regression model by the corrected area of the image taken from different angles, the corrected area of detail image about a single rice ear, the ear length, and the seed test parameters of the single rice grain, so as to realize the estimation of the yield. For the convenience of description, we stipulate that the symbol TPCA is used to represent the corrected area of the bottom-view of the rice ear plot, the symbol OPCA is used to represent the corrected area of the top-view of the rice ear plot, the symbol SPCA is used to represent the corrected area of detail image about a single rice ear, the symbol SPL is used to represent the ear length of detail image about a single rice ear, the symbol SRSR is used to represent the grain setting rate of a single rice, and the symbol “SRTW” indicates the 1000-grain weight of rice grains per plant.

The purpose of regression is to obtain the target value of numerical data according to the empirical value. Mathematically speaking, regression is to calculate a regression equation so that the predicted output can be obtained for each input. The goodness-of-fit (R^2) is often used to evaluate the results of regression, and its calculation expression is shown in formula (1):

$$R^2 = 1 - \frac{SSE}{SST} \quad (1)$$

Among them, SSE represents the sum of squares of sample residuals, also known as $2L$ normal form; whereas SST is the sum of the total squares of samples, and the calculation expressions of SSE and SST are shown in formulas (2) and (3):

$$SSE = \sum_{i=1}^N (y_i - y_i^*), \quad (2)$$

$$SST = \sum_{i=1}^N (y_i - \bar{y}). \quad (3)$$

where y_i^* indicates the prediction of the i th sample, y_i represents the true result of the i th sample, and \bar{y} is the average of all true values of the sample. Actually, when the number of independent variables in the model increases, the R^2 of regression fitting will also change, so the number of independent variables should be considered when analyzing the R^2 of regression model. In SPSS, the adjusted goodness-of-fit (adjusted R^2) is generally used to characterize the fitting results after comprehensive investigation of independent variable degrees of freedom. The expression of adjusted R^2 is shown in formula (4):

$$\text{Adjusted } R^2 = 1 - \frac{N-1}{N-M-1} (1 - R^2), \quad (4)$$

where N represents the total number of samples, and M represents the degree of freedom of independent variables (i.e., the number of sample independent variables).

3.3.1. Yield Prediction Model Based on Rice Ear Area.

The area of images taken from different angles can reflect the number of ears per unit area to a certain extent, and there must be a certain correlation between it and rice yield. Therefore, this section mainly studies the regression analysis between TPCA, OPCA, and field plot rice yield (FPRY).

The scatter diagram between the corrected area of plot rice spike image and plot rice yield is shown in Figure 2. In Figure 2(a), the goodness-of-fit R^2 between TPCA and FPRY is 0.2883, whereas the goodness-of-fit R^2 between OPCA and FPRY can reach 0.412, as shown in Figure 2(b). This shows that the corrected area of rice ear image in plot is a significant yield prediction feature.

Figure 2 shows the correlation analysis of univariate linear model and further carries out the prediction of rice yield based on univariate nonlinear model. Among them, the regression goodness-of-fit R^2 of the index model for predicting rice yield, top view and bottom view is 0.2987 and 0.4136. The regression goodness-of-fit R^2 of the numeric model for predicting rice yield from top to bottom is 0.2937 and 0.4389. The regression goodness-of-fit R^2 of the cubic polynomial model for predicting the rice yield under the top and bottom conditions is 0.3015 and 0.4524. The power function model is used to predict the rice yield and the regression goodness-of-fit R^2 of the top and the bottom views is 0.3487 and 0.5133.

Table 2 shows the goodness-of-fit R^2 , adjusted R^2 , F -test, and model significance test of linear and nonlinear univariate regression prediction in the form of data tables. Through the analysis of Table 2, it shows that the optimal univariate regression model of TPCA, OPCA, and FPRY adopts the form of power function. In this form, the goodness-of-fit R^2 of TPCA and FPRY can reach 0.345, while the goodness-of-fit R^2 of OPCA and FPRY is 0.511. The relationship between the corrected rice ear area and the total yield in this plot should conform to the structure of power function.

If TPCA and OPCA are used as the input independent variables of the regression model (Model 1), and the regression equation is constructed by linear model, as shown in Table 3, then the adjusted R^2 of model 1 is only 0.385, which is lower than the regression prediction $R^2 = 0.412$ of single variable OPCA and FPRY. After the introduction of TPCA, FPRY's results decreased, which indicated that OPCA was more reasonable than TPCA in predicting rice plot yield.

3.3.2. Yield Prediction Model Based on the Characteristics of Rice Ear Area and Single Ear Detail Image.

The features extracted from the detailed image of a single rice ear mainly include two features, namely the single spike corrected area (SPCA) and the single spike length (SPL). These two characteristics are related to the number of grains per panicle among the four factors of rice yield to some extent, so SPCA and SPL also have guiding significance for the prediction of FPRY.

Pearson correlation coefficient and significance between SPCA and SPL and FPRY are shown in Table 4. Pearson

correlation coefficient shows that the correlations between SPCA, SPL, and FPRY are 0.470 and 0.376, respectively. The results show that there is a certain correlation between SPCA and SPL, and FPRY. If these two variables are added to Model 1 as input features, Model 2 consisting of four independent variables will be formed. According to linear regression analysis, it can be seen from Table 5 that the goodness-of-fit R^2 of model 2 is 0.413. The goodness-of-fit R^2 of comparison model 1 is 0.385, and the results show that the introduction of detailed features of a single panicle can improve the predictive ability of FPRY.

3.3.3. Yield Prediction Model Based on Rice Ear Area, Detailed Image Features, and Single Seed Test Characters.

Among the seed test characters of a single plant, the most important are the two characteristics of grain setting rate and 1000-grain weight. As the varieties of 20 rice plants in the same plot are the same, the seed setting rate and 1000-grain weight of each plant can reflect the seed setting rate and 1000-grain weight of the whole plot to a certain extent.

Table 6 shows the introduction of two characteristics of single rice setting rate (SRSR) and single rice 1000-grain weight (SRTW) into Model 2, and the fitting of regression model (Model 3) is constructed with six variables. It can be seen from Table 6 that the adjusted goodness-of-fit R^2 of Model 3 is 0.456. The goodness-of-fit r^2 of comparative model 2 is 0.413. The results showed that the introduction of single seed test traits could improve the predictive ability of FPRY.

3.3.4. Screen of Characteristic Traits Based on Stepwise Linear Regression.

There are two methods to screen common characteristic variables, one is to use all subsets regression, the other is to use stepwise linear regression. In this article, stepwise linear regression is used to screen characteristic variables with characterization ability. In the stepwise linear regression, first, the characteristic variables with the highest correlation are screened, and then new variables are introduced one by one. While every time a variable is introduced, F -test and significance T -test of the selected characteristic variables should be carried out. Specifically, assuming that the number of characteristic variables is n , the independent variables of the regression equation can be expressed as $x_1, x_2, x_3, \dots, x_n$, for dependent variable y , the regression expression is shown in formula (5):

$$Y = \alpha + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_i x_i + \dots + \alpha_n x_n, \quad (5)$$

where $i = 1, 2, 3, \dots, n$ α_i is a regression coefficient of independent variable x_i , assuming the maximum value of F -test statistic of α_i is F_{α_i} , then for the significance level of 0.05, suppose that F_{α_i} is greater than or equal to the critical threshold, so x_i , the corresponding characteristic independent variable of α_i can be introduced. Repeat the above process of introducing characteristic variables, but after each introduction of variables, it is necessary to carry out significance T -test on the selected characteristic variables, which is used to ensure that all the selected characteristic

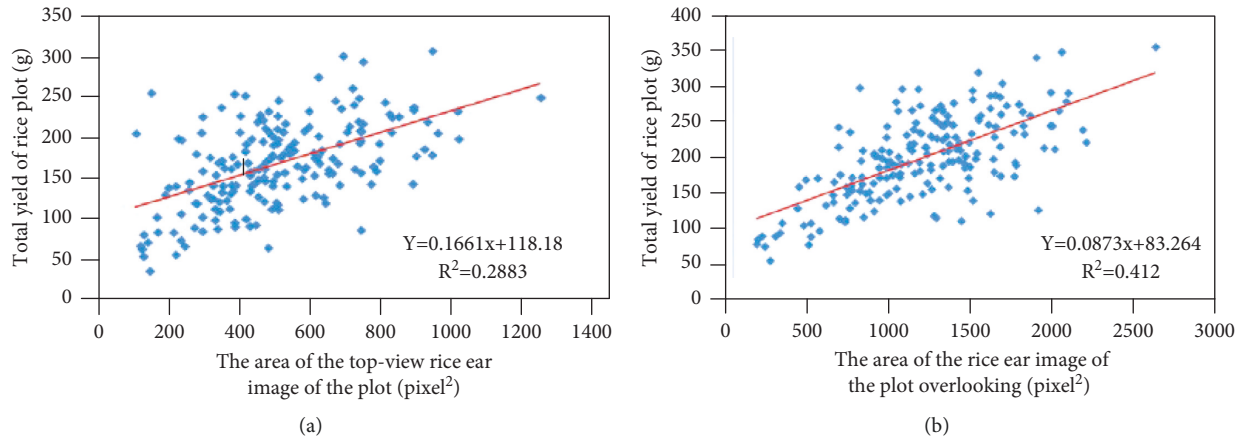


FIGURE 2: Correlation analysis between the TPCA, OPCA, and FPRY.

TABLE 2: Parametric predictive model analysis of the single variables.

	Univariate predictive models	R^2	Adjusted R^2	Standard estimate error	F Test	Salience
TPCA	Linear	0.288	0.285	53.353	76.549	0.000
	Index	0.299	0.295	0.317	50.456	0.000
	Logarithm	0.294	0.290	53.148	78.600	0.000
	Polynomial	0.301	0.290	53.137	26.904	0.000
	Power function	0.349	0.345	0.306	101.171	0.000
OPCA	Linear	0.412	0.409	49.298	134.519	0.000
	Index	0.414	0.411	0.299	135.426	0.000
	Logarithm	0.439	0.436	48.156	150.185	0.000
	Polynomial	0.452	0.444	47.823	52.324	0.000
	Power function	0.513	0.511	0.272	202.493	0.000

TABLE 3: Two-variable regression: Model 1.

	R^2	Adjusted R^2	Significance of the F value	Durbin-Watson
Model 1	0.391	0.385	0.000	1.696

TABLE 4: Correlation between single panicle detail image traits and plot yields.

	Pearson correlation coefficient (R)	Correlation coefficient dominance
Single Inaho image correction surface (SPCA)	0.470	0.000
Single ear length (SPL)	0.376	0.000

TABLE 5: Plot rice yield regression: Model 2.

	R^2	Adjusted R^2	Significance of the F value	Durbin-Watson
Model 2	0.425	0.413	0.000	1.699

TABLE 6: Plot rice yield regression: Model 3.

	R^2	Adjusted R^2	Significance of the F value	Durbin-Watson
Model 3	0.475	0.456	0.000	1.895

TABLE 7: Filtering traits using stepwise linear regression.

	R^2	Adjusted R^2	Significance of the F value	Durbin-Watson
Number of rounds 1a	0.380	0.376	0.000	
Number of rounds 2b	0.425	0.419	0.000	1.790
Number of rounds 3c (Model 4)	0.462	0.453	0.001	

a, predicted value: (constant), OPCA, b, predicted value: (constant), OPCA, SRSR, c, predicted value: (constant), OPCA, SRSR, SPCA.

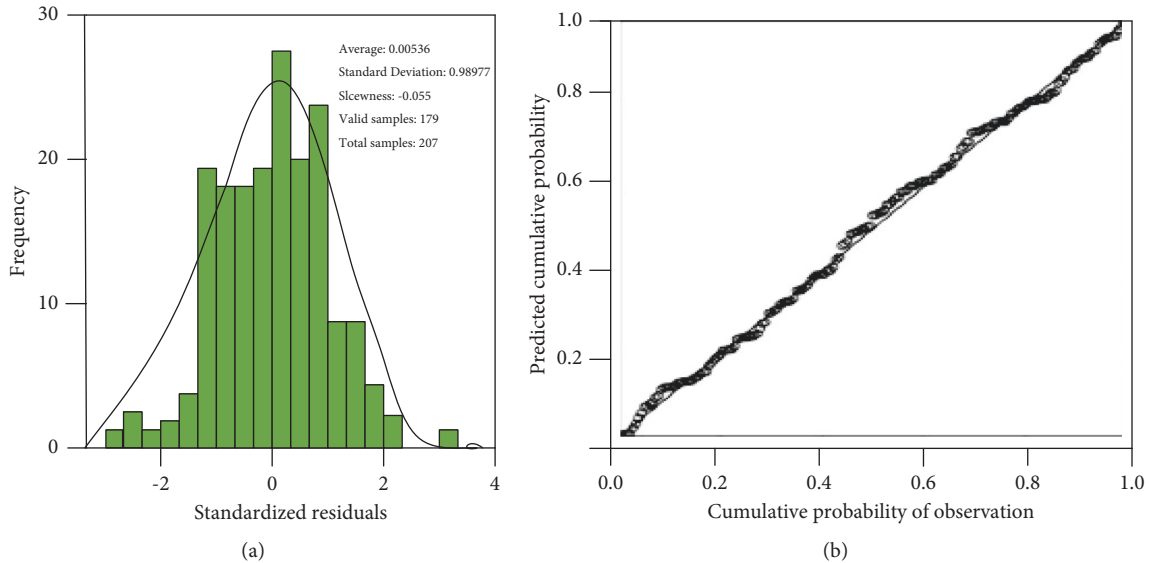


FIGURE 3: Prediction of the residual distribution histogram and P-P graph for Model 4.

variables are significant. If a characteristic variable is introduced, the previously introduced variable is not significant, then the characteristic variable should be eliminated, and finally repeat this process until no new variables can be introduced.

The main advantage of stepwise linear regression is that while constructing regression equation, it can also realize the screening of characteristic variables, which is convenient for people to understand existing models and make corresponding changes. Among them, the first characteristic variable screened out has the highest importance, followed by the second, and so on. After finding the appropriate number of feature variables, the collection of unimportant features can be stopped.

The results of stepwise linear regression after inputting six characteristic parameters (TPCA, OPCA, SPCA, SPL, SRSR, and SRTW) are shown in Table 7. After all six parameters were entered, three variables were screened out, among which OPCA was the first variable screened out, then SRSR, and finally SPCA. That is to say, all these three parameters have passed the T -test of the model, and all of them have strong ability in yield characterization. In Table 7, the adjusted goodness-of-fit R_2 of the yield regression model can reach 0.453, which is called Model 4 in this article. Model 4 consists of three independent variables, namely OPCA, SRSR, and SPCA. Compared with the regression Model 3 consisting of six independent variables, the adjusted R_2 of

Model 4 only dropped from 0.456 to 0.453. The above results show that among the six characteristic variables selected by the four factors of simulated yield, the relationship among OPCA, SRSR, SPCA, and FPRY is closer, so the following regression analysis focuses only on these three selected characteristic variables. The residual n histogram and P-P diagram according to Model 4 are shown in Figure 3.

In Figure 3(a), the histogram skewness of residual distribution in FPRY's prediction is -0.055 , which is close to 0. At the same time, from the P-P diagram of cumulative probability distribution (Figure 3(b)), it can be seen that the observed values are all distributed on the diagonal line, and these results show that it is feasible to use this linear Model 4 to predict FPRY.

4. Conclusion

With the increase of population and the continuous improvement of people's living standards, the demand for food is also increasing. As the main food crop in China, rice has always been the main research object of breeders, and yield prediction has always been an important research orientation of rice. The research structure of this article shows that with a single independent variable, the correlation between OPCA and the total output of rice plot is much higher than that between TPCA and the total output of rice plot. The accurate segmentation of rice ears is of great significance to

the accurate estimation of yield, and the image taken from the perspective of overlooking plays a more obvious role in the estimation of yield.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by <https://doi.org/10.13039/501100012166> National Key Research and Development Program of China, the project number is 2017YFD0100504, and the Key Development Plan of Science and Technology Project of Jilin Province in China, the project number is 20190301061NY.

References

- [1] M. Feng and S. Huang, "Study on prediction model of early rice yield in Nanchang," *Tianjin Agricultural Sciences*, 2017.
- [2] Z. Chu and J. Yu, "An end-to-end model for rice yield prediction using deep learning fusion," *Computers and Electronics in Agriculture*, vol. 174, Article ID 105471, 2020.
- [3] S. J. Varghese, S. Surendran, A. B. Pillai, K. Rajendran, and A. Kitoh, "Future changes in rice yield over Kerala using climate change scenario from high resolution global climate model projection," *Journal of Earth System Science*, vol. 129, no. 1, 2020.
- [4] S. Traore, L. Zhang, A. Guven, and G. Fipps, "Rice yield response forecasting tool (yieldcast) for supporting climate change adaptation decision in sahel," *Agricultural Water Management*, vol. 239, 2020.
- [5] P. Pochanard, "The application of a representative volume element (rve) model for the prediction of rice husk particulate-filled polymer composite properties," *Journal of Materials Science and Applications*, vol. 10, no. 1, p. 26, 2019.
- [6] H. Zhang, G. Zhou, L. Liu, B. Wang, D. Xiao, and L. He, "Climate-associated rice yield change in the northeast China plain: a simulation analysis based on cmip5 multi-model ensemble projection," *The Science of the Total Environment*, vol. 666, pp. 126–138, 2019.
- [7] L. Pan, "Systems model-guided rice yield improvements based on genes controlling source, sink, and flow," *Botanical journal*, vol. 60, no. 12, pp. 1154–1180, 2018.
- [8] T. L. Dammalage, T. Sirisena, and J. Susaki, "MODIS satellite data based rice yield-forecasting model for Sri Lanka: a pilot study on kurunegala district," *Dammalage*, 2018.
- [9] H. Zhong, C. Liu, W. Kong et al., "Effect of multi-allele combination on rice grain size based on prediction of regression equation model," *Molecular Genetics and Genomics: MGG*, vol. 295, no. 2, pp. 465–474, 2020.
- [10] M. Kamruzzaman, S. Hwang, S.-K. Choi et al., "Prediction of the effects of management practices on discharge and mineral nitrogen yield from paddy fields under future climate using apex-paddy model," *Agricultural Water Management*, vol. 241, Article ID 106345, 2020.
- [11] K. Yoshiyuki and M. Yuji, "Impact assessment of climate change on rice yield using a crop growth model and activities toward adaptation: targeting three provinces in Indonesia," *Adaptation to Climate Change in Agriculture*, pp. 67–80, 2019.
- [12] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, and F. B. Fritschi, "Soybean yield prediction from UAV using multimodal data fusion and deep learning," *Remote Sensing of Environment*, vol. 237, 2019.
- [13] S. J. Jang, J. H. Lee, T. W. Kim, J. S. Kim, and J. B. Lee, "A wafer map yield model based on deep learning for wafer productivity enhancement," in *Proceedings of the 2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, IEEE, Saratoga Springs, NY, USA, April 2018.
- [14] S.-J. Jang, J.-S. Kim, T.-W. Kim, H.-J. Lee, and S. Ko, "A wafer map yield prediction based on machine learning for productivity enhancement," *IEEE Transactions on Semiconductor Manufacturing*, vol. 32, no. 4, pp. 400–407, 2019.
- [15] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar, and A. K. Singh, "Crop yield prediction integrating genotype and weather variables using deep learning," *PLoS One*, vol. 16, no. 6, Article ID e0252402, 2020.
- [16] B. Alhnaity, S. Pearson, G. Leontidis, and S. Kollias, "Using deep learning to predict plant growth and yield in greenhouse environments," *Acta Horticulturae*, vol. 1296, pp. 425–432, 2019.
- [17] X. U. Lulu and D. Zhang, "Grain yield prediction model based on deep belief network," *Journal of Henan University of Engineering(Natural Science Edition)*, 2019.
- [18] J. Sun, L. Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep cnn-lstm model," *Sensors*, vol. 19, no. 20, 4363 pages, 2019.
- [19] W. Sheng, S. Sun, L. Zhen, R. Zhang, J. Xu, and S. Avner, "Accurate de novo prediction of protein contact map by ultra-deep learning model," *PLoS Computational Biology*, vol. 13, no. 1, Article ID e1005324, 2017.
- [20] S. Khaki and L. Wang, "Crop yield prediction using deep neural networks," *Frontiers of Plant Science*, vol. 10621 pages, 2019.
- [21] X.-B. Jin, N.-X. Yang, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, "Hybrid deep learning predictor for smart agriculture sensing based on empirical mode decomposition and gated recurrent unit group model," *Sensors*, vol. 20, no. 5, 1334 pages, 2020.
- [22] X. Ni, C. Li, H. Jiang, and F. Takeda, "Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield," *Horticulture Research*, vol. 7, 2020.
- [23] S. Khaki, H. Pham, and L. Wang, "Yieldnet: A convolutional neural network for simultaneous corn and soybean yield prediction based on remote sensing data," *Scientific Reports*, vol. 11, no. 1, 2020.
- [24] S. Wu, J. Yang, G. Cao, Y. Qiu, and J. Dong, "Elevating prediction accuracy for mechanical properties of hot-rolled strips by using semi-supervised regression and deep learning," *IEEE Access*, no. 99, p. 1, 2020.
- [25] J. S. Kim, Y. Cho, and T. H. Lim, "Prediction of the location of the glottis in laryngeal images by using a novel deep-learning algorithm," *IEEE Access*, no. 99, p. 1, 2019.
- [26] Y. Adam, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A deep learning mammography-based model for improved breast cancer risk prediction," *Radiology*, vol. 292, 2019.
- [27] A. Zlokapa and A. Gheorghiu, "A deep learning model for noise prediction on near-term quantum devices," *Computer Science, Physics*, 2020.

- [28] I. Okwuchi, "Machine learning based models for fresh produce yield and price forecasting for strawberry fruit," *Mathematics*, 2020.
- [29] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Ensembles of deep learning models and transfer learning for ear recognition," *Sensors*, vol. 19, no. 19, p. 4139, 2019.
- [30] Y. Yue, J.-H. Li, L.-F. Fan et al., "Prediction of maize growth stages based on deep learning," *Computers and Electronics in Agriculture*, vol. 172, Article ID 105351, 2020.
- [31] B. Saravi, A. P. Nejadhashemi, and B. Tang, "Quantitative model of irrigation effect on maize yield by deep neural network," *Neural Computing & Applications*, vol. 32, no. 14, Article ID 10679, 2020.
- [32] E. Sahin, C. J. Saul, E. Ozsarfatı, and A. Yilmaz, "Abalone life phase classification with deep learning," in *Proceedings of the 2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMi)*, IEEE, Nairobi, Kenya, November 2018.

Research Article

Research on Impulse Power Load Forecasting Based on Improved Recurrent Neural Networks

Chenyang Feng , Kang Xu, and Haoyun Ma

College of Electrical Engineering & New Energy, China Three Gorges University, Yichang, Hubei 443002, China

Correspondence should be addressed to Chenyang Feng; 201908521021015@ctgu.edu.cn

Received 16 March 2022; Revised 28 March 2022; Accepted 7 April 2022; Published 23 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Chenyang Feng 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.

Deep learning is good at extracting the required feature quantity from the massive input information through multiple hidden layers and completing the learning through training to achieve the task of load forecasting. The impulse power load data contain a lot of noise, burrs, and strong randomness. As an improved recurrent neural networks, the output of long short-term memory (LSTM) network is not only related to the current input, but also closely related to the historical information, which can effectively predict the impact power load. An impulse power load forecasting model based on improved recurrent neural networks is proposed. To solve the training difficulties caused by deep networks, database is divided into training data set and test data set. To accelerate running speed and improve accuracy and reliability, parameter setting in deep learning neural network is analyzed. The proposed load forecasting model is verified by simulation and compared with the existing methods. Taking the average relative error as the standard, the effectiveness of the proposed model for the forecasting of impulse power load connected to the bus is verified.

1. Introduction

Grid load forecasting refers to periodic load forecasting for power equipment such as generator sets, power users, and power loads, so as to make reasonable planning to ensure the safe and stable operation of power system. Short-term load forecasting (STLF) is one of the important tasks of energy utilization, power planning, dispatching, and other management services [1–4]. In fact, considering the different operation laws and performance of each power system, on the premise that the transmission accuracy requirements are not discounted, and considering the different influencing factors (including natural environment, human development, social change, etc.), if we can accurately predict the change and fluctuation of power load in the future, it will promote planned energy and power use management.

STLF is beneficial to the preadjustment of power grid operation mode and the arrangement of unit maintenance plan [3, 4]. It can save coal and fuel and is conducive to the compression of power generation cost, which is conducive to

the rational development of energy and the formulation of construction plan.

At present, the research direction of power load forecasting mainly focuses on the utilization of algorithm, the establishment and application of mathematical model, and the determination of main influencing factors [2, 4]. Due to the strong nonlinearity and many random factors of power short-term load, the previously proposed load forecasting theory has shown limitations and deficiencies. The development of new technologies, new ideas, and new methods is the driving force for the long-term and sustainable development of power load forecasting.

With the implementation of energy conservation and emission reduction policies, new energy power plants and ultrahigh voltage (UHV) power grid system are developing rapidly, with a wide variety of power sources, so the load change is more difficult to predict. At the same time, alternating current (AC) and direct current (DC) in UHV are very complex, and the operation and maintenance of power grid is more difficult and faces great challenges. In new

energy power stations, wind power plants and photovoltaic power generation have sprung up under the incentive of policies. However, due to the uncertainty of wind power and solar power generation, the power grid will be violently turbulent. Therefore, how to realize the impact load forecasting of multiple power sources is an urgent problem to be solved.

2. Characteristic Analysis of Impulse Power Load Forecasting

In some areas, due to the existence of large iron and steel enterprises, the impact load accounts for the majority of the overall load, which seriously reduces the accuracy of local load forecasting results. On the premise that some breakthroughs have been made in each new load model, at present, the development of power market is gradually restricted by the accuracy of power system load forecasting and the forecasting state gradually tends to saturation [5]. The application of artificial intelligence in STLF can effectively improve the forecasting accuracy under complex environmental factors. Adding deep learning to power system, STLF can form a higher level of deep intelligence, which can not only manage the security of transmission system, but also play a key role in cost budgeting and power distribution.

Internal and external characteristics are collectively referred to as power load characteristics. The law that the load power changes with the frequency of the load system or the voltage on the load terminal is called the internal power load characteristic, which can be divided into frequency characteristic and voltage characteristic. The inherent load characteristics are mostly used for the analysis of system stability. Usually, the load curve is often used to describe the power load changing with time and reflect the change law of load data in a period of time [6]. However, the daily load curve of power grid in iron and steel impact area is often irregular and even cannot find any regularity all day.

There are various types of steel load, usually including stainless steel, milk line, electric arc furnace, oxygen production, section steel, etc.; and it has heavy single line load, ranging from 30 MW to 80 MW. In the power system load in a region, the impact load capacity of iron and steel enterprises is huge, with high randomness and poor regularity. Some enterprises need to arrange the production plan according to the preferential section of peak and valley prices. Sometimes the production plan will be adjusted due to events, and the accuracy of load forecasting will be greatly reduced.

3. LSTM Model

Long short-term memory (LSTM) [7–9] is a variant form developed and improved by recurrent neural networks (RNN). The structure of “gate” of LSTM neural network is used to control the state of nucleus. A selective gate function information through which cell core states can be added or removed. The ancestor macros form a set of cyclic subarrays called memory blocks. The memory or forgetting of key node

information is selected to realize the combination of long-term memory and short-term memory. This improves the spatiotemporal task which is difficult to be controlled by RNN artificially. The “gate” of LSTM neural network is utilized by a sigmoid neural network layer through point-to-point multiplication operation.

An output value of “0” for the sigmoid layer means no information can be passed, but a value of “1” means all information can be passed. LSTM neural network has three gate structures, including “input gate,” “output gate,” and “forgetting gate.” They are used to control and protect the state of the core of the unit. Figure 1 shows the LSTM structure.

The detailed derivation of the arithmetical expression of LSTM neurons follows. Let us set the time variable to t . The input variable of LSTM (long short-term memory recurrent neural network) set as the sequence input of time t is x_t , the LSTM output of time $t-1$ is h_{t-1} , and the gating state quantity of time $t-1$ is s_{t-1} . LSTM output t is the output value h of LSTM at time t . State s_1 of the door control unit at time t . In LSTM, the oblivion gate works by turning the last moment’s HT. The x_t at this time is regarded as the input to the sigmoid layer. It outputs values between 0 and 1 and sends them to s_{t-1} . Determining the influence of x_t on its s_t is within the function range of the input gate. It is the function of the output gate to control the influence of s_t on h_t . The arithmetic expressions of forgetting gate, input gate, and output gate are as follows:

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i), \\ f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f), \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), \end{aligned} \quad (1)$$

where f_t , i_t , and o_t represent the calculation results of LSTM forgetting gate, LSTM input gate, and LSTM output gate state, respectively.

W_f , W_i , and W_o are the weight matrices of LSTM forgetting gate, LSTM input gate, and LSTM output gate, respectively. b_f , b_i , and b_o are bias items of LSTM forgetting gate, LSTM input gate, and LSTM output gate, respectively. The final output of the LSTM is determined by the output of the output gate and the output of the cell state.

$$\begin{aligned} \hat{s}_t &= \tan h(W_s \cdot [h_{t-1}, x_t] + b_s), \\ s_t &= f_t \cdot s_{t-1} + i_t \cdot \hat{s}_t, \\ y_t &= o_t \cdot \tan(s_t), \end{aligned} \quad (2)$$

where s represents the input unit state of time t , W_s is the weight matrix of the input cell state, b_s is the offset term of the input unit, and $\tan h$ is the activation function type.

4. Deep Learning Adaptability of Bus Load Forecasting Problem

In the early years, foreign experts and scholars have done the work of substation bus load prediction [10–12]. Due to the increasing demand of power grid dispatching and the mature development of dispatching technology, bus load

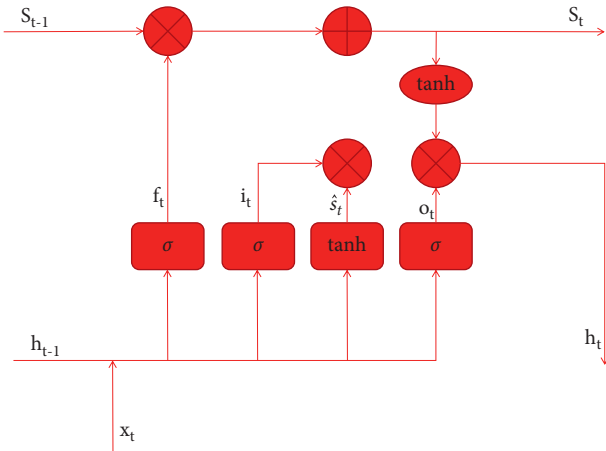


FIGURE 1: LSTM neuron structure.

forecasting is also widely used to solve the security and stability analysis of power system, reactive power optimization of power plant, and system dynamic state estimation.

Accurate bus load prediction is a prerequisite for the safety and stability of power system. In order to adjust generation plan reasonably, realize safety check, and implement energy-saving generation, it is very important to improve the accuracy of bus load forecast.

4.1. Characteristics of Bus Load Prediction. Compared with the traditional system load forecasting, bus load forecasting has its own features:

- (1) As the basic unit of system load prediction, bus load prediction has frequent changes compared with system load prediction. It has the characteristic of small prediction base, and the cardinality is much lower than the system load, which leads to the error of bus load prediction and difficult to improve the prediction accuracy.
- (2) Bus load prediction nodes are large, and the number of regional buses is large. For Laiwu power grid, there are 11 220 KV substation busbar loads to be processed. Each substation has a dual bus connection, resulting in 22 predictive nodes.

Large proportion of power grid in Laiwu area is connected to steel load. Its impact is very strong and difficult to find rules, and uncertain production plans and maintenance plans make local load forecasting work more difficult.

4.2. Bus Load Forecasting Process. The busbar load forecast of the substation shall be reported again after adjustment by the provincial company and the prefectural company. Weather, politics, social activities, holidays, and other factors affecting bus load change should be considered when adjusting. According to the accumulated historical load changes, the correlation between various indicators and load changes is comprehensively analyzed, and the factors affecting load changes and historical load data are used as the input of the

model. The target load, that is, the load value to be predicted, is the output target of training.

As mentioned above, deep learning is good at extracting required feature quantities from massive input information through multiple hidden layers. Finally, the training completes the task of load prediction, especially for short-term power load prediction. Load data are values that change with time. It is both nonlinear and sequential. In the process of load forecasting, the problem of long-term dependence on historical information often appears. The application of LSTM neural network in load prediction can prove its effectiveness theoretically. As a variant of RNN, the output of LSTM is not only related to the current input but also closely related to historical information. Therefore, LSTM has strong applicability in time series. In addition, the impact load data of steel studied in this article contain a lot of noise, burrs, and strong randomness. In theory, mathematical statistics are not as good as machine learning.

To sum up, deep learning method is suitable for charge prediction. Deep learning deals with the time series of power load itself and the nonlinear relationship between various influencing factors and charge.

5. Case Study

5.1. Experimental Data Sample. According to the preliminary investigation, the large steel enterprise is supplied by a number of substations with different capacities. In order to carry out this analysis, a total of 35,040 load data of several typical Laiwu 220 KV busbars connected to the largest energy consuming equipment or production line in 2018 were selected as the data set for simulation. A 15-minute interval is taken as a sampling interval, and 96 intervals of load data are taken for 24 hours per day. The change of load data needs to consider its busbar load operation mode and maintenance plan adjustment, seasonal change characteristics. Usually, 1–3 months is used as the interval time of model training. Due to the particularity of the historical load of the busbar in the area connected to large steel enterprises, the training time of the model is selected from 2 to 3 months. According to the typical seasons, the sample data set is the historical load data from April to May, June to July, and October to December.

5.2. Evaluation Indicators of Forecasting Method. Different power load forecasting models and methods usually have different performance. In order to compare the advantages and disadvantages of various prediction methods and the performance of models, it is important to design reasonable and feasible evaluation indexes for quantitative prediction performance. In this article, mean absolute percent error (MAPE) and root mean square error (RMSE) are selected as two reference indexes to judge accuracy.

The average absolute error evaluates the average absolute error between the predicted value and the real value, and its value range is $[0, +\infty)$. The smaller the value is, the better the model fitting effect is. The model is perfect if the error is zero.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_t(i) - f_p(i)|, \quad (3)$$

where $f_t(i)$ represents the simulated predicted value at time I , $f_p(i)$ represents the real load value at time I , and n indicates the input of load data.

Mean absolute percentage error. The reason why the mean absolute percentage error can describe the accuracy is that the mean absolute percentage error itself is often used to measure the accuracy of the statistical indicators, such as the prediction of time series.

The size of the mean square error is calculated from the sum of the squares of the errors. It measures the error between the predicted value of a model and the actual value of the load by means of the mean square error. The value is between $[0, +\infty)$, and the smaller the index value is, the better the model effect is.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n (f_t(i) - f_p(i))^2. \quad (4)$$

The root mean square error (RMSE) of large error has a stronger influence on the index, and the index will be more sensitive. This is a good indicator of the accuracy of the measurement. Therefore, RMSE is widely used as a standard for error prediction. The value range is $[0, +\infty)$. The smaller the value of the exponent, the better the effect of the model.

$$\text{RMAE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_t(i) - f_p(i))^2}. \quad (5)$$

In the essence of mathematical expression, mean absolute error, relative error, absolute error, and MAPE are the same. They are linear expressions of the predicted deviation from the true value. In this article, RMSE is selected as the performance evaluation index of the prediction method.

The error evaluation standard in this article is based on the load prediction evaluation standard of Shandong Electric Power Company of State Grid.

The formula for calculating the accuracy of the average daily load forecast is as follows:

- (1) Reference error of single bus load in period K
Reference error = (actual load – predicted load)/load reference value * 100%
The load reference value is 305 MW temporarily
- (2) Regional errors of all busbars in k period
Area error = root mean square of reference error of all buses
- (3) Accuracy of all busbars at all times of a day
Accuracy = (1 – root mean square of regional error for all time periods) * 100%

5.3. Deep Learning Model. The programming and simulation work in this article is based on MATLAB R2018b platform. The so-called network and deep learning algorithm is a deep

learning toolkit from MATLAB. The historical load data used in load forecasting have both long-term dependence and short-term dependence of time series, and the influence of other aspects cannot be ignored. This chapter uses historical load database and influence factor processing to carry out the learning and prediction work based on deep learning network.

5.3.1. Model Building. Because LSTM has a memory structure, it can well reflect the load data relationship in time series. When using deep learning LSTM networks in extremely short-term load prediction scenarios, an important step is to preprocess a large amount of historical load information. This solves the training difficulties caused by deep networks and many parameters. This chapter divides database into training data set and test data set, with a ratio 7 : 3. The former is used for deep learning network model learning process, and the latter is used for model performance test after training. Figure 2 shows the flowchart of load forecasting network based on LSTM.

5.3.2. Input and Output. In order to better meet the load forecasting demand of large steel enterprises, on the basis of general power load forecasting, the high sensitivity of industrial load to electricity price is taken into account. Large steel mills use lower loads during the day, when the electricity price is high, than at night at the cutoff point of the electricity price change at 8 a.m. The load it uses is significantly reduced. At the same time, steel load and steel market price changes are positively correlated. When the steel market booms, the steel load will increase accordingly. On the contrary, when the steel market downturn, steel load will be reduced accordingly. After a thorough examination, historical load data, changes in steel futures levels, date types, electricity price fluctuations, and weather changes are taken as influencing factors for load forecasting. Based on the historical data input in the LSTM network model, the influencing factors of the predicted date were used as the feature input template. Because every day there are points with large load fluctuations, the timing and amplitude of these points are characterized by very irregular. The bus load forecasting in Laiwu city can be used for reference. Two sets of input feature sets for load forecasting are summarized after considering various realistic situations. The former feature set network inputs the load data of 96 points in the 7 days before the forecast day and the date type, temperature, electricity price, steel price, and other information of the forecast day. The output data are the forecast daily load data of 96 points. The input data of the latter feature set network are the load data of 96 points 1 day before the forecast day and the date type, temperature, electricity price, steel price, and other information of the forecast day. The output data are the 96-point load data of the day to be predicted. The input feature set is shown in Table 1.

Input 1 is the historical load factor. The historical load information input of model 1 is 96 points of load data per day for 7 days starting from the day before the forecast date. The input of historical load information in model 2 is the

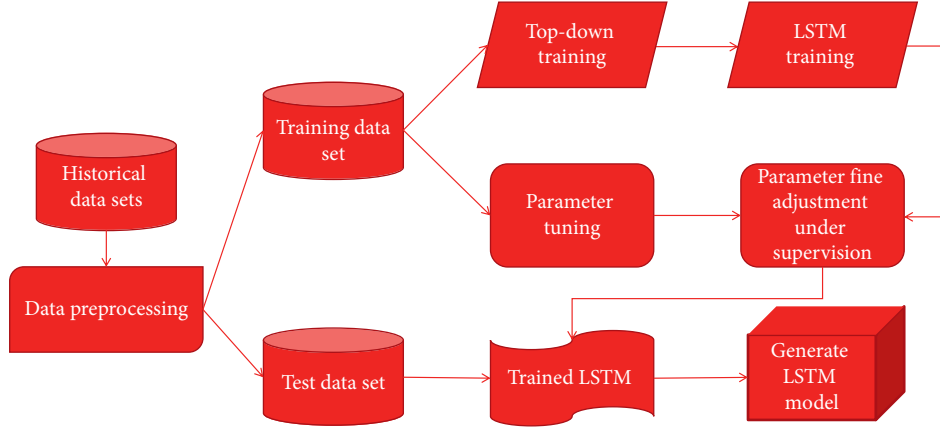


FIGURE 2: Load forecasting network based on LSTM.

TABLE 1: Input feature set.

Input sequence no.	The input character description	
	Model 1	Model 2
1	96-point daily load curve for 7 days before forecast	96-point daily load curve from 1 day before the forecast date
2	Date type of the day to be predicted	Date type of the day to be predicted
3	96-point temperature fluctuation on the predicted day	96-point temperature fluctuation on the predicted day
4	—	Weather conditions for the forecast day
5	96-point electricity price on forecast date	96-point electricity price on forecast date
6-7	Forecast the day before the price trend of steel	Forecast the day before the price trend of steel

load data of 96 points on the day before prediction. Input 2 indicates the date type (holiday or not) of the day to be predicted. Input 3 is meteorological condition, which mainly represents the 96-point fluctuation data of temperature on the day to be predicted. Input 4 is only reflected in Model 2, which is the weather condition of the day to be predicted. Input quantity 5 is the fluctuation of electricity price, which inputs the electricity price in different time periods of each day according to the price of peak and valley. Inputs 6-7 are the price change of steel futures, which shows the price fluctuation of rebar and H-beam.

Since the deep LSTM network has a storage unit, the loading data structure information can be encoded in the parameters of the LSTM network to preserve the timing of data. Let the historical load and influence factor data set be the matrix X of $m \times n$.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix}. \quad (6)$$

In the matrix, M is the step size of the training sample information, n is the number of input characteristic information, and X_{ij} represents the JTH characteristic information of the i^{th} input value. A day of data contains 96 lines. Take 1 day's data as the previous progress. The forward progress length of the longitudinal dimension input is 7, and the data volume of 7 days is entered into the model simultaneously. The $672 * 12$ matrix of the output data and the

96-point load data on the predicted day are the output of the LSTM network model.

5.3.3. Network Structure of the Model. In order to create a deep learning network structure suitable for different topics, we need to conduct centralized testing and debug the network parameters to the best, so that the network effect can meet the needs. When the number of layers and hidden units of LSTM network is too small, that is, the number of neurons in the network is too small, the network model to model building and feature extraction will not occupy a great advantage. When the number of neurons are too large, the learning efficiency of the network will decrease and over-fitting phenomenon will occur. Online learning time will also increase significantly. Therefore, when considering the STLF work in reality, the parameters of LSTM network used should be reasonably selected. Many experiments are carried out in this article, and the best parameters are selected according to the results. Figure 3 shows the LSTM network structure.

5.4. Simulation and Verification

5.4.1. Simulation Settings. In the process of building LSTM neural network, the neural layer and input layer of the network and the expression of Loss need to be defined first. Optimizer minimizes loss. The appropriate training data should be selected first to train the model. Selecting appropriate network parameters during training can accelerate

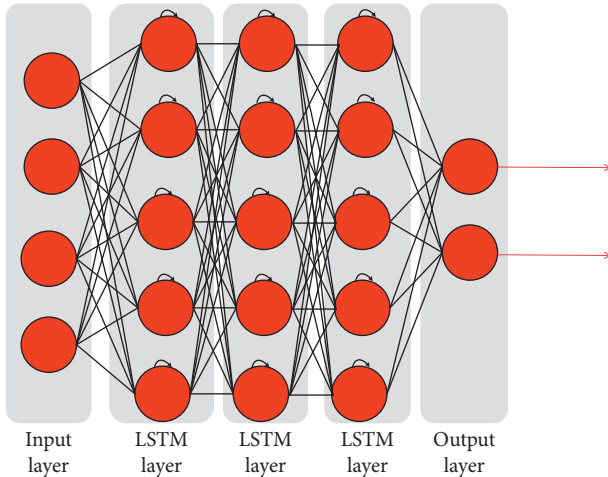


FIGURE 3: LSTM network structure diagram.

the running speed of recursive neural network in deep learning and improve the accuracy and reliability of load prediction. The analysis of parameter setting in deep learning neural network is as follows:

- (1) Input layer and output layer: the number of nodes in the input and output layers is related to the type of historical power load data and the type of influencing factors. In this article, the data set used in the simulation was selected for a 15-minute sampling interval and 96 load data per day. There are many factors that affect the prediction of power load. On the basis of the preprocessing of the historical load analysis, the data are normalized and the main factors affecting the load are taken as the input of the model.

Based on the analysis of the historical load prediction accuracy of bus lines in Laiwu area, the bus lines connected with steel shock load are divided into two types of research. The first bus has a certain steel impact load. Taking bus no. 2 of Shuanglong Station and bus No. 1 of Fangxia Station as examples, certain periodic regularity can be found, and the historical prediction accuracy is fair. The second bus is almost fully loaded with steel shock load. It is difficult to find cyclical regularity and is very sensitive to changes in the production plans of iron and steel enterprises. The accuracy of historical load prediction is not ideal for the #2 bus line of Gangcheng station and #2 bus line of Huihe Station.

On the basis of the above model input study, this article proposes two network models to improve the accuracy of load prediction of impact load bus of access steel.

First, the historical load factor of the first type model selects the load data of the day to be predicted, the date type of the day to be predicted, the 96-point temperature data of the day, the weather type of the day to be predicted, electricity price fluctuation data, and steel price data.

Second, in the second type of model, the historical load factor selects the load data of the 7 days before the forecast day, the date type of the forecast day, the 96-point temperature data of the same day, electricity price fluctuation data, and steel price data.

According to a lot of simulation experience, the first type of model performs better in bus load prediction which is more sensitive to external factors. The typical characteristics of this kind of busbar are that it is more affected by the change of production plan of iron and steel enterprises, it is difficult to find the periodic law, and the accuracy of historical load prediction is low. The second model performs better in bus load prediction with stable steel generation and production process. This kind of bus has a certain impact load of steel, and its historical regularity of load data is easier to grasp than the former one.

- (2) The number of nodes in the hidden layer is based on other researchers' research experiments. More hidden layers have better network training effect, but the training time will be longer. Since the training example in this article is a short-term charge prediction example, multiple hidden layers can be selected to deal with more complex cases. The actual network structure usually has no fewer than 100 neural units.
- (3) Learning rate of deep learning network is the convergence time of control function. It depends on the network read value and weight. With the increase in learning rate, the model training speed is accelerated. The stronger the effect of corresponding output error on parameters is, the more likely it is to oscillate and diverge. The learning rate decreases, the model training speed slows down, and the function is prone to over-fitting and over-converging.
- (4) Time step determines the learning span, that is, the input data depend on several consecutive input data.
- (5) The training of loss function (Loss) is to reduce the Loss value of training set and verification set. When the loss value is lower than a certain threshold or reaches over fitting, the training ends.
- (6) Training duration (EPOCHS): by reducing or increasing the number of trainings, under-fitting or over-fitting can be avoided.

Generalization ability of deep learning model is also a problem that needs to be considered in model training. Generalization ability refers to the ability of deep learning model to respond reasonably to updated data. A good model must have strong generalization ability. It is common to find that the model is too fit or not trained enough in the process of model training. Model over-fitting reflects the generalization ability of the model. In the case of too many training times or too little training data, the deep learning model can fit the training set data well. It cannot fit the test set data efficiently and well. In order to solve this problem, it is necessary to plan the duration and cycle of model training

TABLE 2: Typical bus load forecasting accuracy.

Forecasting object	Model 1 (%)	Model 2 (%)
Shuanglong station #2	97.07	95.98
Fangxia station #1	93.98	92.30
Gangcheng station #2	93.59	92.47
Huihe station #2	94.59	93.54

TABLE 3: Gangcheng station #2 load forecasting accuracy.

	Model 1 (%)	Model 2 (%)
Day 1	94.68	92.97
Day 2	95.84	93.49
Day 3	94.12	93.71
Day 4	93.41	92.26
Day 5	91.96	91.97
Day 6	93.42	93.22
Day 7	91.68	90.25

TABLE 4: Load forecasting accuracy (from April to May).

Forecasting object	Forecasting method of deep learning (%)	Existing methods (%)
Shuanglong station #2	96.66	96.41
Fangxia station #1	93.98	93.38
Gangcheng station #2	92.67	93.14
Huihe station #2	96.89	96.88

TABLE 5: Load forecasting accuracy (from June to July).

Forecasting object	Forecasting method of deep learning (%)	Existing methods (%)
Shuanglong station #2	97.07	96.61
Fangxia station #1	93.98	93.02
Gangcheng station #2	93.59	94.01
Huihe station #2	94.59	94.07

TABLE 6: Load forecasting accuracy (from October to December).

Forecasting object	Forecasting method of deep learning (%)	Existing methods (%)
Shuanglong station #2	97.74	96.38
Fangxia station #1	96.22	91.63
Gangcheng station #2	97.25	96.12
Huihe station #2	98.32	96.19

reasonably and effectively. Training must be stopped before the inflection point of test loss and training loss occurs.

According to the simulation experience, relatively small learning rate is better to ensure the stability of the system. The selection range of learning rate is between [0.01, 0.8]. This is to observe the variation trend of training set and validation set error and the accuracy of load prediction results. Through a large number of simulation experiments in the early stage, the parameters of the deep learning model were constantly adjusted and finally the learning rate was set as 0.01 and the training duration as 800. In this LSTM network structure, the hidden layer is set as three layers, namely three LSTM layers, each layer is set with 150 hidden nodes, namely 150 long short-term memory units. The last layer is the full connection layer, which serves as the output layer of the model.

5.4.2. Simulation Results. Two models are used to predict the typical bus, respectively, and the results are as shown in Table 2.

Taking Gangcheng station #2 bus line as an example, the variation of the load prediction model on the forecasting error in 7 days is analyzed, which is shown in Table 3.

Considering the changes of bus load in the area of steel shock load, such as operation mode adjustment, maintenance plan, equipment start-up and maintenance of iron and steel enterprises, and seasonal influence, three data sets are set according to the typicality, which are the historical load data sets of each bus in Laiwu area from April to May, from June to July, and from October to December in 2018, respectively. It carries on training and testing work with this input model. The results are compared with the current load

forecasting method of Laiwu power grid obtained from preliminary investigation, as shown in Tables 4–6.

The existing regional forecasting method only considers the correlation between time factor and load. After the forecast, operators need to further adjust the forecast results according to the production plan of steel enterprises to obtain the load prediction curve. The internal and external factors affecting load variation including the characteristics of large steel enterprises are not considered. Therefore, the load prediction accuracy of the deep learning LSTM prediction algorithm used in this article, which considers a variety of influencing factors, is significantly improved compared with the traditional method.

The load prediction accuracy of Shuanglong bus No. 2 and Fangxia bus No. 1 increased by about 1%. The load carried by the busbars of the two substations fluctuates less than 50 MW. It has some regularity. Compared with the other three data sets, the improvement in the accuracy of Huihe 2# bus shows obvious difference, which is related to the fluctuation of quarterly load. When the fluctuation is small, the accuracy will be significantly improved, and when the fluctuation is large, the accuracy will be slightly improved. With regard to Gangcheng Station no. 2, the busbar carries the load of Dongling station and Section steel station of large iron and steel enterprises, so the load fluctuates greatly. In some special conditions, the instantaneous load fluctuation can exceed 100 MW, so the prediction method alone cannot meet the requirement of improving the prediction accuracy.

To further improve the accuracy of the prediction, the following direction of production planning and manual intervention is considered:

- (1) For the rapid load changes caused by the impact load superposition of large steel enterprises, special workers need to refer to the production process and production plan of steel mills. It modifies or replaces wave point data with boundary values.
- (2) Sudden load changes caused by accident trip and maintenance of electrical equipment require field operators to revise the prediction results according to the actual situation.

6. Conclusions

At present, the forecasting of steel impact load is mainly STLF, and the identification of influencing factors is mainly limited to production plan, maintenance plan, and other factors related to the production of iron and steel enterprises. Due to the complexity and high nonlinearity of influencing factors in the process of STLF, the traditional forecasting method has simple mathematical model, difficult to adjust parameters flexibly, and poor adaptability and relatively weak ability to reflect load changes. Therefore, it is difficult to further improve the prediction accuracy.

Compared with the mainstream artificial neural network algorithm and support vector machine algorithm, the

impulse power load forecasting based on improved recurrent neural networks (LSTM forecasting algorithm based on deep learning network) proposed in this article can improve the accuracy of load forecasting. This point also directly proves the advanced nature of deep learning and its characteristics for dealing with load forecasting problems.

Nonlinear load is one of the main pollutants affecting the quality of power system. The characteristic power of high fluctuation load changes rapidly, and the load curve presents sawtooth wave. When the amplitude of load change is larger than the system capacity, it will cause continuous oscillation of system frequency and large fluctuation of voltage, which will have an adverse impact on the power system. The regularity of this kind of system load is usually very poor. It is very important to improve the accuracy of this kind of load forecasting. Future research will focus on this aspect.

Data Availability

The data set can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] I. Moghram and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1484–1491, 1989.
- [2] N. Amjady and F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," *Energy*, vol. 34, no. 1, pp. 46–57, 2009.
- [3] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines [J]," *IEEE Transactions on Power Systems*, vol. 28, 2013.
- [4] L. Li, C. Mu, S. Ding, Z. Wang, R. Mo, and Y. Song, "A robust weighted combination forecasting method based on forecast model filtering and adaptive variable weight determination," *Energies*, vol. 9, p. 20, 2016.
- [5] L. Yunfei, Z. Peng, C. Pengfei et al., "Exploration and practice of short term load forecasting based on large data mining under impulse load characteristics," *Power Systems and Big Data*, vol. 22, no. 4, pp. 80–86, 2019.
- [6] J. Yang and J. Stenzel, "Application of two-dimensional support vector machine in short-term Load forecasting," *Power Tech*, IEEE, Russia, 2005.
- [7] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *Computer Science*, vol. 5, no. 1, p. 36, 2015.
- [8] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proceedings of the 15th Annual Conference of the International-Speech-Communication-Association (INTERSPEECH 2014)*, Singapore, Singapore, September 2014.
- [9] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proceedings of the 2015 IEEE*

International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Queensland, Australia, 19 April 2015.

- [10] N. Amjady, "Short-term bus load forecasting of power systems by a new hybrid method," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 333–341, 2007.
- [11] X. Chen, C. Kang, X. Tong, Q. Xia, and J. Yang, "Improving the accuracy of bus load forecasting by a two-stage bad data identification method," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1634–1641, 2014.
- [12] I. P. Panapakidis, "Clustering based day-ahead and hour-ahead bus load forecasting models," *International Journal of Electrical Power & Energy Systems*, vol. 80, pp. 171–178, 2016.

Research Article

TransEffiDet: Aircraft Detection and Classification in Aerial Images Based on EfficientDet and Transformer

Yanfeng Wang,¹ Tao Wang,¹ Xin Zhou ,¹ Weiwei Cai ,^{2,3,4} Runmin Liu ,^{4,5} Meigen Huang,¹ Tian Jing,¹ Mu Lin,¹ Hua He,¹ Weiping Wang,¹ and Yifan Zhu¹

¹College of Systems Engineering, National University of Defense Technology, Changsha 410082, China

²School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

³Graduate School, Northern Arizona University, Flagstaff, AZ 86011, USA

⁴AiTech Artificial Intelligence Research Institute, Changsha 410000, China

⁵College of Sports Engineering & Information Technology, Wuhan Sports University, Wuhan 430079, China

Correspondence should be addressed to Xin Zhou; zhouxin09@nudt.edu.cn

Received 1 March 2022; Revised 20 March 2022; Accepted 29 March 2022; Published 21 April 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yanfeng 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.

In recent years, analysis and optimization algorithm based on image data is a research hotspot. Aircraft detection based on aerial images can provide data support for accurately attacking military targets. Although many efforts have been devoted, it is still challenging due to the poor environment, the vastness of the sky background, and so on. This paper proposes an aircraft detection method named TransEffiDet in aerial images based on the EfficientDet method and Transformer module. We improved the EfficientDet algorithm by combining it with the Transformer which models the long-range dependency for the feature maps. Specifically, we first employ EfficientDet as the backbone network, which can efficiently fuse the different scale feature maps. Then, deformable Transformer is used to analyze the long-range correlation for global feature extraction. Furthermore, we designed a fusion module to fuse the long-range and short-range features extracted by EfficientDet and deformable Transformer, respectively. Finally, object class is produced by feeding the feature map to the class prediction net and the bounding box predictions are generated by feeding these fused features to the box prediction net. The mean Average Precision (mAP) is 86.6%, which outperforms the EfficientDet by 5.8%. The experiment shows that TransEffiDet is more robust than other methods. Additionally, we have established a public aerial dataset for aircraft detection, which will be released along with this paper.

1. Introduction

Analysis and optimization algorithm based on image data is a hot issue in recent years. Image processing is not only used in the civil field, but also widely used in the military field. In the military field, aerial images and remote sensing images are used to detect aircraft objects in military bases and airports in war. It is of great significance to intelligence deployment and strategic deployment. Through the acquired images, the commander can quickly and accurately understand the number of enemy aircraft on the battlefield and the take-off and landing situation. Location distribution can provide a strong information security guarantee for the follow-up operational decision and play an essential role in

winning the war [1]. Therefore, aircraft detection in images is very popular in the field of military research. In addition, the classification of military and civil aircraft is also very important, which may reduce unarmed civilian casualties. These tasks must be reliable to automate site analysis, especially to export alerts corresponding to abnormal events. Therefore, how to accurately detect and classify aircraft in aerial images has high research value.

Traditional object detection requires manual feature extraction, and classifiers are designed and trained for specific detection objects. It is difficult for this kind of method to obtain robust solid features and it is very sensitive to external environmental noise, so it has significant limitations in engineering applications.

Object detection technology based on convolution neural networks is developing rapidly with the development of hardware. The two-stage method which is represented by R-CNN [2] and the single-stage method which is represented by SSD and YOLO [3] are two mainstream frameworks of object detection techniques based on convolution neural networks.

In recent years, many scholars have optimized and improved the internal backbone network, anchor design, region feature coding, and other submodules based on the two frameworks, which effectively enhance the performance of the object detection method. Some researchers have also proposed an object detection framework based on object key points [4] and have made remarkable achievements in each big data set.

However, it still faces many challenges for aircraft detection in aerial images, for example, the poor image quality due to poor environment, similar shapes of different types of aircraft, and the vastness of the sky background. The factors require more robust technologies and systems that can reliably detect and classify aircraft using characteristics. Artificial intelligence (AI) and Deep Learning (DL) can play an important role. The AI system, which is usually based on the DL, can automatically detect the aircraft, thereby enhancing military situational awareness [5].

In this paper, a robust aircraft detection method is proposed to solve the aircraft detection problems. This method is a hybrid solution based on EfficientDet and Transformer. We improved and optimized the EfficientDet algorithm by using Transformer, which models long-range dependency for feature maps, and then a feature fusion module is proposed to effectively capture both the long-range and short-range context generated by the Transformer module and efficient backbone. In addition, most aircraft detection methods are to detect targets in remote sensing images. However, it is significantly essential to detect aircraft in aerial images. Therefore, we explore aircraft detection in aerial images in this paper.

The novelty can be summarized into the following four points:

- (1) Create a hybrid structure by combining the power of CNNs with Transformers to extract multiscale and multidimensional feature representations. In this way, the long-range dependency of features is modeled.
- (2) Design a feature fusion module which can fuse features extracted from different deformable Transformer layers, so that the backbone can fuse global context information and abundant local information effectively.
- (3) Use the deformable attention modules to reduce the computational complexity. Therefore, our method can handle higher dimensional feature maps.
- (4) Obtain superior performance on the task of aircraft detection in aerial images. In addition, it is a challenging and labor-intensive task to label all the aircraft in an image. To promote the development of

aircraft detection in aerial images, we will release this dataset along with this paper, which contains five types of aircraft: armed helicopters, bombers, fighter jets, early warning aircraft, and passenger aircraft, with a total of 2558 images.

2. Related Work

2.1. Traditional Machine Learning Method. Traditional machine learning methods mainly implement object detection through the following steps: constructing a training data set, region extraction, feature design and extraction, feature processing, similarity measurement selection, classifier design, training, and detection.

The main limitations of traditional machine learning methods are region extraction, feature extraction, feature processing, and classifier design.

In the search strategy of candidate region extraction, the common methods are the sliding window method, such as [6]. Feature extraction refers to feature fusion and dimension reduction. The feature extraction of the object essentially maps the high-dimensional information to low-dimensional feature space, which becomes the basis of object detection [7]. There are generally two methods for feature processing: feature fusion [8] and feature dimensionality reduction [9].

2.2. Artificial Neural Network. The neural networks do not need to manually design features and automatically extract features from samples through a trained network. Using only a single feature as a feature extractor is inefficient and cannot meet the needs of rapid detection in aircraft object detection tasks, and the deviation and asynchrony between different feature extractors will reduce the effectiveness of the training process [10]. Therefore, more research focuses on establishing end-to-end networks. Deep learning research has continued to develop in object detection field. These methods with powerful feature representation capabilities regards object detection as the classification problem of regions of interest using deep features, uses deep network architecture to obtain image features from input data automatically, and classifies images at the output layer. These feature maps extract rich semantic features and have strong feature representation capabilities.

According to the usage of the data set, the network structure, and different application scenarios, deep learning has derived a variety of methods, such as CNN-based image object detection network, Deep Belief Network (DBN) [11], etc.

2.3. Optimization of Object Detection Framework. The object detection framework generally includes backbone network, neck connection layer, anchor, region feature coding, classification and location head, loss function, and other submodules. In addition, different models have their unique submodules. The performance of the object detection method can be effectively improved using reasonable optimization.

Residual structure [12], which is used to increase network depth, can effectively improve accuracy of convolution networks, but the number of parameters increases exponentially. For this reason, Xie et al. [13] integrated the residual structure with the Inception structure [14]. The Inception structure increases the depth and width of the network by sharing the width direction parameters. This method improves the network accuracy while effectively controlling the growth of the number of parameters.

Optimization of Anchor Design: Anchors are rectangular boxes with different sizes and proportions generated in each grid on the feature graph. The single-stage detection framework generates the object boundary box directly based on the anchors, and the two-stage detection framework obtains the candidate box by fine-tuning the positive anchors.

Anchor frames with different scales are suitable for detecting different objects. For this reason, Zhu et al. [15] proposed an anchor frame design strategy based on step size reduction. The high-resolution feature image has a small receptive field and is used to detect small-scale objects. To prevent missed detection, the step size generated by the anchor frame should be reduced to increase the density of the anchor frame. Xie et al. [16] proposed a dimensionally decomposable region recommendation network. This method decomposes the anchors in the dimension, thus effectively solving the detection of proportional special objects.

To optimize the limitations of anchor-based methods in allocating positive and negative samples and dealing with multiscale problems, many scholars have proposed a target detection model without anchors.

Most of these models carry out pixel-level classification and regression on the different scale feature images to replace the anchor box. Tian et al. [17] first calculate the position of each point on the feature map that maps back to original images, then distribute the different samples, and define the center degree to reduce the fractional weight when predicting the edge position of the instance frame. As a result, the influence of the low-quality prediction box on the detection results is suppressed and the detection performance of the model is improved.

Optimization of nonmaximum suppression algorithm: In object detection, nonmaximum suppression means that, in the forward reasoning stage, the highest-confidence candidate box is selected as the final result. Besides, the surrounding candidate boxes whose intersection and union ratio are greater than the threshold are eliminated. For the same detected object, this method can eliminate other nonoptimal candidate results and avoid relocation.

To solve the issues, Bodla et al. [18] proposed an algorithm for soft suppression of SoftNMS. It reduces the confidence of the first n nonoptimal candidate boxes instead of eliminating them directly. The confidence of the candidate box is not strongly related to the intersection and union ratio, only considering that the classification confidence is one-sided. He et al. [19] improved the soft suppression algorithm and incorporated the location confidence into it to indicate the credibility of the

coincidence of the current candidate box and the instance box. It models the candidate box and the instance box, respectively, and uses KL divergence to measure the distance. In addition, Liu et al. [20] designed a subnetwork containing only the fully connected and convolution layer, which is used to determine whether the nonoptimal candidate box whose intersection and merge ratio is greater than the threshold predicts the same target as the optimal candidate box and retains the nonoptimal candidate box with different detection targets, so as to effectively avoid the disadvantages of the traditional methods.

Optimization of positive and negative sample sampling algorithm: There are a large number of anchor frames on large-scale feature maps, and most of the negative sample anchor boxes provide similar gradient information, so all of them are used in classification and regression training to waste computational resources. Therefore, sampling all anchor boxes is required. Only some of them are selected to participate in the training.

Since the number of positive anchor boxes is much smaller than that of negative anchor boxes, it is easy to cause the imbalance between positive and negative training samples by sampling randomly in the whole world. The SSD model sorts the negative samples according to the confidence error and updates the model for the difficult negative samples with low confidence.

Different from the SSD model, Shrivastava et al. [21] select complex negative samples online according to the loss of input samples. They extended the two-stage detection framework, designed another RoI network to calculate the loss of input samples, and reduced the order of input loss, and selected the first n negative samples with the largest loss for model training. The advantage of using input loss as a standard to measure the difficulty of sample learning is that it can consider both the difficulty of classification and regression. Inspired by the above research, Yu et al. [22] adopted a similar method to optimize the positive and negative sample sampling of the single-stage detection framework. It directly filters simple samples and only backpropagates the k samples with the largest loss to update the network parameters. The difficult negative sample can also be represented by the intersection and union ratio with the instance box.

3. Material and Method

This paper proposes an improved object detection and classification architecture named TransEffiDet, which is based on EfficientDet [23] and Transformer [24]. The deep neural network is helpful to extract high-level information. However, gradient vanishing is prone to occur simply by deepening the number of layers in the network. In other words, the EfficientDet cannot model the long-range dependence, because of the network architecture. However, the proposed TransEffiDet has the improved Deformable Transformer module. With the proposed feature fusion strategy, the Transformer can better model the long-term dependence for the generated features, and the performance of TransEffiDet is improved.

3.1. Experimental DATA. To promote the development of aircraft detection in aerial images, we have established a data set, named Military Aircraft Detection in Aerial Images (MADAI), containing four types of military aircraft: fighter jets, armed helicopters, bombers, and early warning aircraft, as shown in Figure 1. In addition to military aircraft, we have added passenger aircraft to the dataset, which helps to distinguish military and civilian aircraft, thereby reducing civilian casualties. Therefore, there are five types of aircraft in the MADAI dataset: fighter jets, armed helicopters, bombers, early warning aircraft, and passenger aircraft. There are a total of 2558 images in this dataset. The typical resolutions are 1600×1024 , 3500×2280 pixels. Table 1 shows the number of training and testing images. The MADAI contains the images and the ground truth. The ground truth, annotated by three experienced experts, indicates the type and location of the aircraft. The MADAI dataset can not only be utilized to measure the accuracy of the aircraft detection algorithm, but also provide ideas for subsequent object strikes in the military field. This dataset and annotations can be obtained from <https://github.com/wangyanfeng231/TransEffiDet>.

Moreover, each image in our dataset may contain multiple types of aircraft rather than just one type, and each image contains a different number of aircraft objects. This is consistent with the actual military use because military aircraft usually operates jointly, such as early warning aircraft and fighters, bombers and fighters, as shown in Figure 2.

3.2. TransEffiDet Architecture. The architecture of TransEffiDet is shown in Figure 3. This architecture follows the paradigm of single-stage objection detection methods.

We employ EfficientDet as the backbone network, which can efficiently fuse the different layer multiscale feature maps. BiFPN, which is detailed in Section 3.3, is unitized as the feature network. We take the 3–7 level features of the network and repeatedly apply feature fusion model BiFPN to extract the context features. Moreover, to model the long-term dependence of different layer features, we add a 12x deformable Transformer between the P5 and P6. The diagram of Transformer and feature fusion module are shown in Figure 4. Finally, object class is produced via a class prediction net, and bounding box predictions are generated by feeding these fused features to box prediction net. Similar to [25], in all levels of features, the class prediction net and the box prediction net shared the same weights.

3.3. BiFPN. BiFPN has made a number of optimizations for multiscale connections to improve the efficiency. First, compared with the PANet [26] (Figure 5(a)), the node with only one input is deleted, because these nodes have little benefit to the fusion of different level features. Second, in each layer, an additional edge is added liked ResNet, which connects the input and output. This setting can integrate more features. Third, the BiFPN is repeated to produce more high-level features. See reference [23] for details.

We used the fast normalized fusion. Since the network does not have the Softmax in fast normalized fusion, it is more efficient.

3.4. Transformer Module. Since the intrinsic locality of convolutional networks, convolutional neural networks cannot effectively model the long-range dependency between pixels. Therefore, deformable Transformer encoder (DeTrans-encoder) layer is constructed to map long-term dependency between pixels to effectively extract local as well as global semantic information. The deformable Transformer introduces a deformable self-attention mechanism (DMSA) which can reduce the hardware requirement. DeTrans-encoder layer consists of an Input-to-Sequence Transformation (IST) layer and a deformable Transformer layer.

IST layer is necessary to first convert the feature map generated by the CNN-encoder {P5} into a one-dimensional sequence, because the Transformer only processes the input by a sequence form. However, directly converting a 3D feature map into a 1D sequence will result in the loss of spatial features. Hence, the sine and cosine functions are used to obtain a 3D position-encoded sequence, shown as follows:

$$\begin{aligned} PE_{(local,2i)} &= \sin\left(\frac{local}{10000^{2i/d_{model}}}\right), \\ PE_{(local,2i+1)} &= \cos\left(\frac{local}{10000^{2i/d_{model}}}\right), \end{aligned} \quad (1)$$

where $local$ is the location, and d_{model} is the dimension of the CNN feature map. We obtain the input sequence of DeTrans-encoder by summing the extracted features with the position-encoded sequence.

3.4.1. DMSA Layer. The self-attention mechanism of the Vanilla Transformer focuses on all possible positions of the feature map, which causes the network to converge more difficult. In addition, the computational and spatial complexities of the self-attentive mechanism grow squarely with the image size, making it difficult to handle high-resolution features. To address these issues, we use a deformable self-attentive mechanism that can significantly reduce the complexity by focusing on only a small number of sampled locations instead of all locations of the feature map based on automatically determined reference points [27].

We have an input feature map $f_i \in R^{C \times H \times W}$, let q be a query element with a feature value $z_q \in R^C$, and p_q is a 2-d reference coordinate point. The deformable attention feature can be calculated as

$$\text{DeformAttn}(z_p, p_q, x) = \sum_{m=1}^M W_m \left[\sum_{k=1}^K A_{mqk} \cdot W'_m x(p_q + \Delta p_{mqk}) \right], \quad (2)$$

where m is the attention heads, k is the sampled keys, and K represents the total sampled key number, where $K \ll D \times H \times W$ thereby not only speeding up the convergence but also significantly reducing the computational and spatial complexities. The $\Delta p_{mqk} \in [0, 1]$ and A_{mqk} ($\sum_{k=1}^K A_{mqk} = 1$) are the sampling offset and attention weight of the k -th



FIGURE 1: The MADAI dataset we released. MADAI contains four types of military aircraft (fighter jets, armed helicopters, bombers, and early warning aircraft) and one type of civil aircraft (passenger aircraft).

TABLE 1: The detailed numbers of each type of aircraft.

	Bomber	Passenger aircraft	Early warning aircraft	Fighter	Armed helicopter	Total
Training	348	378	318	481	494	2019
Testing	100	97	97	135	110	539
Total	448	475	415	616	604	2558

sampling point in the m -th attention head, respectively. For a point $p_q + \Delta p_{mqk}$ that does not fall in integer coordinate positions, we use bilinear interpolation to obtain the final result $x(p_q + \Delta p_{mqk})$. Δp_{mqk} and A_{mqk} are obtained by performing a linear transformation of z_q .

3.4.2. Deformable Transformer Layer. Based on the above image-sequence transformation and the deformable attention mechanism, the attentional feature map can be obtained, and the deformable Transformer layer can be

obtained by a feedforward network. Layer normalization operations are performed after both the attention layer and the feedforward layer, and skip connections are used in each layer to prevent the gradient from vanishing. The deformable Transformer layer is obtained by cascading multiple deformable Transformers.

3.4.3. Feature Fusion Module. Figure 6 shows the detailed dimension changes for the Transformer, which is designed to solve the dimensional inconsistency between feature maps of

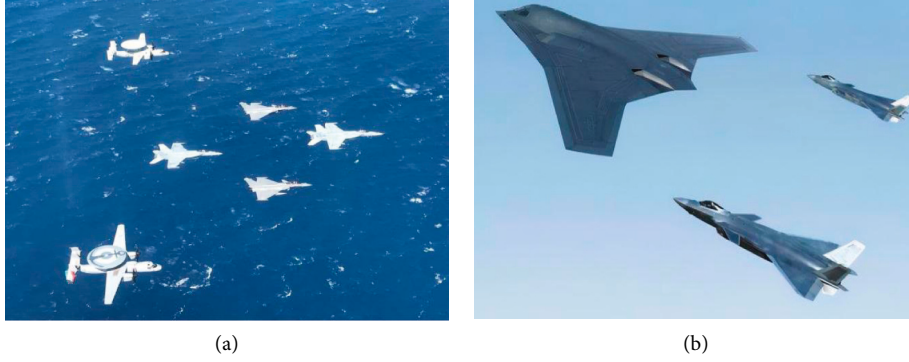


FIGURE 2: Joint operations of different types of military aircraft. (a) Joint military operations of early warning aircraft and fighter jets. (b) Joint military operations of bombers and fighter jets.

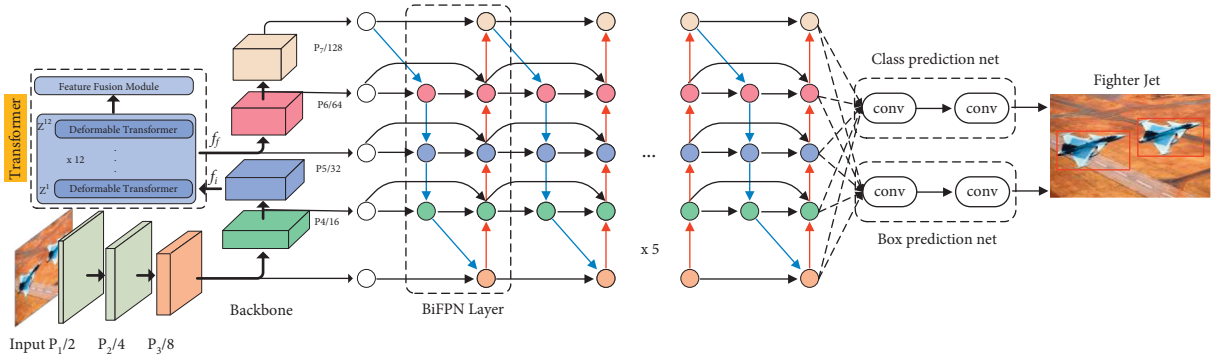


FIGURE 3: Diagram of TransEffiDet. EfficientDet [23] is backbone; BiFPN is feature extraction network. The Transformer modules are added between the P5 and P6 layers.

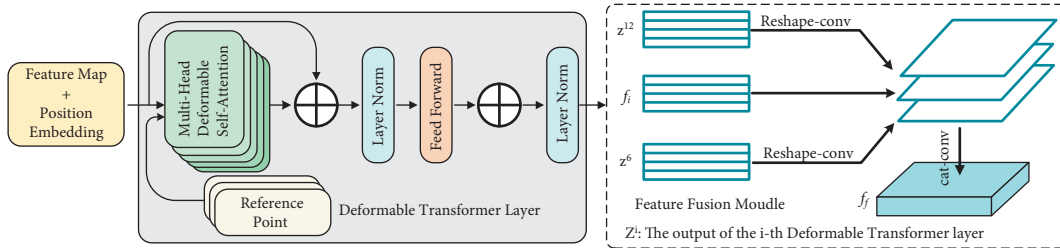


FIGURE 4: Diagram of Transformer and feature fusion module.

the Transformer and CNN backbone. Firstly, the feature map f_i of level 5 (P5) of CNN-encoder is flattened into a 1D sequence. Then the flattened sequence is sent to the Deformable Transformer introduced above. The output of the Transformer is also the 1D sequence, whereas the CNN backbone receives a 2D feature map as the input. Therefore, we reshape the 1D sequence to the original dimension of P5. To effectively fuse the global context generated by different Transformer layers and local information extracted by EfficientDet backbone, we designed a feature fusion module to capture both the global and local context. Specifically, assuming that the Transformer module is composed of L_e ($L_e = 12$) layers, we take out N features uniformly $\{Z^n\} (n \in \{L_e/N, 2L_e/N, \dots, NL_e/N\})$ with a step length L_e/N as the input of the feature fusion module, and N sets to 2 in this paper. For each feature sequence outputted by deformable Transformer layer, we first reshape it

into a 2-dimensional feature map with the same size of P5. And then, the convolutional operation is employed to each 2-dimensional feature map, and the output channel of this convolution is halved. In order to combine the global context information modeled by Transformer and the rich semantic information extracted by EfficientDet backbone, we concatenate all the feature maps with halved channels and the input feature map P5 to obtain a feature map f_o . Finally, the feature map f_o is fed to a convolutional layer and the same channel to perform adaptive feature calibration to obtain the final fusion feature f_f .

In this paper, the output of the sixth and twelfth Transformer layers and f_i are used to produce the input of the layer {P6}, which can better obtain the characteristics of different layers and keep the balance between the calculation and the efficiency. Specifically, these two resized feature

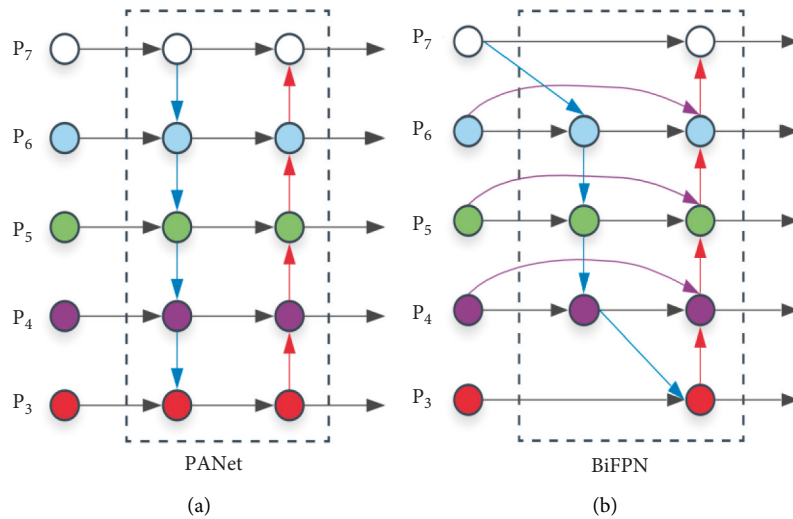


FIGURE 5: PANet and BiFPN architectures. (a) The PANet. (b) The BiFPN used in TransEffiDet.

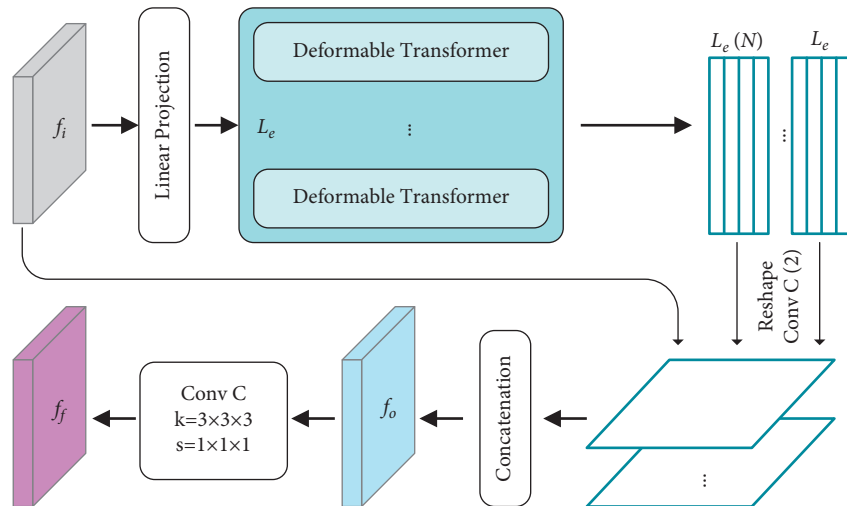


FIGURE 6: The detailed dimension changes for Transformer.

maps are concatenated to produce f_o . Finally, the feature f_f is obtained by a convolution function.

3.5. Implementation Details

3.5.1. Data Augmentation. We used data augmentation method to realize different feature learning by adding different feature variables to the image. To expand the training set while preserving the basic features, data augmentation is carefully applied to get some new images. Various random changes were included, including movement, rotation, zooming, and horizontal/vertical flip.

3.5.2. Pretrained Weights. In EfficientDet, the networks are pretrained using the ImageNet. Following this tradition, all models of TransEffiDet are pretrained on ImageNet. Then it is fine-tuned on our datasets. Since the high-order features are learned on the ImageNet, which is different from the

MADAI dataset, we retrained some convolution blocks to fine-tune the weight of the classification task.

3.5.3. Other Details. The images are resized to 768×768 pixels to reduce memory requirements during training. The TransEffiDet architecture is implemented using PyTorch. We used SGD optimizer with 0.9 momentum. The validation set is 25% of the training data. To facilitate a fair comparison, the metrics provided in this paper correspond to the best performance in the validation and training dataset, so the performance of the proposed method is not, in any way, optimized for the test datasets. The source code is publicly available at <https://github.com/wangyanfeng231/TransEffiDet>.

4. Results and Discussion

4.1. Experimental Results. In this section, the performance of TransEffiDet is evaluated on our dataset. Both EfficientDet and the proposed TransEffiDet have been retrained using the

TABLE 2: The performance measures for the EfficientDet and TransEffiDet.

	AP					mAP
	Bomber	Early warning aircraft	Fighter	Armed helicopter	Passenger aircraft	
TransEffiDet	55.4	98.9	84.9	98.2	95.3	86.6
EfficientDet	37.7	98.8	76.5	95.7	95.2	80.8

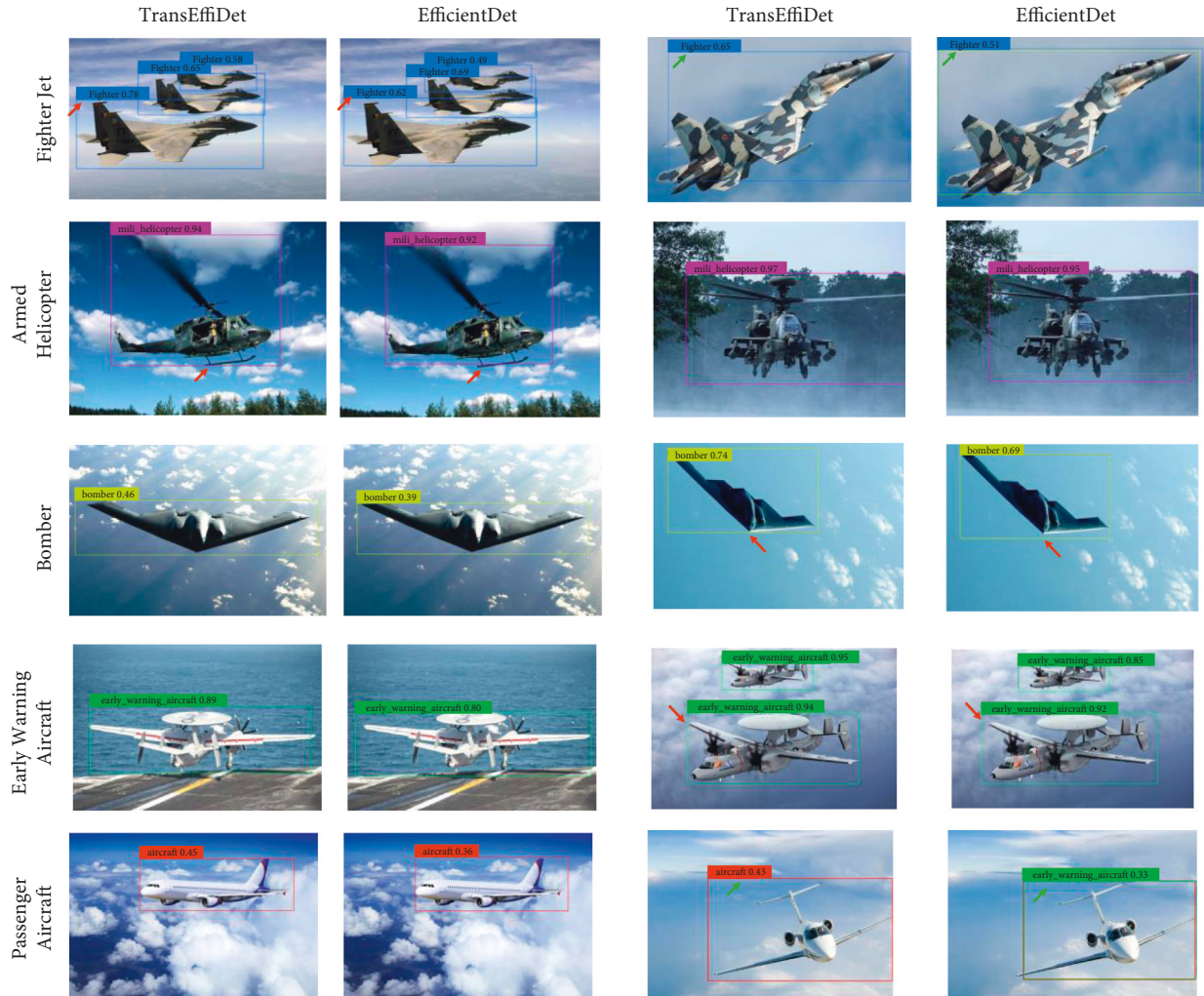


FIGURE 7: Examples of aircraft detection results in the MADAI dataset.

MADAI dataset. Detection performance measures are shown in Table 2. Compared with EfficientDet, the Average Precisions (AP) of Bomber, Fighter, and Armed Helicopter are improved by 17.7%, 8.4%, and 2.5%, respectively. In total, the mean Average Precision (mAP) is improved by about 5.8%. From the quantitative results, we can see that the performance of TransEffiDet is better than EfficientDet.

The aircraft detection results of TransEffiDet and EfficientDet are illustrated, as shown in Figure 7. Each type of aircraft is marked with different color boxes, and the numbers next to the boxes represent confidence. We can see that most of the aircraft are accurately detected, especially the early warning aircraft. This shows the advantage of this method; that is, it can detect aircraft accurately.

Compared with EfficientDet, TransEffiDet can achieve more accurate detection and the detected box can achieve better precision (red arrows in Figure 7). The detection box produced by EfficientDet is larger or smaller than the real object, which leads to lower accuracy. Moreover, the method EfficientDet will produce some false positives of aircraft. This is mainly because some types of aircraft are similar such as bomber and fighter, early warning aircraft, and passenger aircraft, and so on. Therefore, it is hard for the network to detect these similar objects. However, the proposed TransEffiDet can handle this problem well, because the Transformer can provide the long-term relationship and further make the network focus on the aircraft's features.

TABLE 3: The ablation study of the proposed method on all test datasets.

Models	Ablation type										mAp (%)
	Input	Input/half	Z^6	Z^6 /half	Z^{12}	Z^{12} /half	Z^4, Z^8, Z^{12} /half	Cat	Add	Cat/Add	
Model 1					✓						76.46
Model 2				✓		✓		✓			81.03
Model 3	✓		✓		✓				✓		82.67
Model 4	✓		✓		✓			✓			82.86
Model 5	✓			✓		✓				✓	82.84
Model 6		✓		✓		✓		✓			81.67
Model 7	✓						✓	✓			85.06
Model 8 (proposed)	✓			✓		✓		✓			86.55

4.2. Ablation Study. The ablation study is conducted on all test datasets to illustrate the effectiveness of the feature fusion module, as shown in Table 3. We explore the fusion i.e., concatenation (Cat), add (Add), of the input and Transformer outputs of different layers (Z^4, Z^6, Z^8, Z^{12}) to obtain the optimal fusion feature representation. All outputs Z^l of the Transformer are fed into a convolution layer (kernel size 3×3) with ($*/\text{half}$) and without the feature channels halved operation. $*$ represents the input and the outputs of different Transformer layers.

We investigated the contribution of the outputs of different Transformer layers and the input. From the results of Models 1, 2, and 8 in Table 3, we can see a clear trend that the more the feature maps are added, the better the performance is. Compared with Model 1, Model 2 significantly improves the performance by adding a feature map of the Transformer middle layer. The further improved performance was obtained by Model 4 by introducing the input into the final fusion feature.

Furthermore, to explore the influence of the feature fusion way, Models 3, 4, and 5 were built. We can see that the concatenation fusion way can obtain relatively good detection results. This is because simple addition cannot integrate different feature maps well. Finally, we built Models 6 and 7 to further demonstrate the effectiveness of the proposed feature fusion module. Combining the long-term modeling capabilities of the Transformer and the abundant local information of the feature maps results in the good performance of Model 8.

5. Conclusion

This paper proposes an aircraft object detection method named TransEffiDet in aerial images based on EfficientDet and Transformer methods. We improved the EfficientDet object detection algorithm by combining it with the Transformer which models the long-term correlation of the features. The mAP of the proposed TransEffiDet in aerial images can reach 86.6%, which outperforms the EfficientDet by 5.8%. The experimental results show that TransEffiDet has good robustness and is more suitable for aircraft detection and classification tasks in military field than the compared methods. Additionally, we have established a public aerial dataset for aircraft detection and classification, which will be released along with this paper. In the future, we will explore the application of this method to target detection in military field, which may require faster detection speed.

In this study, our proposed method is employed to detect aircrafts, but so far it is not easy for this method to accurately detect fighter jets, bombers, and passenger aircraft. A possible explanation is that the shape features of these aircrafts are not very obvious, so the feature extraction network cannot effectively extract these features for classification. In our future work, we try to use swim-transformer to extract rich global and local features to improve detection accuracy.

Data Availability

Data are available at <https://pan.baidu.com/s/11UORs4eaKKPZNscIrtbISg>; Extraction code: data.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Yanfeng Wang contributed to methodology, writing the original draft preparation, and conceptualization. **Tao Wang** provided software and performed data acquisition and investigation. **Runmin Liu** contributed to model guidance. **Xin Zhou** performed validation and project administration. **Meigen Huang** performed formal analysis. **Tian Jing** performed visualization. **Mu Lin** contributed to reviewed and edited the article. **Hua He** performed data curation. **Weiping Wang** contributed to model guidance. **Yifan Zhu** contributed to model guidance. Yanfeng Wang and Tao Wang contributed equally to this work.

Acknowledgments

The authors are grateful to all members of College of Systems Engineering of National University of Defense Technology for their advice and assistance in the course of this research. This study was supported without any funding.

References

- [1] G. Wang, G. Liu, N. Yu, and X. Liu, "An intelligent algorithm for infrared object recognition," in *Proceedings of the 2016 IEEE International Conference on Information and Automation (ICIA)*, pp. 1818–1821, Ningbo, China, August 2016.
- [2] 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. 1137–1149, 2017.
- [3] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, IEEE, Honolulu, HI, USA, July 2017.
 - [4] H. Law and J. Deng, “Cornersnet: detecting objects as paired keypoints,” in *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, pp. 734–750, Springer, Munich, Germany, September 2018.
 - [5] A. Rastogi, R. Singh, R. Sharma, and S. D. Kalony, “The survey of digital image analysis with artificial intelligence-DCNN technique,” in *Proceedings of the 2020 Ninth International Conference System Modeling and Advancement in Research Trends (SMART)*, pp. 209–211, Moradabad, India, December 2020.
 - [6] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, “Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325–3337, 2015.
 - [7] F. Zhang, B. Du, and L. Zhang, “Saliency-guided unsupervised feature learning for scene classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, 2015.
 - [8] L. Cao, J. Luo, L. Feng, and T. S. Huang, “Heterogeneous feature machines for visual recognition,” in *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision*, pp. 1095–1102, Kyoto, Japan, October 2009.
 - [9] B. Hariharan, J. Malik, and D. Ramanan, “Discriminative decorrelation for clustering and classification,” *Computer Vision - ECCV 2012*, vol. 7575, pp. 459–472, 2012.
 - [10] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, “State-of-the-art in artificial neural network applications: a survey,” *Heliyon*, vol. 4, no. 11, Article ID e00938, 2018.
 - [11] X. Chen, S. Xiang, C. Liu, and C. H. Pan, “Aircraft detection by deep belief nets,” in *Proceedings of the 2013 2nd IAPR Asian Conference on Pattern Recognition*, pp. 54–58, Naha, Japan, November 2013.
 - [12] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, IEEE, Las Vegas, NV, USA, June 2016.
 - [13] S. N. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, “Aggregated residual transformations for deep neural networks,” in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1492–1500, IEEE, Honolulu, HI, USA, July 2017.
 - [14] C. Szegedy, W. Liu, Y. Q. Jia et al., “Going deeper with convolutions,” in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, IEEE, Boston, MA, USA, June 2015.
 - [15] C. C. Zhu, R. Tao, K. Luu, and M. Savvides, “Seeing small faces from robust Anchor0s perspective,” in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5127–5136, IEEE, Salt Lake City, UT, USA, June 2018.
 - [16] L. Xie, Y. Liu, L. Jin, and Z. Xie, “DeRPN: taking a further step toward more general object detection,” vol. 33, pp. 9046–9053, in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*, vol. 33, AAAI, Honolulu, Hawaii, USA, February 2019.
 - [17] Z. Tian, C. H. Shen, H. Chen, and T. He, “FCOS: fully convolutional one-stage object detection,” in *Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9627–9636, IEEE, Long Beach, CA, USA, November 2019.
 - [18] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, “Soft-NMS—improving object detection with one line of code,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision (CVPR)*, pp. 5561–5569, IEEE, Venice, Italy, October 2017.
 - [19] Y. H. He, X. Y. Zhang, M. Savvides, and K. Kitani, “Softer-NMS: rethinking bounding box regression for accurate object detection,” 2018, <https://arxiv.org/abs/1809.08545>.
 - [20] Y. Liu, L. Q. Liu, H. Rezatofighi, T.-T. Do, Q. Shi, and I. Reid, “Learning pairwise relationship for multi-object detection in crowded scenes,” 2019, <https://arxiv.org/abs/1901.03796>.
 - [21] A. Shrivastava, A. Gupta, and R. Girshick, “Training region-based object detectors with online hard example mining,” in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 761–769, IEEE, Las Vegas, NV, USA, June 2016.
 - [22] H. Yu, Z. N. Zhang, Z. Qin et al., “Loss Rank Mining: a general hard example mining method for real-time detectors,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Springer, Rio de Janeiro, Brazil, July 2018.
 - [23] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: scalable and efficient object detection,” in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Article ID 10778, Seattle, WA, USA, June 2020.
 - [24] A. Vaswani, N. Shazeer, N. Parmar et al., “Attention is all you need,” 2017, <https://arxiv.org/abs/1706.03762>.
 - [25] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.
 - [26] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768, Salt Lake City, UT, USA, June 2018.
 - [27] Y. Xie, J. Zhang, C. Shen, and Y. Xia, “CoTr: efficiently bridging CNN and transformer for 3D medical image segmentation,” 2021, <https://arxiv.org/abs/2103.03024>.

Research Article

Research on the Transformation from Financial Accounting to Management Accounting Based on Drools Rule Engine

Rui Liu,¹ Yuqin Wang ,² and Jing Zou¹

¹School of Accounting, Wuhan College, Wuhan 430212, Hubei, China

²The Centre of Finance Research, Wuhan University, Wuhan 430212, Hubei, China

Correspondence should be addressed to Yuqin Wang; 8659@whxy.edu.cn

Received 22 February 2022; Revised 11 March 2022; Accepted 18 March 2022; Published 20 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Rui 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 development of Internet economy and the advent of artificial intelligence era, the transformation from financial accounting to management accounting has become an inevitable trend of financial management among companies. In this paper, the significance of the transformation from financial accounting to management accounting is expounded under the background of artificial intelligence, and the current situation, problems and reasons of the transition are analyzed. In addition, the development of self-management accounting system based on Drools rule engine is put forward, including rule management, rule engine and so on. The core processes such as pattern matching and auto-billing are analyzed, and then the core subsystems such as data receiving and processing, accounting rule engine are designed in detail. Finally, the technological advancement and structural stability of this system are discussed, which can provide reference for promoting the transformation from financial accounting to management accounting under artificial intelligence.

1. Introduction

The deep combination of artificial intelligence and financial management has promoted the development of financial management among enterprises towards automation and intelligence. Under this trend, traditional financial methods have been difficult to apply to modern financial management. Financial management must undergo subversive changes to fully adapt to the development of artificial intelligence [1]. Under the background of artificial intelligence, financial management needs to actively use information technology, integrate financial data, and provide guarantee for various business management decisions of enterprises, so as to avoid the lag of traditional financial management. Considering the current situation, with the help of a new generation of scientific and technological revolution and industrial transformation, artificial intelligence has achieved good application in many professional fields. Combined with relevant survey [2], in the next 10 years, the accounting industry will probably be the professional field most affected

by the application of artificial intelligence. From the perspective of accounting management, the concept of traditional financial accounting management has been weakened by the management accounting, and the repetitive labors in financial work can be reduced by combining the artificial intelligence, which ensures the improvement of financial quality [3–5]. It is not difficult to see that with the support of artificial intelligence technology, the transformation and development from financial accounting to management accounting will be further promoted.

With the rapid development of information technology and the global popularity of artificial intelligence, the management accounting system has gone through a process from manual accounting to computerized accounting to informatization. After entering the twenty-first century, under the impact of the rapid development of artificial intelligence and e-commerce, the emergence and development of accounting information system has been regarded as the inevitable trend [6]. Network accounting information system is based on Internet technology, which reflects the

accounting and supervision of financial resources within the whole enterprise, which realizes comprehensive, timely and dynamic accounting supervision, prediction and management of the whole enterprise [7]. By providing enterprises with accounting methods and financial management modes under the network environment, accounting management is gradually becoming standardized, and the supervision and accounting functions of accounting are fully exerted. Therefore, Promoting the informatization, networking and automation of management accounting system in the era of artificial intelligence is helpful to the transformation from traditional financial accounting to management accounting, which has become the new direction of development in accounting system.

2. Analysis of Financial Accounting and Management Accounting under Artificial Intelligence

2.1. Definition of Financial Accounting and Management Accounting

2.1.1. Financial Accounting. The so-called financial accounting refers to the general term of economic management activities for investors and creditors who have economic interested relationship with the development of enterprises outside the enterprise and information about financial status and profitability provided by relevant government departments [8, 9]. In the development of modern enterprise, financial accounting is one of the most important management items. Generally speaking, through a series of financial accounting procedures, financial management can provide useful decision-making information for managers, and can promote enterprises to run efficiently in the process of serving the market economy.

2.1.2. Management Accounting. The so-called management accounting refers to a branch of accounting [10, 11] which is separated from traditional accounting, integrates accounting and financial management, and focuses on providing economic forecast, investment decision-making, management improvement and economic benefit service for enterprise managers. Theoretically, the concept of management accounting enriches the related research on accounting and management in academic circles. From the practical function point of view, where management accounting mainly has the following functions: first, it can put forward objective and practical suggestions on the development of enterprises based on data from the financial point of view, so that the development of enterprises can be more scientific and reasonable; Second, it is able to grasp the capital status of an enterprise in the process of development in time, which makes the context of the enterprise's capital clear; Third, to a certain extent, it can reduce the cost in the process of business operation and expand the profits of enterprises. At the same time, it can also make full preparations for further investment to a certain extent, and ensure the steady progress of enterprises.

2.2. Differences between Financial Accounting and Management Accounting. In recent years, with the rapid development of artificial intelligence technology, management accounting system has been successfully integrated into various professional fields for practical application, and achieved remarkable results. For financial management, with the support of decision-making of artificial intelligence technology, the differences between financial accounting and management accounting are becoming more and more obvious [12].

On the one hand, under the background of "artificial intelligence," the financial accounting work of enterprises is more focused on serving the management of enterprises. Generally speaking, in the process of business management, enterprises will involve a large number of capital transactions and financial data. If only the previous financial accounting methods are used for management, in terms of financial integration, it is easy to make mistakes. By using informatization management, enterprise financial data information can be integrated and given feedback to managers in the form of financial statements, which can effectively reduce errors of management [13]. At the same time, in the management of economic activities, the financial department of an enterprise can complete the decision-making analysis of financial data by means of management accounting, in which managers can make forward-looking deployment for the direction in a certain period according to the information data fed back by the financial report [14].

On the other hand, the methods of financial management become more flexible and diverse. Compared with the traditional financial management model, the financial accounting of modern enterprises pays more attention to reflecting the diversified characteristics, requiring financial managers to flexibly adjust the selected management methods according to the actual operation of enterprises. The essential purpose of financial accounting and management accounting is to realize high-quality process of financial management, but there are still some differences in responsibility [15]. For example, financial accounting reports are more strict in preparation than management accounting, and legal responsibility is also relatively strict. In other words, Accounting is more obvious in terms of financial constraint, while management accounting can actively combine artificial intelligence and other technical contents to realize systematic treatment of financial management, and timely adjust financial management according to the financial operation, with more obvious decision-making characteristics.

2.3. Problems in the Current Management Accounting System.

In order to promote the transformation from financial accounting to management accounting, first, the current management accounting system must be optimized. At present, the management accounting system used by most companies is a software system under computerized accounting. The whole process of the traditional management accounting system is: transaction, voucher, accounting, reconciliation, trial calculation, financial statements, most of which need manual participation, thus exposing some urgent problems [16–19].

2.3.1. Inefficient Manual Accounting. The basic function of accounting is calculation and supervision, so in the daily business of the company, every transaction involving capital changes needs accounting. From the traditional cycle process, it can be seen that from the transaction to the voucher accounting, reconciliation, trial calculation, and the issuance of financial statements, manual intervention is needed because it is mainly designed for traditional invoicing enterprises, while the turnover rate of inventory funds in traditional enterprises is slow, so it has no need for real-time auto-billing. However, with the in-depth development of artificial intelligence, the Internet-based financial enterprise does not have the concept of inventory, but provides online payment or financial services, with rapid transactions. Therefore, their capital turnover rate is fast, and manual accounting is obviously inefficient and cannot meet the actual business needs of the company's financial management.

2.3.2. Low Accuracy Rate of Data. The current management accounting system cannot reflect the detailed changes of the whole business or individual business of the company. The data recorded by accountants are generally the data after statistics of various business systems, due to the changes of transaction data in real time. Moreover, the financial personnel can only record the total statistics of each business system in advance, so the changes of individual transactions cannot be reflected in the current financial system, and all the business systems are connected to the management accounting system.

2.3.3. Nonnetworking Based on Single Machine. At present, the company's management accounting system is based on the client stand-alone mode, and the financial personnel need to install the client first. Without the client's personal PC or mobile devices such as mobile phones and IPAD, the financial personnel cannot count the financial statements, and managers cannot analyze and make decisions according to the financial statements. In addition, the management accounting system under computerized accounting is not based on the Internet. Under the circumstance that enterprises are constantly building various information systems, it will be out of touch with the major information systems of enterprises. For example, the data format of each business information system will be inconsistent with the format required by the management accounting system. Therefore, the computerized accounting software based on single machine can easily lead to the formation of information islands.

3. Design and Implementation of Management Accounting System Based on Drools

Management accounting system is undoubtedly very important during the normal operation of management activity. It can help managers to know the profit and loss situation of enterprises in time, so as to formulate reasonable strategies. From the above analysis, it can be seen that in order to promote the transformation from traditional

financial accounting to management accounting, first of all, it is necessary to solve the problems that the current accounting management system needs to manually record the accounting vouchers of each transaction, as well as it cannot provide business rules to automatically match transaction orders with large amounts of data, and cannot automatically make trial balance [20]. Therefore, a management accounting system based on Drools rule engine is designed in this paper, so as to better realize the faster and better transformation of financial accounting.

3.1. Introduction of Drools Rule Engine Technology

3.1.1. Basic Concepts of Rule Engine. Drools is a popular rule engine component at present, while its basic components include business rules and rule engines are. Here, the definition, composition and workflow of rule engines are briefly introduced.

Definition of business rules: in essence, business rules can also be understood as a set of conditions and operations, which include conditions and actions. When a business fact meet certain conditions, certain actions will be executed, which is similar to the meaning expressed by conditional statements if...then... in program [21]. The rules files of this system are managed by Drools Guvnor.

Definition of rule engine: rule engine is the environment where rules are to be executed. Its main functions contain describing rules, compiling rules, rule execution and resolution of rules conflict, which is an independent component that can be embedded into programs [22]. At present, the rule engines of comparative processes are: Drools of Jboss Company of the United States, VisualRules of Qizheng Company of China and ILOG of IBM abroad.

Composition and workflow of rule engine: "How to make computers think like humans?" This is a problem in the field of Artificial Intelligence, which includes [23, 24]: Neural Networks, Genetic Algorithms, Decision Trees, Frame Systems and Expert Systems. Rule engine is one of Expert Systems, which is also known as knowledge-based system. It is a system based on rule base, transferring facts and drawing conclusions. The structure of the rule engine is shown in Figure 1:

It can be seen from the figure that the rule engine consists of three parts [25]: Production Memo, Working Memory, and inference Engine, while the inference engine consists of Pattern Macher, Agenda, and is also known as Conflict Resolution. When one or more facts are met by multiple rules, conflicts will occur. At this time, the Agenda is responsible for coordinating the execution sequence of Activations. Each activation consists of a rule and a fact or multiple facts. The Execution Engine is responsible for executing the rules selected by the Agenda. The actions triggered by the rules may produce new facts, which will be readded to the Working Memoir fact base.

3.1.2. Matching Algorithm of Rete Pattern. Drools is an open source rule engine that coded in Java, and adopts Rete algorithm to calculate rules [26]. The implementation of

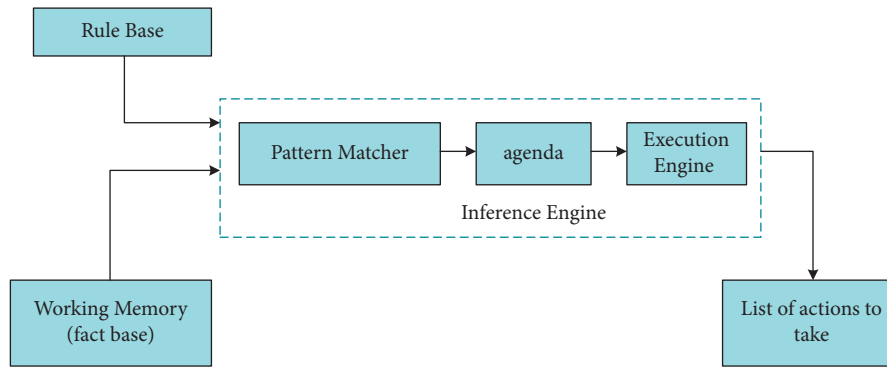


FIGURE 1: Structure of rule engine.

DroolsRete is called ReteOO which means that Drools has enhanced and optimized the implementation of Rete algorithm of object-oriented system. Drools allows users to express rules of business logic declaratively. Rules in non-xml native language can be coded which is easy to learn and understand [27] and Java can be embed directly into the rules file. In addition, Drools has other advantages:

- (1) Supported by an active community
- (2) Convenient to use
- (3) Fast execution
- (4) More and more popular among Java developers
- (5) Comply with Java rule engine API(JSR94)
- (6) Free

Drools uses a rule-based method to implement an expert system. It is more accurately classified as a production rule system [28]. The production rule system is Turing complete, focuses on knowledge representation, and expresses propositions and first-order logic in a concise and nonfuzzy way. The brain of the production rule system is an inference engine Figure 2. It can be extended to numerous rules and fact reasoning engines to match facts and data according to production rules (also known as production rules or fair rules) to infer the conclusions that lead to actions.

3.2. *Design of Systematic Architecture.* By analyzing and explaining the Drools rule engine based management accounting system, the network architecture, system software architecture and other aspects are analyzes and compared with similar systematic architectures which illustrates the technical advantages of implement based on Drools rule engine.

3.2.1. *Systematic Network Architecture.* The management accounting system is a comprehensive application system with a large scale of data. The design of platform system should fully consider the needs of the actual accounting business, and on this basis, determine the overall goal of the system, so that the system has the characteristics of advanced practicality, high performance, security, scalability and excellent compatibility [29]. The network architecture of the management accounting system is shown in Figure 3.

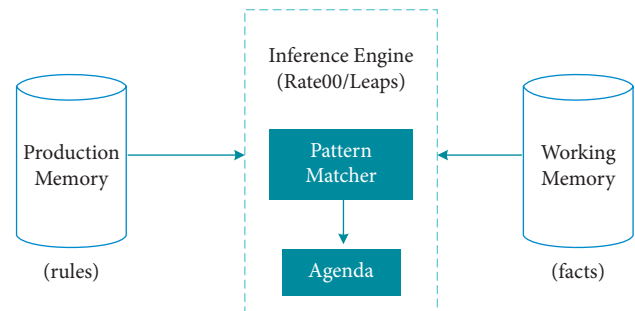


FIGURE 2: Structure of Rete rule matching algorithm.

B/S (Bowser/Server) architecture is adopted where the client does not need to be installed, and users only need to pass the comparison on personal computers, laptops and IPAD, then Internet browsers such as Rome and Safari can realize the interaction with the management accounting system [30]. First, users initiate online transactions to the company's business system through HTTP/TCP requests. Besides recording data in the databases of their respective systems, at the same time, the ActiveMQ message is sent to ApacheActiveMQ server through TCP/IP protocol, the data receiving and processing subsystem reads the message of ActiveMQ server through TCP/IP protocol, and then calls the interface of rule engine subsystem through Netty communication framework of TCP/UDP protocol. Afterwards, the data receiving and processing subsystem converts the transaction message into accounting voucher and then the accounting voucher is sent to the rule engine subsystem for processing. After receiving the accounting voucher object, the rule engine subsystem converts the accounting voucher into the corresponding accounting record according to the predefined accounting rules.

The other way is completed by the batch processing task of the data receiving and processing subsystem, because each business system needs to access the management accounting system step by step for a period of time, and at the same time, it is necessary to keep an account of the historical transaction data of each business system, so the accounting batch processing is needed to process the batch processing task through ORACLE's DBlink technology, so as to let the

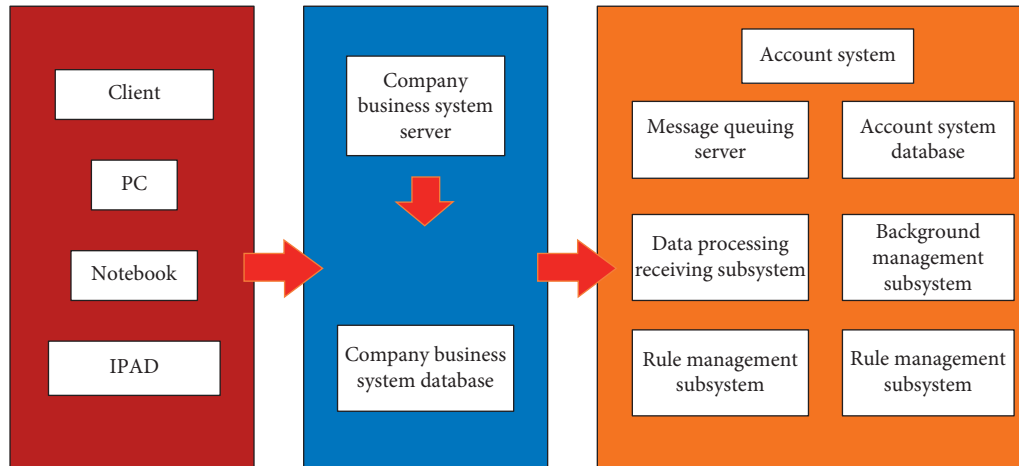


FIGURE 3: Systematic network architecture.

database of the management accounting system be directly connected to the database reserve of each business system.

3.2.2. Systematic Software Architecture. In the design process, the mainstream three-tier architecture system is adopted, and the whole architecture system is divided into data access layer, business logic layer and presentation layer, which makes the software design convenient for modularization and standardization, and is conducive to the development, maintenance and future expansion and upgrading of the system. The software architecture of management accounting system is shown in Figure 4.

- (1) Performance layer: it is provided for external end users, and is used by internal staff through the entrance of interoperation with major business systems. For example, it is provided for financial personnel with the back-office management subsystem of management accounting system. After the back-office management personnel log in to the subsystem, audit accounts, bank gateway, bank rate, query and report, etc will be set up. This layer is all displayed in the way of web browser, and users can use the most common browsers without installing any client program, which brings convenience to them.
- (2) Business logic layer: the main business logic of the management accounting system can be realized at this layer, which is mainly composed of four main subsystems: rule management subsystem, data receiving and processing subsystem, rule engine subsystem and accounting background management subsystem. At the same time, the short message system and mail system is externally connected. After the system error or some business is completed, For example, the data receiving and processing subsystem generates all the transaction data into accounting vouchers and accounting records, short messages and emails will be used to inform relevant developers and financial personnel. In addition, relevant data in Memocached cache system can be

selected in order to optimize system performance, and ApacheActiveMq message queue system is used to receive transaction data messages of major business systems.

- (3) Data access layer: the business logic layer is connected to the data access layer through JDBC. This system uses the relational database Oracle10g, and manages the basic data of the accounting system, such as basic data used in accounting rules, accounting subjects, bank gateway, bank rates, etc. Because it involves confidential data of company transactions, the production data is not allowed to be accessed by the test environment. At the same time, The release of DDL and DML in the database only allows special DBA to query and export related statements such as balance sheet, income statement, cash flow statement, etc., and is restricted to financial personnel and company management. The company's business system and management accounting system take Oracle10g, where DBlink technology can be used to realize high-speed access between business system and accounting management system database.

The business module has been independent, and loose coupling is realized at the code level. Therefore, when adding or modifying accounting rules, there is no need to recompile the code and release a new system version, just edit the xml or drl rule file online, and all subsystems can adopt the Saturn framework developed by the company. The architecture of framework is clear and easy to maintain and expand, which has superior performance. Saturn framework is shown in Figure 5.

3.3. Systematic Implementation

3.3.1. Implementation of Rule Management Subsystem. The main function of DroolsGuvnor rule management subsystem is to manage the rule files used by the rule engine subsystem, and to create and edit the rule files, check the grammar correctness, test and publish the rules.

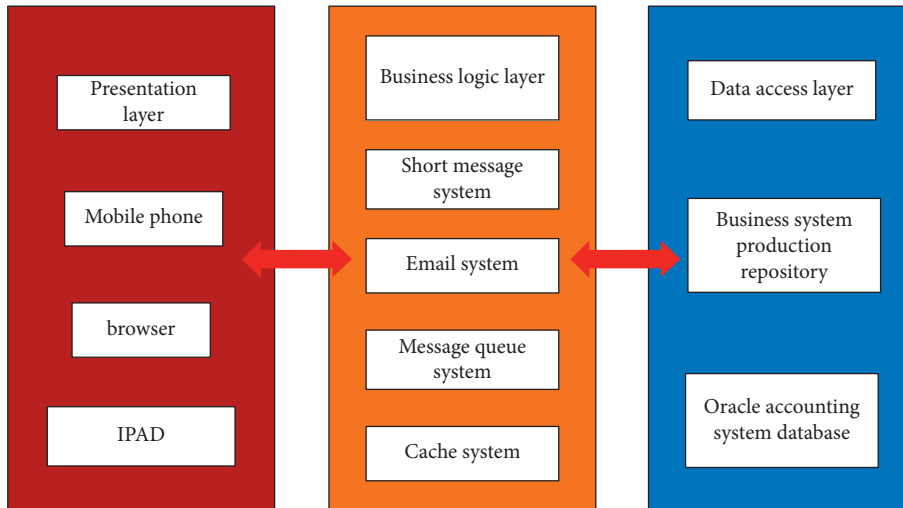


FIGURE 4: System software architecture.

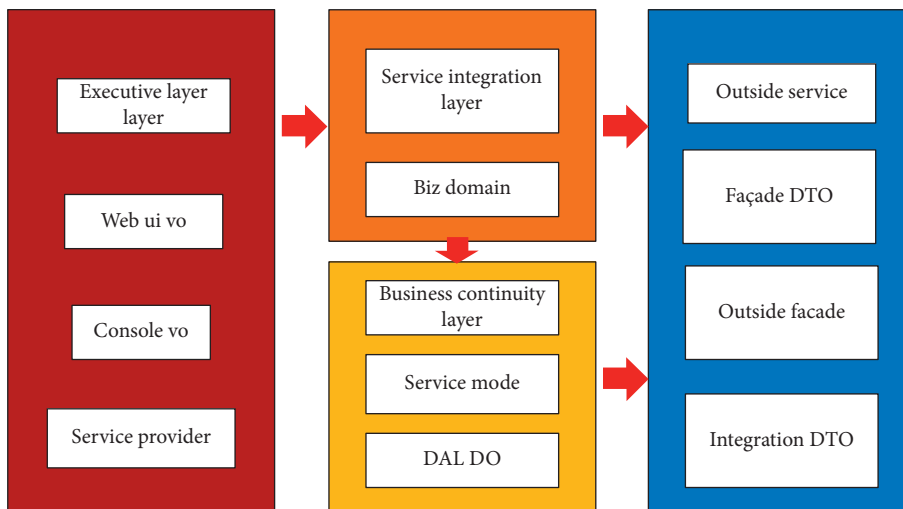


FIGURE 5: Saturn framework.

Add repository: after the new repository is added, the role of the repository is to store rule file items. Click Authoring Administ Rational, then click Newreposito to add a new repository, and then fill it down step by step. Enter the name of the repository and the organizational unit to which it belongs. The relative location of the repository is git://localhost: 9418/AccountRepository. The version of the project file is controlled by git tool.

Add items: after the repository is created, create the Project of Drools rule file, and enter the description such as Project Name, Project Description, GroupID, ArtifactID, VersionID, etc. to create the rule file management project of the management accounting system.

Create a new data model:

after the project is created, it needs to create a data model, that is, the Fact model in the rule engine, which is created by DataModeler under the tools menu. Fill in the name of the model and the name of the package where it is located, then add the field name and field type for the

new object type, and then click “Save” to save the object model.

Create a rule file: after the relevant data model is created, continue to create the rule file, and send it to create the rule file through the DRLfile in theNewItem, or through the GuideRule or GuideRuleTemplate option.

3.3.2. Implementation of Rule Engine Subsystem. This system uses Maven technology to manage the dependency and construction. The project dependency on DroolsExpert [31] is configured in the root pom xml of the project as follows:

- (i) <dependencies>□□
- (ii) <dependency>□□
- (iii) <groupid>org.drools</groupid>□□
- (iv) <artifactId>drools-core</artifactId>□□
- (v) <version>5.5.0.FinaK</version>□□
- (vi) </dependency>□□

- (vii) <dependency>□□
- (viii) <groupId>org.drools</groupId>□□
- (ix) <artifactId>drools-compiler</artifactId>□□
- (x) <version>5.5.0.Final</version>□□
- (xi) </dependency>□□
- (xii) <dependency>□□
- (xiii) <groupId>com.thoughtworks.xstream</groupId>□□
- (xiv) <artifactId>xstream</artifactId>□□
- (xv) <version>1.3.1</version>□□
- </dependency>□□
- </dependencies>□□

3.3.3. Implementation of Data Receiving and Processing Subsystem. The main function of the data receiving and processing subsystem is to receive the order transaction data of major business systems, convert the transaction data into accounting vouchers, and then call the rule engine subsystem to record the accounting vouchers into the management accounting system database. At the same time, there are two ways for data receiving and processing subsystem to receive data: single transaction and batch transaction.

The first way is to provide ApacheActiveMQ message queue address. Every transaction with cash in online business system will send a message to ApacheActiveX queue in real time. Then, the consumer asynchronously extracts the message, after extracting the transaction order data message, the data in JSON string format are converted into the accounting voucher object, and then the relevant interface of the rule engine subsystem is called to transfer the accounting voucher object to the rule engine subsystem.

The second way is to process the historical transaction data of major business systems. Firstly, task scheduling plan must be set in task scheduling. The batch-running program that runs once every time adopts javax.jms.MessageListener technology. After a scheduled task is triggered at a certain point in time, the MessageListener method of MessageListener will receive the information of this scheduled task. Therefore, the batch-running program that triggers the timed batch-running task directly goes to the database backup of the business system by adopting DBLink technology between the management accounting system and the business system database. The rule engine subsystem is responsible for converting each accounting voucher into accounting entries, and recording the accounting vouchers and accounting entries into the database in turn.

4. Transformation of Financial Accounting under the Management Accounting System Based on Drools

With the rapid development of the company's business, relying on the accounting method of financial accounting can no longer meet the needs of the company's current

accounting management. In order to achieve real-time accounting, Drools rule engine is introduced in this system so that the transaction records can automatically match accounting rules after passing through the rule engine subsystem, which has the following advantages to realize the transformation of accounting mode:

4.1. Advanced Technology. The efficiency of automatic accounting is high, and the probability of occurring error is very small. All that needed is to configure the correct entry generation rules in the rule management subsystem. The accounting entry generation rules set up double-entry accounting entries according to the information of transaction data such as transaction platform number, transaction type, payment gateway number, merchant number, etc., which not only realizes automation in accounting, but also in many aspects, such as the generation of trial balance daily-cut accounting statements that is conducive to the transformation from financial accounting to management accounting.

4.2. Reasonable Structure. This system adopts B/S(Browser/Server) architecture, without installing any client. Users only need a terminal that can be connected to the network to access the background management system of the rule management subsystem to manage the accounting system, set accounting rules and manage relevant basic data of the accounting system, or to view and download related financial statements. Compared with the financial accounting system, there is no need to install SqlServer and other related third-party components before running. The current management accounting system provides structural advantages for the transformation of accounting modes.

5. Conclusion

In order to conform to the development of the artificial intelligence, the financial management of modern enterprises needs to be actively based on the background of artificial intelligence management, and solve the problem of lag existing in the current financial management. In this paper, the strategy of transformation from financial accounting to management accounting is analyzed. By designing a management accounting system based on Drools rule engine to realize the transformation from financial accounting to management accounting under the background of artificial intelligence, the architecture of the management accounting system is specifically analyzed, including: the network architecture and software architecture. Then the implementation of the core subsystem related to Drools technology application is studied. The system can realize automation in many aspects such as the generation of trial balance and daily accounting statements, thus promoting the transformation from financial accounting to management accounting under artificial intelligence, and advancing the long-term sustainable development of enterprises and social economy.

Data Availability

All data generated or analysed during this study are included in this published article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was funded by the Construction Plan of Scientific Research and Innovation Platform of Wuhan College (no. KYP202001) and supported by the Research and Innovation Team of Wuhan College (no. KYT201903). The authors would like to thank Research Center for Value Evaluation and Creation of Private Enterprises and Audit Value Innovation Research Team in the New Era which contribute to this study greatly.

References

- [1] C. Xian, "Research on the transformation from financial accounting to management accounting in the era of artificial intelligence," *Accounting Learning*, no. 16, 2021.
- [2] W. Wang, "Transformation from financial accounting to management accounting in the era of artificial intelligence," *Guide to public investment*, no. 7, 2021.
- [3] Q.in Zhu, "Discussion on the transformation from financial accounting to management accounting in the era of artificial intelligence," *Taxation*, vol. 15, no. 7, 2021.
- [4] M. Zhang, "Research on the countermeasures of the transformation from financial accounting to management accounting in the era of artificial intelligence," *Enterprise Reform and Management*, no. 4, 2021.
- [5] C. Yun, "Probe into the transformation from financial accounting to management accounting in the era of artificial intelligence," *Business News*, no. 34, 2020.
- [6] Na Zhang, "Probe into the transformation from financial accounting to management accounting under the background of big data," *Modernization of shopping malls*, no. 3, pp. 152-153, 2019.
- [7] H. Yao, "Exploring the path from financial accounting to management accounting," *Accounting Studies*, no. 25, p. 135+137, 2019.
- [8] Y. Zhou, "Analysis of the transformation from enterprise financial accounting to management accounting under the background of big data," *Tax collection*, vol. 14, no. 4, pp. 108-110, 2020.
- [9] J. Cai, "Research on the transformation strategy from financial accounting to management accounting under the background of big data," *Tax collection*, vol. 14, no. 4, p. 138, 2020.
- [10] H. Dong, "Thinking and analysis of the transformation from financial accounting to management accounting in the era of artificial intelligence," *Finance and Economics*, no. 28, pp. 131-133, 2020.
- [11] M. Tong, "On the transformation from financial accounting to management accounting under the new situation," *China Management Informatization*, vol. 21, no. 22, pp. 14-15, 2018.
- [12] L. Jing, "Using the integration of financial accounting and management accounting to strengthen enterprise management," *Small and medium-sized enterprise management and science and technology (Shangxun Journal)*, no. 8, 2021.
- [13] B.'s Li, "Research on the strategy of transforming financial accounting into management accounting under the background of Great wisdom moves cloud," *Journal of Jiamusi Vocational College*, no. 7, 2021.
- [14] K. Wang, "Thoughts on the transformation from financial accounting to management accounting under the background of big data," *Accounting Studies*, vol. 6, no. 18, 2021.
- [15] J. Zhao, "On the integration trend and strategy of enterprise management accounting and financial accounting," *Accounting Studies*, no. 18, 2021.
- [16] Y. Wu, "Probe into the application of management accounting in financial management of commercial banks," *China Sankei*, no. 12, 2021.
- [17] J. Integration, "Of enterprise management accounting and financial accounting under the new situation," *Journal of Science and Technology Economy*, no. 18, 2021.
- [18] Drools, "The,Business,Logic,Integration,Platform,[Eb/OI]," 2001, <https://www.drools.org/>.
- [19] A. Gao, *Drool Smart Rule Engine Development Tutorial*, Shanghai Ruidao Information Technology Co., Ltd., Shanghai, 2009.
- [20] L. Charles, "Forgy. Rete: A Fast Alorithm for the Many Pattern/Many Object Pattern Match Problem," *Artificial Intelligence*, Department of Computer Science, Carnegie-Mellon University, no. 19, , pp. 17-37, Pittsburgh, PA15213,U.S.A, 1982.
- [21] JBoss Drools Team, *Drools Guvnor User Guide*, Red Hat,Co. Ltd., North Carolina, 2012.
- [22] B. Paul, *JBoss Drools Business Rules*, pp. 12-18, Packt Publishing, Birmingham, 2009.
- [23] JBoss Drools Team, *Drools Documentation*, Red Hat Co.,Ltd., North Carolina, 2013.
- [24] M. Bali, *Drools JBoss Rules 5.X Developer ' s Guide*, pp. 90-200, Packt Publishing, Birmingham, 2013.
- [25] T. Davenport, *Process Innovation: Reengineering Work through Information Technology*, p. 6, Harvard Business Review Press, New York, 1993.
- [26] J. Rovtar, *Implementation of business rules using business rules management system*, 2017.
- [27] P. Browne, *JBoss Drools Business Rules*, Pakt Publishing, Birmingham, 2009.
- [28] B. V. Halle, *Business Rule Revolution*, Google Books, 2011.
- [29] T. Morgan, *Business Wesley, Rules and Information Systems*, Addison, Boston, 2002.
- [30] I. Graham, *Business Rules Management and Service Oriented Architecture. A Pattern Language*, Wiley, Hoboken, New Jersey, 2006.
- [31] A. Poniszewska-Maranda, D. Kaczmarek, N. Kryvinska, and F. Xhafa, "Studying usability of AI in the IoT systems/paradigm through embedding NN techniques into mobile smart service system," *Journal of Computers*, vol. 101, no. 11, pp. 1661-1685.

Research Article

Research on Intelligent Target Tracking Algorithm Based on MDNet under Artificial Intelligence

Yu Wang 

Chengyi University College, Jimei University, Information Engineering School, Xiamen 361000, China

Correspondence should be addressed to Yu Wang; ruoque1001@jmu.edu.cn

Received 23 February 2022; Revised 17 March 2022; Accepted 21 March 2022; Published 19 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Yu 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.

Target tracking is an important subject in computer vision technology, which has developed rapidly in recent ten years, and its application have become wider and wider. In this process, it has transferred from a simple experimental tracking environment to a complex real scene where more challenges need to be solved. The rapid development of deep learning has promoted the research progress of digital vision. Target tracking technology is an important foundation of digital vision research, which makes it develop from academia to industry. In this paper, a method of target tracking using MDNet is introduced. Starting with the attention mechanism, two attention mechanisms are added to extract and integrate the better features. Case partitioning is used to reduce the investment of tracking module and minimize the network size during tracking, and its result can be prevented from getting worse. Finally, the experiment is analyzed in detail.

1. Introduction

Target tracking is an important subject in computer vision technology, which is widely used in human-computer interaction, vehicle navigation, automatic monitoring and so on. The general tracking problem is considered as a task of online learning, that is, to estimate the motion trajectory of subsequent targets according to the information of the first object in a video. Visualization of target tracking is a challenging task, and it is difficult to predict a large number of rapid changes well.

Generally speaking, target tracking refers to inferring the subsequent unknown target information according to the known target state information in the video sequence, and obtaining the most likely location of the target. The ability to use deep learning techniques to improve computer vision problems [1], especially in image classification [2], target detection [3], target tracking [4], semantic segmentation [5], etc.

2. Target Tracking Method Based on Deep Learning

Three kinds of tracking algorithms based on deep learning: depth-based tracking algorithm based on association filtering,

detection-based tracking algorithm and template-based tracking algorithm [6].

Automatic tracking algorithm based on deep learning combined with correlation filtering has the same model training and prediction method as the tracking algorithm based on correlation filtering, except that the feature used is the depth feature extracted by convolutional neural network [7]. In addition, convolution features can be combined with traditional features to improve the expressive ability of features. Convolution feature is a feature of hierarchical representation, that is, deep convolution features encode more semantic features, while shallow convolution features contain fine details. By utilizing semantic class data from the profound layer and segregation data from the shallow layer, Ma prepared a straight relationship channel on a bunch of fixed convolution layers, and the coarse-to-fine technique is utilized to anticipate the area of the objective [8]. By extending the correlation filter of spatial regularization, Weber and Kanarachos et al. put forward DeepSRDCF tracker, and furthermore explored the impact of convolutional highlights on target tracking [9].

In the detection-based target tracking method, the target detection network is used to detect the target, and then, the

matching calculation is utilized to relate the distinguished target frames. Ren et al. proposed a multi-target tracking method that is suitable for crowded situations. This method employ density map to predict the number of targets, detect their positions and track them. Through experiments in cells, fishes and people, it has achieved excellent tracking results. When tracking this kind of algorithm, it needs to go through two steps of detection and matching, and the matching algorithm is complicated, which makes high time complexity of this algorithm. In addition, it has strong adaptability to the changes of appearance and tracking environment [10, 11].

Target tracking technology based on template matching, which can make network evaluate the similarity of sample map and retrieval map. The most representative network architecture is Twin Network Structure [12]. Taking SiamFC as an example, the structure of twin network is introduced. The algorithm uses two branches of twin network to extract the features of sample image and image search. Then, the obtained features are applied for convolutional operation, and the response diagram is output. However, it is difficult for twin trackers such as SiamFC to distinguish similar objects, even if the objects have obvious chromatic aberration. This is because SiamFC is insensitive to the underlying characteristics [13]. By using convolutional network, the tracker can not only obtain semantic information from high-level convolutional features, but also the local features such as texture can be acquired from low-level convolutional features. Therefore, a common practice is to integrate the features of target image and background image learned from different convolutional layers to obtain richer hierarchical information [14]. However, such integration method for convolutional feature is implicit, Because the information of contained feature may be diluted by high-level features, and it cannot fully express the characteristics of low-level features.

3. Analysis of Target Tracking Algorithm Based on Multi-Domain Convolutional Neural Network MDNet

Through experimental analysis, it can be found that in the field of savvy transportation, the mind boggling foundation, enlightenment change, impediment and scale change will all influence the following aftereffects of moving focuses in video sequences. The difficulty of tracking targets lies in how to overcome the above-mentioned factors that cause the deterioration of tracking results, on this basis, the similarity problem is solved, and the accuracy and robustness are ensured. In a complex environment, target tracking requires both accuracy and accuracy, but also insure the real-validity. And if occlusion occurs, the tracking result will gradually deteriorate [15]. Therefore, in order to obtain a more robust tracking result, In this paper, improvement of target tracking using multi-domain convolution neural network, The principle work is to consolidate the case division technique with MDNet following calculation. By adding two consideration instruments to the calculation , it extricates better highlights and guarantees that the objective adjusts to the

appearance change and significantly further develops the following exhibition. In addition, finally, this method is compared with the existing mainstream methods, which proves its good tracking effect.

3.1. Framework of Algorithm. Opromolla et al. proposed a target algorithm with less training data and lower precision is pred in this paper [16]. In 2019, the SiamRPN++ algorithm proposed by Bo Kang et al. combined the deeper neural network with the twin network, which further promoted the development of visual tracking and solved the problem of applying the deep network to target tracking. The MDNet algorithm network consists of a fully connected layer and a domain-specific layer. Offline pretraining shallow layer networks [17]. A new convolution neural network layer based on pretraining, a new two-class layer is used to construct the network, thus improving the tracking efficiency in force.

The methods above still have great drawbacks in solving practical problems. Improved MDNet algorithm improves the robustness of target tracking, and solve the problem of target similarity. In this algorithm, based on the characteristics of convolutional neural network, large data sets are trained off-line before target tracking, so that the initial target tracking algorithm can process predefined objects. On the basis of MDNet tracker, two attribute models (spatial attribute and channel attribute) are added to reduce the influence of attribute difference on attribute learning. On this basis, a method based on multi-domain learning is proposed, which separates domain-related information from domain-related information and combines it with instance segmentation and target tracking.

Firstly, MDNet based on multidomain convolution algorithm is used to improve the tracking performance by adding two models of spatial attention and channel attention. This method is combined with MaskR-CNN's instance segmentation method. Through instance segmentation, the candidate regions of the tracked object can be reduced effectively, and the candidate regions can be obtained. This method supplements the characteristics of vehicle and adjusts the network precisely, thus improving the tracking of target. Because of using the sample segmentation method, a narrower candidate tracking region can be obtained, so a smaller network structure is adopted. This method can not only separate background from foreground object effectively, but also improve tracking accuracy effectively and eliminate the redundancy of the algorithm.

Figure 1 shows the basic framework of the tracing algorithm presented in this article. First, the video boundary is segmented using MaskR-CNN, and afterward the up-and-comer areas acquired from the divided pictures are utilized as the contribution of our superior calculation [18]. In this way, the foreground tracking target can be expressed more clearly and tracking range can be reduced. Finally, the network online is fine-tuned during training and testing where the input of the tracker network is an The RGB image size is 107×107 and contains the Conv1-Conv3 convolution layer, the fc4-fc5 full connection layer, and two attention

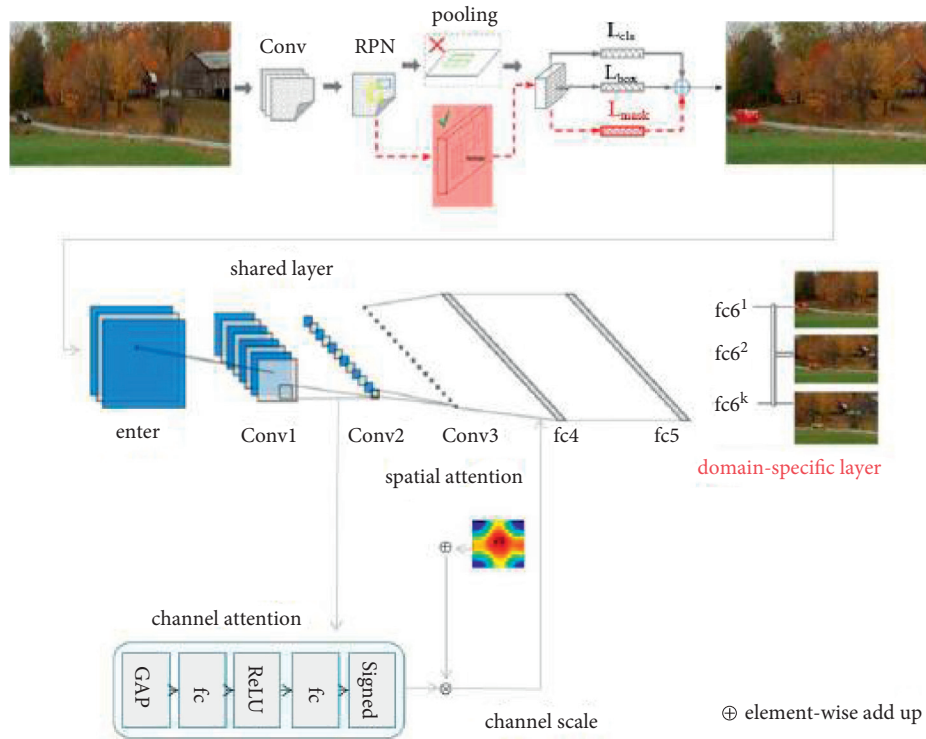


FIGURE 1: Flowchart of the algorithm in this paper.

components. Although the network size is small, it can still get a better tracking effect.

3.2. Instance Segmentation Module. Instance segmentation distinguishes different sample categories according to the representation of pixel feature. Then the targets can be found out in images of different sequence through detection, and then they are classified and regressed to distinguish different instance of the same kind.

3.2.1. Architecture of Mask R-CNN. Based on FasterR-CNN, the overall architecture of MaskR-CNN is extended [19]. As shown in Figure 2, the whole experiment of instance segmentation includes three tasks: classification, regression and segmentation. The framework consists of two parts. The first part is convolution of scanning images to extract different features. The second part is to generate bounding box and mask. When pixel-level tasks are handled in Mask R-CNN, different kinds of targets have different pertinence, which is why it is applied to the tracking algorithm in this paper.

As shown in Figure 3, firstly, the data is marked on the video frame; Secondly, the selected ROI is input to the RPN network for secondary classification. Then filter out the unwanted ROI and narrow it down. The RPN network outputs ROI coordinates in the form of $[x, y, w, h]$, and then puts them into the ROI library to get a 7×7 -size feature map. Finally, the ROI Pooling locale is handled through the ROI Align run. The ROI Align layer will get appropriate fixes, and loss of mean binary cross entropy is adopted in the training of MaskR-CNN algorithm, which can be described as:

$$L_{\text{final}} = L(\{P_i\}, \{t_i\} + (L_{cls} + L_{\text{box}} + L_{\text{mask}})), \quad (1)$$

where, L_{cls} and L_{box} are used for classification and regression, L_{mask} represents the partial loss value of Mask, it includes the size of the output $K \times m \times m$, which represents the size of the associated region produced in the image. To maintain area $m \times m$, it is necessary to ensure that the feature map of the region of interest is aligned with the original picture. Finally, the mean value of the cross entropy of all candidate regions is L_{mask} .

3.2.2. Network Training. In the training stage, each batch of GPU trains two pictures, and positive and negative samples, are classified according to the proportion [20]. Follow-up operation zooms the training pictures according to the set pixel level, samples n candidate regions for each picture in the training, and Set the number of positive and negative samples at a ratio of 1:3. Do this one time before each practice. Stochastic disturbance to the training data can improve the convergence of the tracking model and improve the model of the test set. Figure 4 shows the output of MaskR-CNN, which can be seen that even with many complicated challenges, MaskR-CNN still present excellent segmentation.

3.3. Attention Mechanism Module. The attention model (AM) is an important part in neural network structure [21], which is widely used in artificial intelligence related processing field [22]. It can learnt from the human visual system for explanation that when people observe a certain scene, the target range is the whole visible scene, but when it is

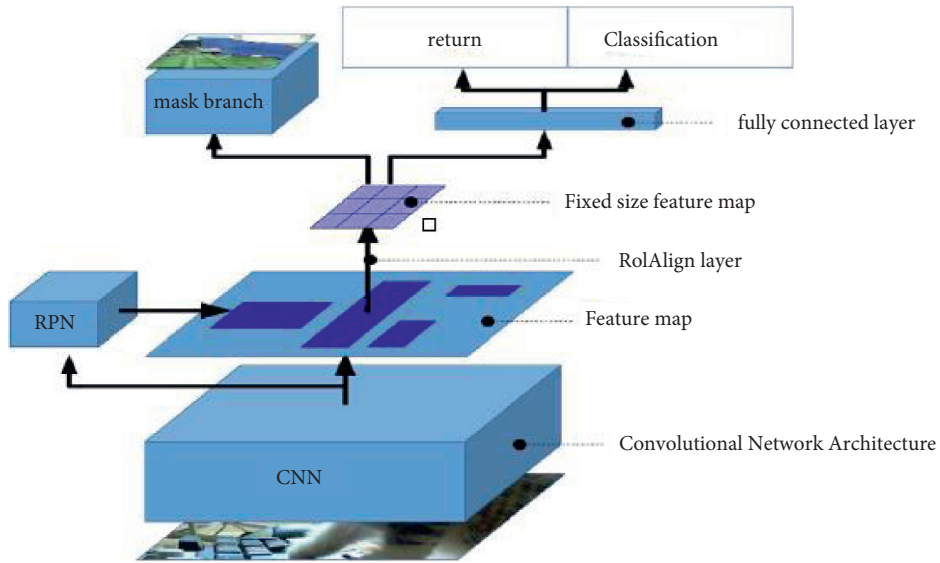


FIGURE 2: Mask R-CNN network structure.

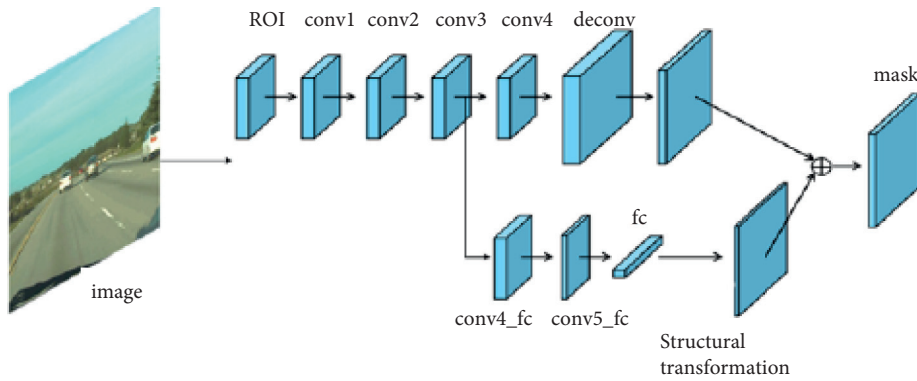


FIGURE 3: Mask generation flow chart.



FIGURE 4: Segmentation result.

necessary to judge a certain target more deeply, people’s attention is focused on the target [23].

The main reasons why attention mechanism is widely used are as follows: (1) Attention model can well solve

problems of multitasking, such as semantic translation and image analysis. (2) Improve the feature extraction performance of target tracking algorithm and improving performance of convolution neural network. (3) Overcome the

challenge of recurrent neural network well, and have operate better in image classification, target detection and other tasks [24]. Therefore, spatial attention and channel attention are applied to guide the task.

Attention mechanism is to transform the spatial original picture information, and save keywords information through attention mechanism. By finding out the regions that need attention in the image information, the important information of the local image is extracted [25]. TVary with input, preprocessing that suitable for tracking tasks can be completed which can be described as:

$$E \in R^{C \times H \times W}: E_j = \alpha \sum_{i=1}^N (S_{ji} D_i) + A_j, \quad (2)$$

where α is the scaling coefficient, S_{ji} is the influence of the feature of the i th position on the j th position, and D is the feature block obtained in convolution.

The channel attention mechanism can be described as an image sequence represented by three channels(R, G, B), where each image channel changes after corresponding convolution. The image matrix of the new channel is transformed that W and H indicate width and height of the image. Moreover, in the module, dot multiplication and Fourier transform are performed on the height and width to decompose the signal of image channel. It can be describe as formula (3):

$$E \in R^{C \times H \times W}: E_j = \beta \sum_{i=1}^c (X_{ji} A_i) + A_j, \quad (3)$$

where β represents the scale coefficient, which is initialized to 0, E represents the feature of a individual channel, and A is the weight between features of channels.

In this paper, the structure of attention mechanism is shown in Figure 5. In computer vision, channel attention mechanism is regarded as a process of extracting various features from semantic association. In practice, the goal is to keep the adaptive deformation and overcome the problem of target similarity, which can greatly improve the tracking efficiency and obtain excellent features by associating channel attention and spatial attention mechanism.

3.4. Tracking Module. The principle undertaking of target following is to recognize foundation and closer view focuses in video groupings, which is moderately less complex than general objective arrangement or target acknowledgment [26]. Therefore, the algorithm adopts a simpler network model that is separated from VGG-M [27]. The structure of tracking module is shown in Figure 6. Before tracking the target, The target is segmented by MaskR-CNN algorithm, and the candidate regions of vehicle target are obtained. Usually, as the network develops further, the multidimensional information of the tracked object will be diluted, which will result in serious consequences or even failure. Simultaneously, the element of the competitor locale got by example division is a lot more modest than that of the original image, which contributes to excellent tracking results.

Although attention mechanism module has great influence on tracking results, overlapping attention mechanism

module will lead to the performance degradation of the model. A module mechanism of space and channel focus is added to the tracking algorithm, and the external information is used to improve the influence on the robustness of the whole tracking algorithm [28]. Formula (4) is used to code the general learning of all training samples:

$$f_1(q_t, K, V) = \sum_{s=1}^m a(q_t, k_s) v_s, \quad (4)$$

There is a query vector t , K is the critical matrix, and V is the value matrix.

In the process of target tracking, simple network architecture is always adopted. Learning from the excellent target detection method and the idea of Feature Pyramid Networks(FPN), the moving target is tracked with large-scale feature map after convolution [29]. And self-adaption gradient descent method (SGD) is taken to train convolutional neural network. In batch training, the error of the network model is calculated, and then different domains in the iteration are managed. For each repetition, 8 frames are selected sequentially. Of the 8 frames, 4 positive samples (IoU >0.7) and 12 negative values (IoU <0.5) are obtained for each frame. On this basis, 32 positive numbers and 96 negative numbers are studied.

The tracking uses both long-term and short-term updates. When long-term update is adopted, it is updated according to the set specified interval. However, if the long-term update fails, the score of the positive sample of the predicted target will be less than 0.5, and then the short-term update mode will be adopted. How to predict the target location region in subsequent frames is a difficult problem. One approach is to allocate multiple matching templates score before the current frame in the video sequence $f^+(x^i)$, $f^-(x^i)$ according to the positive and negative samples stored in our network. Finally, in light of the mean adoption strategy, in video sequence X^* , select the sampling region with the highest score as the optimal tracking object state:

$$X^* = \arg \max_{x^i} f^+(x^i). \quad (5)$$

Using Gaussian distribution theory, 256 candidate regions are generated near the position of the predicted object in the previous frame, and are represented by (x, y, w, h) . Finally, the candidate frames in the original image are clipped and adjusted to 107×107 , which is used for model input. Afterwards, the highest score can be obtained by forward propagation in the previously candidate regions. Finally, The candidate region with that high score is randomly selected, and the average of candidate region scores is the score of the tracking target framework. Moreover, in the algorithm, 20 positive samples and 100 negative samples according to the borders obtained are generated here, where the positive sample region will be resampled if the number of video frames is larger than 50. Otherwise, if the frame number is larger than 20, the negative sample region will be resampled. In addition, a threshold can be set to compare the calculated score. According to the relationship between the score and threshold, the tracking result is judged.

TABLE 1: Configuration of experimental platform used for algorithm simulation.

Configuration content	Model
Processor model	2.30 GHz Intel (R) Xeon (R) gol <i>d</i> 5118
Storage configuration	RAM 16G, 512 G SSD
Graphics card configuration	NVIDIA GTX 1070
Operating system	Windows
Software platform	Pycharm
CUDA	Cuda8.0.61 for ubuntu 16.04
Deep learning platform	Tensorflow1.0

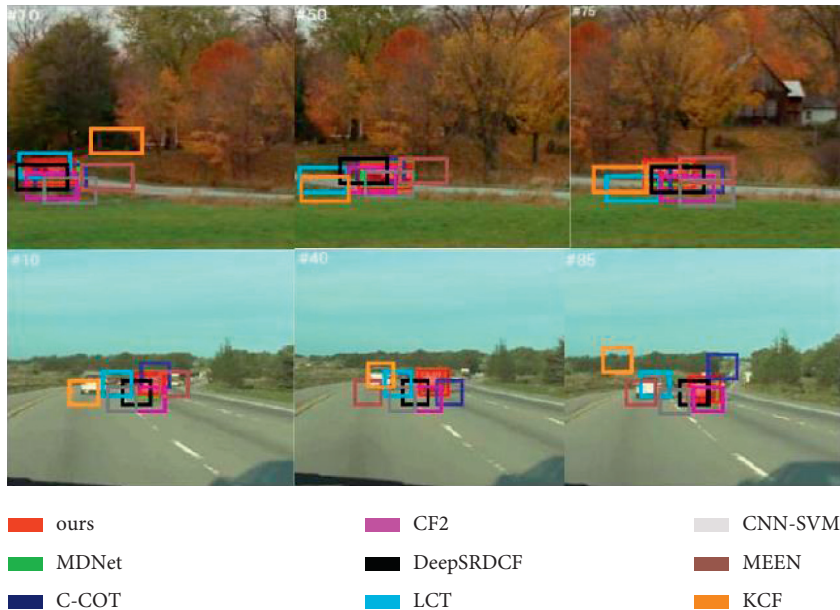


FIGURE 7: Comparison of experimental results.

4.2. *Analysis of Results.* On the OTB data set, results in this paper compares with eight advanced target tracking algorithms, including MDNet, C-COT, DeepSRDCF, CF2, CNN-SVM, MEEM, LCT and KCF. Figure 7 shows the comparison between them:

Through comparative experiments, we can see the improved algorithm can still keep the tracking stable when the target drifts and moves rapidly. At the same time, it can also capture the target when they are greatly deformed and moving rapidly.

The algorithm proposed in this paper performs well in both measures. Such excellent robust performance is mainly attributed to the fact that the features extracted by attention mechanism is combined with traditional features such as appearance and texture to complement each other, which can converge to a more robust result in training. From Table 2, we can see that this method can realize the intelligent filtering of a certain level of features to a certain extent, effectively suppress noise, effectively solve the problem of vehicle target similarity, and improve the tracking performance of the system.

In the experiment, an independent MDNet target tracking model was first established; Then join the channel attention mechanism module to judge finally, two feature modules are added and segmented with examples, to

TABLE 2: Comparison of experimental data.

s	Accuracy	Success rate
Algorithm	0.918	0.678
MDNet	0.909	0.678
C-COT	0.898	0.671
DeepSRDCF	0.854	0.635
CF2	0.837	0.562
CNN-SVM	0.814	0.562
MEEM	0.781	0.554
LCT	0.762	0.530
KCF	0.692	0.475

improve tracking efficiency and solve the problems such as occlusion of vehicle targets. We analyzed the differences between the two models, calculate these indicators by cross-validation method, and evaluate the robustness of the algorithm, in which FIM represents the harmonic average of accuracy and recall rate, and Auc represents its accuracy. The results are shown in Table 3. You can see the accuracy from the comparison, is improved from 90.9% to 90.13% after adding the channel attention mechanism module. In addition The variance decreased from 3.78% to 1.27%, which has better stability. Generally, after adding two attention mechanisms, this method is combined with the image

TABLE 3: Comparison of different models.

Method	Accuracy (%)	Variance (%)	FIM (%)	GM (%)	Auc (%)
MDNet	90.90	3.78	82.10	85.04	67.80
Add channel attention module	90.13	1.27	83.50	87.09	67.80
Algorithm	91.80	0.57	86.79	89.37	67.80

TABLE 4: Comparison of the results of different video sequences.

Video sequence	Accuracy (%)	Variance (%)	FIM (%)	GM (%)	Auc (%)
1	78.35	4.12	77.19	75.21	46.18
2	87.51	2.70	78.35	80.68	59.88
3	89.60	0.66	83.99	87.28	62.98
4	91.80	0.57	86.79	89.37	67.80

segmentation method, so that it has higher tracking accuracy and better stability.

Furthermore, the robustness and accuracy of video data are explored in different time series, and we cut off the test video in units of length ratio with 1, 2, 3 and 4, which forms three groups of test data. The results are shown in Table 4 as follows: the shorter the video sequence is, the better the tracking performance is, and the longer the video sequence is, the better the tracking effect. From the above analysis, it can be known that long video sequences are of great significant to our training tracker.

5. Conclusion

Target tracking is a very important application in computer vision, which has a very wide prospect. Target tracking can be widely promoted in smart traffic supervision, public safety monitoring, unmanned driving and other fields. Because of the outstanding performance of deep learning related technologies in target detection and classification, many researchers have introduced deep learning technology into the field of target tracking. Therefore, in this paper, the instance segmentation algorithm is used to segment the target to be tracked, which is used as the input of the target tracker. [30–32].

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Young Teachers Education and Research Projects of Fujian Province, the project number is JAT191160.

References

- [1] C. Kwan, B. Chou, and L. Kwan, "A comparative study of conventional and deep learning target tracking algorithms for low quality videos," in *Proceedings of the International Symposium on Neural Networks*, June 2018.
- [2] C. Kwan, B. Chou, and L. Kwan, *A Comparative Study of Conventional and Deep Learning Target Tracking Algorithms for Low Quality Videos*, Springer, NY, USA, 2018.
- [3] B. Yang, X. Cao, C. Yuen, and L. Qian, "Offloading optimization in edge computing for deep learning enabled target tracking by internet-of-UAVs," *IEEE Internet of Things Journal*, vol. 8, 2020.
- [4] V. Chandrakanth, A. Murthy, and S. S. Channappayya, "Target tracking in blind range of radars with deep learning," in *Proceedings of the 2020 21st International Radar Symposium (IRS)*, Warsaw, Poland, October 2020.
- [5] L. Zhou, Q. Liu, F. Wu, and Y. Wei, "Deep learning based sensing resource allocation for mobile target tracking," in *Proceedings of the 2020 IEEE 20th International Conference on Communication Technology (ICCT)*, October 2020.
- [6] C. Zhang and Y. Guan, "Research on related technologies of vision target tracking based on discrete differential algorithm for deep learning," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 3, pp. 1–8, 2019.
- [7] T. Roggen, M. Bobic, N. Givehchi, and S. G. Scheib, "Deep Learning model for markerless tracking in spinal SBRT," *Physica Medica*, vol. 74, pp. 66–73, 2020.
- [8] F. Bi, X. Ma, W. Chen et al., "Review on video object tracking based on deep learning," *New Media Magazine*, vol. 1, no. 2, p. 12, 2019.
- [9] Y. Weber and S. Kanarachos, "The correlation between vehicle vertical dynamics and deep learning-based visual target state estimation: a sensitivity study," *Sensors*, vol. 19, no. 22, p. 4870, 2019.
- [10] B. Tuncer, M. Kumru, and E. Zkan, "Extended target tracking and classification using neural networks," in *Proceedings of the 2019 22th International Conference on Information Fusion (FUSION)*, pp. 1–7, IEEE, Ottawa, ON, Canada, July 2019.
- [11] J. Bai, G. Y. Nie, W. Song, Y. Liu, and Y. Wang, "Study of 3D target replacement in AR based on target tracking," in *Proceedings of the 2019 12th Asia Pacific Workshop on Mixed and Augmented Reality (APMAR)*, Ikoma, Japan, March 2019.
- [12] X. P. Xie, X. Wei, X. Wang, X. Guo, J. Li, and Z. Cheng, "Abnormal target tracking and localization algorithm for UAV PV inspection scenarios," *IOP Conference Series: Materials Science and Engineering*, vol. 768, Article ID 072068, 2020.
- [13] J. Doellinger, V. S. Prabhakaran, L. Fu, and M. Spies, "Environment-aware multi-target tracking of pedestrians," *IEEE*

- Robotics and Automation Letters*, vol. 4, no. 2, pp. 1831–1837, 2019.
- [14] D. Sudha and J. Priyadarshini, “An intelligent multiple vehicle detection and tracking using modified vibe algorithm and deep learning algorithm,” *Soft Computing*, vol. 24, no. 21, pp. 1–13, 2020.
- [15] S. M. Marvasti-Zadeh, L. Cheng, H. Ghanei-Yakhdan et al., *Deep learning for visual tracking: a comprehensive survey*, IEEE Transactions on Intelligent Transportation Systems, 2021.
- [16] R. Opromolla, G. Inchingolo, and G. Fasano, “Airborne visual detection and tracking of cooperative UAVs exploiting deep learning,” *Sensors*, vol. 19, no. 19, 2019.
- [17] W.-j. Kang, C. Liu, and G.-l. Liu, “A quantitative attribute-based benchmark methodology for single-target visual tracking,” *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 3, pp. 405–421, 2020.
- [18] F. Jia and S. Wang, *A Hybrid Catheter Localisation Framework in Echocardiography Based on Electromagnetic Tracking and Deep Learning Segmentation*, 2020.
- [19] R. J. . Soldin, “SAR target recognition with deep learning,” in *Proceedings of the 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, October 2018.
- [20] Z. H. Huang, J. Zhan, H. M. Zhao et al., “A review of plant phenotypic image recognition technology based on deep learning,” *Electronics*, vol. 10, no. 1, p. 81, 2021.
- [21] D. Jian and H. Liu, “Research of moving target tracking technology based on LRCN,” in *Proceedings of the 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*, 2018.
- [22] S. H. Silva, P. Rad, N. Beebe, K.-K. R. Choo, and M. Umaphathy, “Cooperative unmanned aerial vehicles with privacy preserving deep vision for real-time object identification and tracking,” *Journal of Parallel and Distributed Computing*, vol. 131, no. SEP, pp. 147–160, 2019.
- [23] J. Choi, J. Kwon, and K. M. Lee, “Deep meta learning for real-time target-aware visual tracking,” in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, pp. 911–920, 2019.
- [24] A. Zgaren, W. Bouachir, and R. Ksantini, “Coarse-to-fine object tracking using deep features and correlation filters,” in *Proceedings of the International Symposium on Visual Computing*, pp. 517–529, Springer, Cham, Germany, 2019.
- [25] W. Zhao, L. Shen, and Y. Wu, *Automatic Target Positioning and Tracking for Image-Guided Radiotherapy without Implanted Fiducials*, 2019.
- [26] A. Rohan, M. Rabah, and S. H. Kim, “Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2,” *IEEE Access*, vol. 7, no. 99, 2019.
- [27] X. Huang, K. Wang, H. Yin, S. Zheng, X. Meng, and S. Zhang, “Learning a reliable decision making policy for robust tracking,” in *Proceedings of the 2019 IEEE Visual Communications and Image Processing (VCIP)*, IEEE, Sydney, NSW, Australia, December 2020.
- [28] X. Zhang, P. Ye, S. Peng, J. Liu, K. Gong, and G. Xiao, “SiamFT: an RGB-infrared fusion tracking method via fully convolutional siamese networks,” *IEEE Access*, vol. 7, 2019.
- [29] W. Zhang, Y. Du, Z. Chen, D. Jianhua, and L. Peizhong, “Robust adaptive learning with Siamese network architecture for visual tracking,” *The Visual Computer*, vol. 37, no. 9, 2020.
- [30] M. S. Bahraini, A. B. Rad, and M. Bozorg, “SLAM in dynamic environments: a deep learning approach for moving object tracking using ML-RANSAC algorithm,” *Sensors*, vol. 19, no. 17, p. 3699, 2019.
- [31] M. M. Rahman, M. Fiaz, and S. K. Jung, “Efficient visual tracking with stacked channel-spatial attention learning,” *IEEE Access*, vol. 8, no. 99, 2020.
- [32] R. Yan, D. Xiong, and Y. U. Qinghua, “Object tracking algorithm based on parallel tracking and detection framework and deep learning,” *Journal of Computer Applications*, vol. 39, no. 2, p. 343, 2019.

Research Article

Aircraft Image Recognition Network Based on Hybrid Attention Mechanism

Yanfeng Wang ¹, Yinan Chen ², and Runmin Liu ^{3,4}

¹College of Systems Engineering, National University of Defense Technology, Changsha 410082, China

²College of Computer & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China

³College of Sports Engineering & Information Technology, Wuhan Sports University, Wuhan 430079, China

⁴AiTech Artificial Intelligence Research Institute, Changsha 410000, China

Correspondence should be addressed to Yanfeng Wang; wangyanfeng@nudt.edu.cn

Received 8 March 2022; Revised 28 March 2022; Accepted 30 March 2022; Published 18 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Yanfeng 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.

With the deepening of deep learning research, progress has been made in artificial intelligence. In the process of aircraft classification, the precision rate of aircraft picture recognition based on traditional methods is low due to various types of aircraft, large similarities between different models, and serious texture interference. In this article, the hybrid attention network model (BA-CNN) to implement an aircraft recognition algorithm is proposed to solve the above problems. Using two-channel ResNet-34 as a characteristic extraction function, the depth of network is increased to improve fine-grained characteristic extraction capability without increasing the output characteristic dimension. In the network to introduce a hybrid attention mechanism, respectively, between the residual units of two ResNet-34 channels, channel attention and spatial attention modules are added, more abundant mixed characteristics of attention are obtained, space and characteristics of the local characteristics of the channel response are focused, the characteristics of redundancy are reduced, and the fine-grained characteristics of learning ability are further enhanced. Trained and tested on FGVC-aircraft, a public fine-grained pictures dataset, the recognition precision rate of the BA-CNN networks model reached 89.2%. It can be seen from the experimental results, the recognition precision rate of the original model is improved effectively by using this method, and the recognition precision rate is higher than most of the existing mainstream aircraft recognition ways.

1. Introduction

Aircraft picture recognition has been a research hotspot of fine-grained picture recognition of machine vision field. In recent years, with the continuous deepening of artificial intelligence, aircraft recognition ways based on deep learning have been adopted in airport management, military, and other fields. Aircraft picture recognition is a typical representative of fine-grained picture classification. The fine-grained nature of aircraft pictures leads to large interclass [1] similarity and intraclass variability among aircraft models, which in turn affects the precision rate of aircraft classification. How to effectively use the information about aircraft pictures and research a better performance aircraft recognition algorithm is not only a certain promotion significance

of the application of aircraft recognition system but also a certain reference values for the solution to other picture recognition problems [2].

The current fine-grained aircraft picture recognition algorithms mainly include two directions: one is the recognition algorithm based on traditional picture processing and the other is the recognition algorithm based on deep learning. The recognition algorithms based on traditional picture processing mainly include template matching algorithm [3] and traditional characteristic extraction recognition algorithm [4]. The template matching algorithms mainly analyzes the regions of the target picture, compares the characteristics of each region with those of the template, and determines the category of the target picture according to the degree of similarity. The template matching algorithm

requires low contrast of the picture and has a good ability to deal with the presence of occlusion in the picture, but it has the disadvantages of large computation and poor real-time performance. Since traditional picture recognition algorithms often have difficulty in finding high-quality characteristics, resulting in low recognition precision rate, the mainstream research direction is currently based on deep learning recognition methods, which mainly use CNNs that are well suited for processing two-dimensional picture data, such as Chevalier et al. [5] proposed a deep learning-based LR-CNN network model for picture classification, Huihui Li [6] proposed a PCNN network model for aircraft classification, Malekzadeh [5] proposed a DNN network model for extracting aircraft characteristics, Ting-Bing Xu [7] proposed an “end-to-end” FCN for fast aircraft classification, Tsung-Yu Lin et al. [8] proposed a B-CNN network model.

Although the deep learning way has been utilized to obtain better aircraft recognition results than traditional methods, the current deep learning method has a single network structure, which ignores the channel and spatial relationships that aircraft pictures have, resulting in a certain loss of information and hindering further improvement in recognition precision rate. Moreover, the current convolutional neural network structure dealing with recognition problems of aircraft pictures is to train separate networks according to each input, and then through each separate network, the recognition prediction is performed for each input separately. After analysis, it is known that this method has two main problems: on the one hand, the number of networks used is too many and unrelated to each other, thus increasing the cost of network training and causing information redundancy, resulting in time inefficiency; on the other hand, the inherent relationship between aircraft picture channels and space is ignored, which limits the improvement in recognition results.

Thereby, this article proposes a hybrid attention network model (BA-CNN) to implement the aircraft recognition algorithm.

The dedications to this article are as follows:

- (1) Using two-way ResNet-34 as the characteristic extraction function, adding the depth of the network to improve the fine-grained characteristics extraction capability without adding the output characteristic dimension
- (2) A hybrid attention mechanism is introduced into the network to add the channel attention module and spatial attention module between the residual units of two ResNet-34 channels respectively to obtain richer hybrid attention characteristics, focus on local characteristic channels and spatial response parts in the characteristic map, reduce characteristic redundancy, and further enhance the fine-grained characteristics learning capability at the same time

The recognition precision rate of the BA-CNN network models reaches 89.2% when trained and tested for the publicly available fine-grained picture dataset FGVC-aircraft. Experimental consequences showed that the recognition precision rate of the original model is improved

effectively by using this method, and the recognition precision rate is better than most of the existing mainstream aircraft recognition ways.

2. Materials and Methods

2.1. Data Acquisition. Most of the picture recognition network models are required to have a certain size of dataset to optimize the algorithm. For the aircraft recognition problem, the pictures and labels used in this article are mainly collected from the FGVC-aircraft [9] dataset, which contains 10,000 aircraft pictures, with the size of every picture ranging from 33 KB to 1 MB. Each aircraft picture is uniquely labeled with “manufacturer,” “series,” and “model.” As shown in Figure 1, this aircraft picture is labeled as the 310-300 model of the Airbus 310 series designed by the airbus manufacturer.

Figure 2 shows some of the aircraft pictures in the dataset used in this article. For the aircraft dataset in this article, when aircraft is considered as a large category, they are an object of high intra-class similarity, and the objects of its class all have the basic structure of an aircraft (e.g., fuselage, wings, and engines), so it is inherently difficult to subdivide them internally. It is easier to classify various aircraft from different manufacturers compared to classifying various types of aircraft produced by the same manufacturer. This is because aircrafts produced by the same manufacturer are similar in structure and appearance, thus making recognition more difficult. In addition, factors such as complex backgrounds and aircraft crippling can also affect the precision rate of the classification. In order to obtain the reliable experimental consequence and improve the applicability to the multi-label recognition problem, manufacturers with multiple types of aircraft were selected in the construction of the dataset, while the number of different label data was appropriately balanced, and 7000 pictures were randomly selected from them to form the train set; however, the remaining 3000 pictures were used as the test set, so that the experiments can accurately test the effectiveness of the algorithms in this article.

Considering the quality and quantity of the pictures in the dataset, the dataset is expanded and some of the pictures are enhanced, and the aircraft pictures are preprocessed in a specific way to heighten the recognition of the pictures and protrude the aircraft part of the pictures. In order to facilitate the learning optimization of the algorithmic network in this article, taking into account the needs of the algorithms used in this article. The dataset plays a top-down role in the solution to the whole recognition problem, and it is directly related to the specific representation of the aircraft recognition problem and the structure of the algorithm for solving the problem.

In this article, we mainly use the spatial domain enhancement method to sharpen the relatively blurred and shadowy aircraft pictures and increase the contrast of the pictures. The spatial domain method is mainly a direct operational processing of pixel grayscale values in the spatial domain, and the more common methods include gray-level transformation, histogram correction, picture space

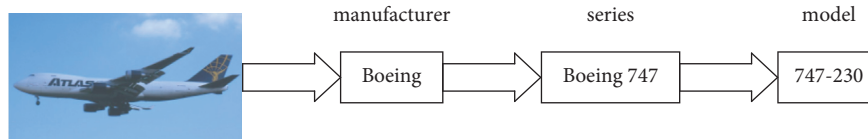


FIGURE 1: Sample labeling diagram.

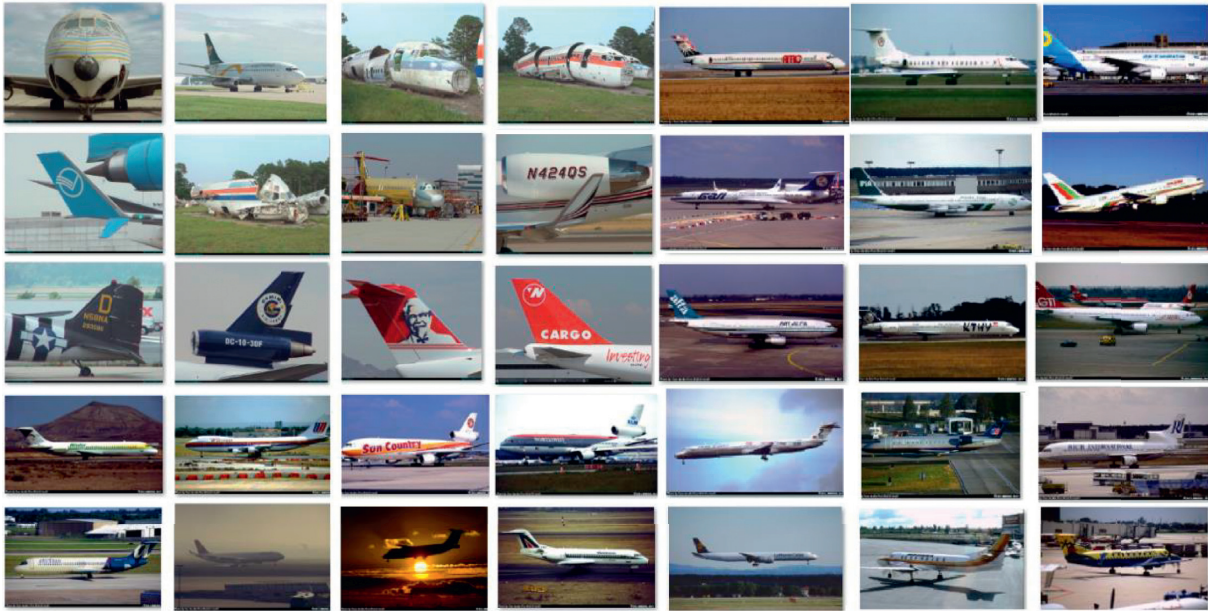


FIGURE 2: Selected aircraft sample pictures.

smoothing, and sharpening processing. Grayscale transformation mainly uses the mapping function to change the gray-level range of the picture, which can be corrected for a certain part of the picture or the whole picture underexposed, in order to strengthen the gray contrast of the picture; histogram correction by using a specific conversion function to change the gray distribution of the picture, so that the high and low brightness areas of the picture gray map have approximately the same intensity, which can make the picture with the desired gray distribution, so as to selectively highlight the desired gray distribution. Thus, the desired picture characteristics are selectively highlighted to meet the needs of a specific task. Figure 3 shows an aircraft picture before and after spatial domain enhancement. The enhancement technique can sharpen the edges of the aircraft picture, highlight the outline of the aircraft, and reduce the background color to improve the contrast between the aircraft and its environment. Although picture enhancement does not increase the inherent information of the data, it increases the dynamic adjustment range of the selected characteristics and facilitates object classification.

2.2. BA-CNN Network. BA-CNN consists of two ResNet-34 networks as the characteristic extraction function and adds the channel attention module and spatial attention module to the two-way characteristic function, and the convolutional characteristic extracted by the two networks are

bilinearly combined to achieve end-to-end weakly supervised classification. Using ResNet-34 with added hybrid attention as the characteristic extraction function, the characteristic representation capability is somewhat enhanced to pay sufficient attention to the influence of discriminative parts of objects on classification. The BA-CNN network combines the two output characteristics by outer product to generate high-dimensional bilinear characteristics.

2.2.1. Network Structure. The hybrid attention network model uses two parallel CNNs to achieve the characteristic extraction process. In this article, the characteristic extraction process selects the ResNet34 network as the characteristic extraction function and replaces the final fully connected and Softmax layers of the two CNNs with a bilinear pooling layer, and the final bilinear characteristic representation vector is obtained by bilinear combination and pooling of the output results of the two eigenfunctions. BA-CNN network models utilize the second-order statistical information about the picture to model the combined interactions between local characteristics of translation invariance and achieve weakly supervised recognition with only picture category labels. Meanwhile, the BA-CNN network simplifies the gradient calculation, making its end-to-end network model easier to be trained, and the architecture of this network model is shown in Figure 4.



FIGURE 3: Aircraft pictures before and after spatial domain enhancement.

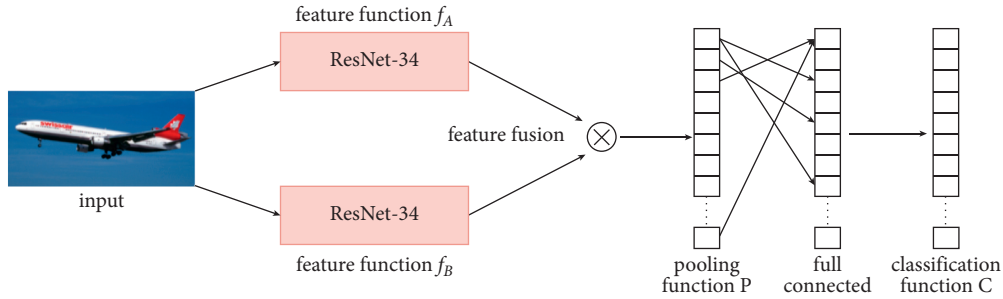


FIGURE 4: Hybrid attention network model architecture.

The BA-CNN network model can be represented by a quadratic function $B = F(f_A, f_B, P, C)$, where f_A and f_B are characteristic functions, P is the pooling function, and C is the recognition function. The characteristic functions f_A and f_B represent a mapping relationship $f: I \times L \rightarrow R^{K \times T}$, where I represents the input picture, $L \in R^K$ represents the location range of the input picture, and f maps them into a $K \times T$ dimensional characteristic map, where K represents the spatial resolution size of the characteristic map and T denotes the characteristic channel dimension. The characteristic vectors m and n are bilinearly combined through the outer product operation (here refers to the tensor product in linear algebra [10]), and the bilinear characteristic $b(l, I, f_A, f_B) = f_A(l, I) \otimes f_B(l, I) = m^T n$ is obtained, where $b \in R^{w \times h \times 1 \times t_2}$, $l \in L$; $L \in R^K \otimes$ represents the outer product operation of the vector and T represents the product of vector m and n -channel dimension $t_1 \times t_2$, and the schematic diagram of characteristic fusion is shown in Figure 5.

To further obtain the picture descriptors, the pooling function P aggregates the bilinear characteristics at each

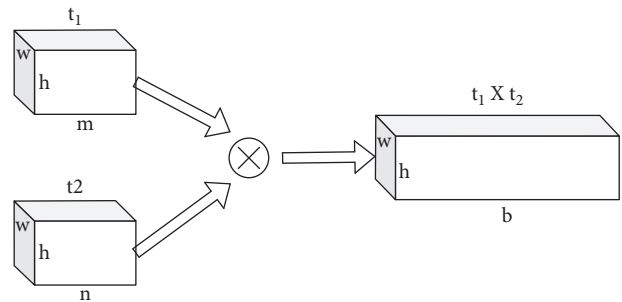


FIGURE 5: Schematic diagram of characteristic fusion.

position in the picture to obtain a global representation of the picture. One pooling approach is to sum all the bilinear characteristics cumulatively, that is, $\phi(I) = \sum_{l \in L} b(f_A, f_B, l, I) = \sum_{l \in L} m^T n$. The pooling function P obtains the vector by $\phi(I)$ converting the bilinear characteristic b into a $t_1 t_2 \times 1$ dimensional column vector, denoted as x will be subjected to the signed open-square

operation $\text{sign}(x)\sqrt{x/y}$, to which a $L2$ regularization constraint $y/\|y\|_2$ is applied to obtain the final representation vector z that will be an input to the recognition function C to complete the classification.

2.2.2. Characteristic Extraction. The aircraft feature extraction part is mainly composed of two channels ResNet-34 as a general classification network. Compared with VGGNet, although it has a certain characteristic representation ability, it has certain limitations of discriminant local characteristic extraction in fine-grained picture recognition [11]. The more network parameters of VGGNet consume a large amount of computational resources, leading to higher memory occupancy, making the network model limited in terms of speed and precision rate, which affects the practicality.

With the development of deep convolutional neural networks, the network depth has an important impact on the picture recognition precision rate. Usually, when there are few layers, increasing the depth can get better characteristic extraction and improve the recognition precision rate; however, when there are many layers (e.g., if more than 30 layers), continuing to increase the depth will bring higher training and testing errors, making it difficult to converge when training the network, but reducing the precision rate [12]. The main reason for the elevated error is the phenomenon of gradient disappearance and gradient explosion when increasing the number of layers, especially the problem of gradient disappearance, which prevents the gradient from being effectively updated to the shallow network of weight adjustment during back propagation. To address these problems, He et al. [13] propose a deep residual network (ResNet). Compared with other convolutional neural networks, ResNet adopts a residual learning structure to transfer the original input information directly to the next layer of the network of jump connections, while the gradients are also directly transferred to the previous layer through jump connections when back propagating. The basic structure of the residual network is the residual unit, and Figure 6 shows the structure of the residual unit.

Let x be the input of the residual unit and $H(x)$ be the expected output of the residual unit. If x is passed directly to the output part as the initial result, the network only needs to learn $F(x) = H(x) - x$ at this time, which is a basic residual unit of ResNet. This is a basic residual unit of ResNet. With this residual unit structure, ResNet is equivalent to changing the learning target for the fully output value $H(x)$ to the distinguish between the output value and the input value $H(x) - x$, which simplifies the network learning target and reduces the learning difficulty. ResNet is proposed to effectively overcome the trouble of disappearing gradients in deep networks, which makes the recognition precision rate significantly improved and has good portability. BA-CNN network model using ResNet-34 has a deeper network structure and can learn local characteristics in fine-grained pictures more finely than VGGNet to improve the recognition precision rate.

Therefore, in this article, the two-way ResNet-34 was chosen as the characteristic function part of the network model, and ResNet-34 contains five groups of convolutional blocks conv1-conv5, 33 convolution layers, and one

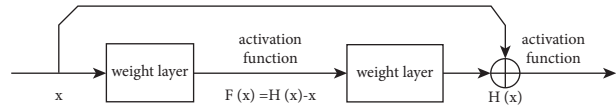


FIGURE 6: Structure of residual unit.

complete connected layer, total of 34 layers. The final fully connected layer is removed from the two-way ResNet-34 as the backbone of the network model, and the output characteristic dimension of the last convolutional layer of the network is 512. Compared with using the VGGNet network of characteristic extraction, the ResNet-34 networks increase the depth, while maintaining the same output characteristic dimension, avoiding the exponential increase in the characteristic dimension after the bilinear combination.

2.2.3. Hybrid Attention Module. The attention mechanism is proposed by the imitation of human brain's special vision signals processing mechanism. When human is to observe and identify objects, there will be a targeted focus on target, while ignoring some significant part of the background and global information, the mechanism of selective attention in fine-grained picture recognition task rely on consistent discriminant characteristics of parts [14]. Therefore, in order to further extract judicious part characteristics, a hybrid attention mechanism is introduced in the network-using the CBAM (convolutional block attention module) algorithm to extract attention weight maps in both channel and spatial dimensions in the two characteristic functions of the backbone network, and to distribute the weights distributed in the original characteristic maps for characteristic fusion, and the fused channel attention and spatial attention modules are added between the convolutional blocks of the first network conv4 and conv5 and the second network conv2 and conv3, respectively, to obtain attention characteristics with different dimensions and more richness.

2.2.4. Channel Attention Module. The convolutional characteristic maps produced by the characteristic functions contain different characteristic channels, and in the fine-grained picture recognition problem, each characteristic channel may represent different information in the picture, some of which contain irrelevant picture background information and are redundant. Therefore, focusing on the characteristic channel including the discriminant sites information and giving it a higher weight distribution can effectively enhance the fine-grained recognition effect. In this article, the channel attention module is added between the conv4 and conv5, and the construction of the channel attention is shown in Figure 7.

The characteristic extraction and characteristic fusion steps of the channel attention module are as follows:

- (1) The convolutional characteristic map f_A generated by the characteristic function is used as the original input F , set $F \in R^{w \times h \times t}$, where $w \times h$ represents the spatial dimension of F , and t represents the number of channels. F is compressed in the spatial

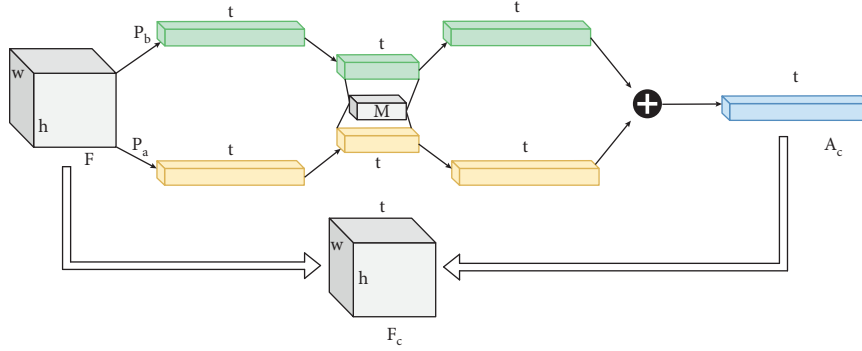


FIGURE 7: Channel attention module.

dimension, and the characteristics of the same channel are compressed into a number for extract channel attention effectively. This step can be achieved through pooling operation.

- (2) Take a way of pooling of multi-scale, respectively using maximum pooling functions average pooling p_m and p_a to dimension reduction of F , get two $1 \times 1 \times t$ characteristic vector, the size of the two input vectors in the same shared network in order to get the attention of the weight distribution channel dimension, sharing network comprises a hidden layer of multilayer perceptron into a unit.
- (3) The two output vectors after reassigning the attention weights are subjected to the corresponding element summation operation, and the combined characteristic vectors are mapped using the Sigmoid activation function to generate the channel attention weights $A_c A_c \in R^{1 \times 1 \times t}$
- (4) Characteristic fusion is carried out between the attention weight A_c and the original characteristic graph F . Here, a fusion method of multiplying corresponding elements is adopted to finally obtain the fused attention characteristic graph F_c and $F_c \in R^{w \times h \times t}$. The original input characteristic F in f_A is replaced by F_c to realize the attention extraction of channel dimension.

2.2.5. Spatial Attention Module. Different from the channel attention module, the spatial attention module pays more attention to the spatial position information of the discriminant part, which is a supplement to the channel attention. Add a spatial attention module between the second channel characteristic functions conv2 and conv3, and the structure of the spatial attention is shown in Figure 8.

The steps of characteristic extraction and characteristic fusion of spatial attention are as follows:

- (1) The convolutional characteristic map f_B produced by the characteristic function is used as the original input $G \in R^{w \times h \times t}$, where $w \times h$ the size of the spatial dimension represented G by t the number of channels, will be compressed G along the channel axis direction to extract spatial attention

information, and a column of channel values is compressed into one channel, which is achieved by pooling of channel dimensions in this step.

- (2) The same multi-scale pooling approach is used, and the maximum pooling function p_m and the average pooling function are adopted p_a G to decrease the dimensionality to obtain two $w \times h \times 1$ size characteristic maps, and the two characteristic maps are stitched together along the channel axis direction using the corresponding element summation way to obtain a new characteristic map of one $w \times h \times 2$ size.
- (3) Convolution of the spliced characteristic map using a 7×7 convolution kernel, again compressing its size to $w \times h \times 1$, $A_s \in R^{w \times h \times 1}$, and mapping the convolved characteristic map using the Sigmoid activation function to produce a spatial attention map A_s .
- (4) Finally, the spatial attention map is fused A_s with the original characteristic map G using the corresponding element dot product way to get the fused spatial attention characteristic map G_s , $G_s \in R^{w \times h \times t}$ and the original input characteristics G in the G_s replacement f_B are used to achieve attention extraction in spatial dimensions.

After adding two attention modules with two dimensions, the network acquires richer attention characteristics. The residual attention construction of the BA-CNN network model in this article is shown in Figure 9. The above two improvement methods make the network model BA-CNN with stronger local characteristic extraction ability, while fewer network parameters make the network model easier to be trained and reduce the overfitting phenomenon; at the same time, by adding attention modules of different dimensions to the residual network to obtain richer information, the BA-CNN network model can focus on and learn fine grained. This is the key to the fine performance of fine-grained recognition.

3. Results

3.1. Experimental Environment. The experimental part of BA-CNN network model is mainly divided into data preprocessing and model training. In this article, we use the

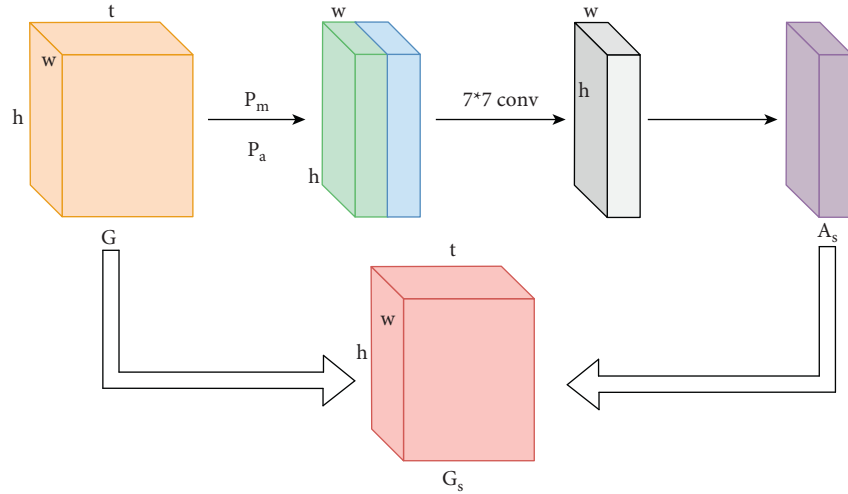


FIGURE 8: Spatial attention module.

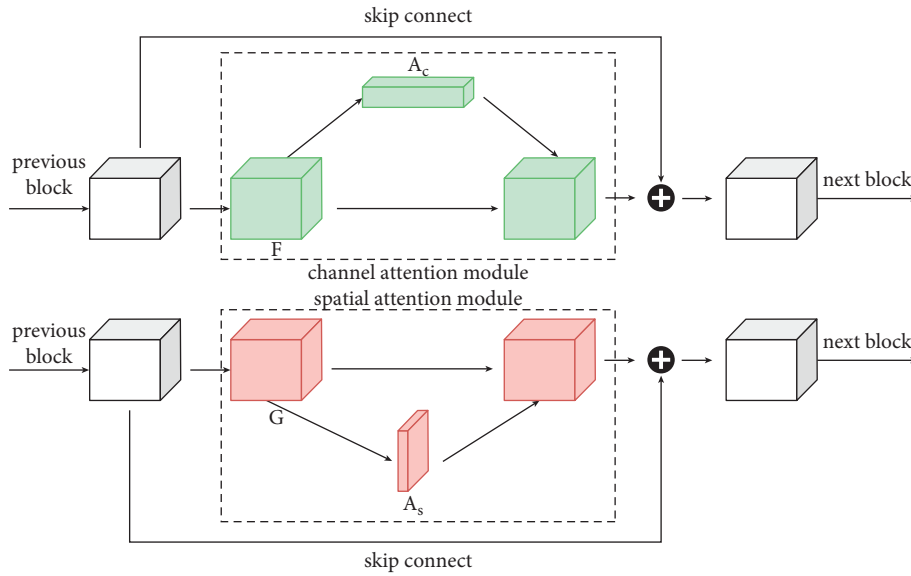


FIGURE 9: Residual attention structure of BA-CNN network model.

PyTorch [15] as the platform, and use one NVIDIA 3090 GPU to train on the aircraft picture dataset by stochastic gradient descent method in parallel. Due to the small size of fine-grained picture dataset and limited training and testing pictures, training directly on the aircraft picture dataset may result in the network failing to converge, so the ResNet-34 network parameters pretrained on the ImageNet dataset are used for initialization, and then the network model is fine-tuned on the aircraft picture dataset. The network model was trained and optimized using the Adam [16] optimizer to train and optimize the network, with the training batch size set to 128, the first- and second-order moment estimation exponential decay rates set to 0.9 and 0.99, respectively, and the learning rate set to 0.001.

Hardware environment: Intel Core i7 12700k; 1T Memory; Nvidia RTX 3090; 32G RAM.

Software environments: CUDA Toolkit 11.1; CUDNN V11.3; Python 3.9; Pytorch 1.8.1; Windows 10.

3.2. *Experiments and Analysis of Results.* In order to comprehensively verify the effectiveness of the method in this article and to be able to better compare the recognition consequences of BA-CNN network model and its aircraft recognition algorithm, several experiments are conducted in this article as follows:

3.2.1. *Ablation Experiments.* In this section of the ablation experimental protocol, the following network structures are compared:

TABLE 1: Experimental analysis of the ablation of the method in this article on the FGVC-Aircraft dataset.

Approach	Backbone	Accuracy (%)
BA-CNN (resnet $\times 2$)	ResNet-34 $\times 2$	85.0
BA-CNN (channel attention)	ResNet-34 $\times 2$ + channel attention	86.2
BA-CNN (spatial attention)	ResNet-34 $\times 2$ + spatial attention	86.5
BA-CNN (channel and spatial attention)	ResNet-34 $\times 2$ + channel and spatial attention	89.2

TABLE 2: Comparison of the recognition precision rate of different weakly supervised algorithms.

Approach	Backbone	Precision rate (%)
Two-level attention	VGG19	77.9
NAC	VGG19	81.01
B-CNN	VGG-M + vgg-d	84.1
ST-CNN	Inception-v2 $\times 3$	84.1
DVAN	VGG-19 $\times 3$	79.0
RA-CNN	VGG-19 $\times 3$	85.3
MA-CNN	VGG-19 $\times 3$	86.5
MAMC	ResNet-101	86.5
BA-CNN	ResNet-34 $\times 2$	89.2

- (1) The network-using only two-way characteristic extraction function is represented by BA-CNN (ResNet-34 $\times 2$)
- (2) The BA-CNN (channel attention) is used to represent the network after adding only the channel attention module
- (3) The BA-CNN (spatial attention) is used to represent the network after adding only the spatial attention module
- (4) The BA-CNN (channel and spatial attention) is used to represent the network after simultaneously adding the channel attention and spatial attention to the two-way characteristic function, respectively

The experimental consequences are listed in Table 1 (backbone indicates the underlying network used for the different methods).

The results from Table 1 show that the recognition precision rate of the network after adding only the channel attention module, the spatial attention module and both modules improve by 1.0%, 1.5%, and 4.2%, respectively, over the original bilinear network model, and the highest recognition precision rate of the network is achieved after adding both attention modules simultaneously.

3.2.2. Comparison Experiments. The BA-CNN aircraft recognition networks model proposed in this article does not require additional labeling information such as object labeling box and part location and only uses category labels to implement a recognition network [17] model based on weakly supervised information. The two-level attention network model, NAC[18], B-CNN, ST-CNN [19], DVAN [20], RA-CNN [21], MA-CNN, and MAMC[22]and other mainstream weakly supervised recognition algorithms in recent years, and the experimental results of the method proposed in this study and the above methods on the FGVC-aircraft dataset are compared, and the results are listed in

Table 2 (backbone denotes the underlying network used by the network model and the precision rate represents the recognition precision rate).

As can be seen from the consequences of the above table, the recognition results of the method proposed in this study for the FGVC-aircraft dataset are all better than the mainstream weakly supervised methods of recent years. The consequences show that the BA-CNN networks in this article in the addition of channel attention and spatial attention modules can focus on distinguish parts in fine-grained pictures, strengthen the extraction of local characteristics, and obtain good recognition results from the fine-grained aircraft picture dataset.

4. Discussion

The BA-CNN hybrid attention networks model proposed in this article uses two-way ResNet-34 as the characteristic extraction function and adds a channel attention module and a spatial attention module among the residual units to achieve the introduction of a hybrid attention mechanism that strengthens the extraction of discriminative local characteristics of fine-grained pictures. The results of ablation and comparison experiments on several fine-grained picture datasets show that the way in this article can effectively improve the precision rate of the aircraft recognition model and outperforms the recognition precision rate of most of the mainstream weakly supervised algorithms in recent years. On the other hand, since the combination of bilinear characteristic vector outer product will greatly increase the characteristic dimensionality and consume computational resources, it is the progress direction of the subsequent work of this article to decrease the dimensionality of bilinear characteristics and improve the practicality of the network model, while minimizing the loss of recognition precision rate.

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.

Authors' Contributions

Yinan Chen was involved in methodology, original draft preparation, conceptualization, and data curation. Guoxiong Zhou was involved in validation and project administration.

Acknowledgments

The authors are grateful to all members of the Computer Science College of Central South University of Forestry and Technology for their advice and assistance in the course of this research.

References

- [1] W. Zhang, G. Zhou, A. Chen, Y. Hu, and Y. Hu, "Deep multi-scale dual-channel convolutional neural network for Internet of Things apple disease detection," *Computers and Electronics in Agriculture*, vol. 194, Article ID 106749, 2022.
- [2] J. Li, G. Zhou, A. Chen et al., "Adaptive linear feature-reuse network for rapid forest fire smoke detection model," *Ecological Informatics*, vol. 68, Article ID 101584, 2022.
- [3] S. Omachi and M. Omachi, "Fast template matching with polynomials," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2139–2149, 2007.
- [4] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.
- [5] M. Chevalier, N. Thome, M. Cord, J. Fournier, G. Henaff, and E. Dusch, "LR-CNN for fine-grained classification with varying resolution," in *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP)*, pp. 3101–3105, IEEE, Quebec City, Canada, 27–30 September 2015.
- [6] H. Li, X. Jin, N. Yang, and Z. Yang, "The recognition of landed aircrafts based on PCNN model and affine moment invariants," *Pattern Recognition Letters*, vol. 51, pp. 23–29, 2015.
- [7] T. B. Xu, G. L. Cheng, J. Yang, and C. L. Liu, "Fast aircraft detection using end-to-end fully convolutional network," in *Proceedings of the 2016 IEEE International Conference on Digital Signal Processing (DSP)*, pp. 139–143, IEEE, Beijing, China, 16–18 October 2016.
- [8] T. Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear cnn models for fine-grained visual recognition," in *Proceedings of the IEEE international conference on computer vision*, pp. 1449–1457, Santiago, Chile, 7–13 December 2015.
- [9] S. Maji, E. Rahtu, J. Kannala, and B. Matthew, "Fine-grained visual classification of aircraft," 2013, <http://arXiv.org/abs/1306.5151>.
- [10] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [11] W. Yang and L. Libo, "Bilinear residual attention networks for fine-grained image classification," *Laser & Optoelectronics Progress*, vol. 57, no. 12, p. 121011, 2020.
- [12] Z. Ming, L. Xiaoqi, and W. Liang, "Multiplicative denoising method based on deep residual learning," *Laser & Optoelectronics Progress*, vol. 55, no. 3, Article ID 031004, 2018.
- [13] 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, 27–30 June 2016.
- [14] 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*, vol. 19, pp. 1–5, 2021.
- [15] A. Paszke, S. Gross, and S. Chintala, *Automatic Differentiation in Pytorch* Proceedings of the 2017 NIPS Workshop, LongBeach, CA, USA, 2017.
- [16] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, <http://arXiv.org/abs/1412.6980>.
- [17] J. Wang, T. Zheng, P. Lei, and X. Bai, "Ground target classification in noisy SAR images using convolutional neural networks," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 4180–4192, 2018.
- [18] M. Simon and E. Rodner, "Neural activation constellations: unsupervised part model discovery with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 1143–1151, Santiago, Chile, December 2015.
- [19] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [20] J. Fu, H. Zheng, and M. Tao, "Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, 21–26 July 2017.
- [21] B. Zhao, X. Wu, J. Feng, Q. Peng, and S. Yan, "Diversified visual attention networks for fine-grained object classification," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1245–1256, 2017.
- [22] M. Sun, Y. Yuan, F. Zhou, and E. Ding, *Multi-Attention Multi-Class Constraint for Fine-grained Image Recognition*, Springer, Cham, 2018.

Research Article

DCCAM-MRNet: Mixed Residual Connection Network with Dilated Convolution and Coordinate Attention Mechanism for Tomato Disease Identification

Yujian Liu,¹ Yaowen Hu,¹ Weiwei Cai ,^{2,3} Guoxiong Zhou ,¹ Jialei Zhan,¹ and Liujun Li⁴

¹College of Computer & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China

²School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

³Graduate School, Northern Arizona University, Flagstaff, AZ 86011, USA

⁴Missouri University of Science & Technology, Department of Civil, Architectural and Environmental Engineering, Rolla, MO 65409, USA

Correspondence should be addressed to Guoxiong Zhou; zhougx01@163.com

Received 22 February 2022; Revised 8 March 2022; Accepted 11 March 2022; Published 15 April 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Yujian 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.

Tomato is an important and fragile crop. During the course of its development, it is frequently contaminated with bacteria or viruses. Tomato leaf diseases may be detected quickly and accurately, resulting in increased productivity and quality. Because of the intricate development environment of tomatoes and their inconspicuous disease spot features and small spot area, present machine vision approaches fail to reliably recognize tomato leaves. As a result, this research proposes a novel paradigm for detecting tomato leaf disease. The INLM (integration nonlocal means) filtering algorithm, for example, decreases the interference of surrounding noise on the features. Then, utilizing ResNeXt50 as the backbone, we create DCCAM-MRNet, a novel tomato image recognition network. Dilated Convolution (DC) was employed in STAGE 1 of the DCCAM-MRNet to extend the network's perceptual area and locate the scattered disease spots on tomato leaves. The coordinate attention (CA) mechanism is then introduced to record cross-channel information and direction- and position-sensitive data, allowing the network to more accurately detect localized tomato disease spots. Finally, we offer a mixed residual connection (MRC) technique that combines residual block (RS-Block) and transformed residual block (TR-Block) (TRS-Block). This strategy can increase the network's accuracy while also reducing its size. The DCCAM-classification MRNet's accuracy is 94.3 percent, which is higher than the existing network, and the number of parameters is 0.11 M lesser than the backbone network ResNeXt50, according to the experimental results. As a result, combining INLM and DCCAM-MRNet to identify tomato diseases is a successful strategy.

1. Introduction

Tomatoes are a globally important vegetable crop [1]. However, diseases can harm tomatoes, reducing their quality and yield. Leaf mold, Septoria leaf spot, yellow leaf curl virus, tomato mosaic virus, target spot, and two-spotted spider mite are all common diseases of tomato foliage. In their early stages, these diseases produce small irregular-shaped spots that are dispersed and difficult to identify. In comparison, the late stage of the disease, with its distinctive spots and large spot areas, is easier to identify. Still, it is discovered late, resulting in a significant loss of tomato quality and yield. As

a result, disease detection technology for tomatoes is critical. However, traditional manual identification and knowledge base-based expert system methods are highly subjective and reliant on farmers and experts [2]. Lesions on tomato leaves vary in shape and are insufficiently characterized. Although certain diseases have distinctive spots in terms of shape and color, they are difficult to distinguish with the naked eye because of the small spot area and require magnifying equipment, such as magnifying glasses or microscopes for observation. Thus, detecting tomato leaf diseases quickly and accurately and implementing appropriate control measures are critical to ensuring tomato production.

Disease-based leaf recognition methods are a popular research direction in computer vision and image processing [3–5]. Numerous studies have successfully combined image processing and traditional machine learning techniques, resulting in significant application value [6, 7]. However, most disease recognition algorithms extract image features via multiple filters [8]. The extraction process is tedious and frequently selects for recognition objects with noticeable disease features and concentrated disease areas. As a result, this traditional recognition method cannot extract disease features from tomatoes. Along with the explosion in data volume and the advancement of computer hardware, deep learning has made significant strides in image recognition [9, 10]. Many researchers prefer convolutional neural networks because of their three primary features: local perception, multiple convolutional kernels, and parameter sharing. For instance, Bedi and Gole [11] proposed a novel hybrid model based on a convolutional autoencoder (CAE) network and convolutional neural network (CNN) for automatic plant disease detection. Reference [12] used a convolutional neural network to extract features from a large dataset containing 14,828 images of tomato leaves infected with nine diseases. They visualized the results, achieving a 99.18 percent accuracy rate. Abbas et al. [13] used GAN and transfer learning to identify and classify tomato plant disease, achieving an average classification accuracy of 99.35 percent. Su et al. [14] separately fed one-dimensional spectral and three-dimensional hyperspectral images of ripe strawberries into a ResNet classification network. Both inputs were more than 84 percent accurate in the ResNet classification network. However, the error persisted even with the prepared dataset, as the shooting environment and equipment constrained the image quality. To ensure successful recognition, the features of spots on tomato leaves are compared to those of typical pests and diseases. It frequently falls victim to locally optimal solutions and gradient disappearance, resulting in low recognition accuracy. As a result, the main problems of the study are as follows: (1) the images in the tomato leaf disease dataset are gathered from a variety of sources, including the internet and the demonstration base of Hunan Vegetable Research Institute, and they suffer from a complex background and uneven quality. When the network is fed the original images, it can extract features from the training set. Nonetheless, the network may extract blurred features from the original images (e.g., speckled objects, such as dust and dirt) as features, resulting in incorrect extraction. (2) Common tomato leaf diseases produce subtle differences in the appearance of leaves, such as spots and slight yellowing, complicating disease recognition. Additionally, the disease areas in the disease images are small and scattered, complicating feature extraction. (3) In cases where some disease features are not readily apparent, more subtle features must be extracted. Increasing the number of network layers in the model can enhance the recognition ability of the network. Nonetheless, the resulting issue is that the network will be more difficult to train. Given that the subject of this study is tomato leaf disease, a prolonged training period will quickly result in additional disease damage to tomatoes. As a result, the network requires high accuracy, few parameters, and rapid model convergence.

Buades et al. [15] proposed the nonlocal mean filtering algorithm to address the issue of tomato leaf disease images being susceptible to background interference and blurred features. Because of its novel comparison of local similarity, it outperforms other traditional algorithms in terms of filtering effect and better preserves image edge details. Recent years have seen a surge in improvements based on this classic filtering algorithm. Dore and Cheriet [16] enhanced the denoising effect of the NL-Means algorithm by incorporating a robust regression with fixed smoothing parameters. It significantly reduces the blurring caused by weight. To address the issue of similarity accuracy degradation of this algorithm in the presence of harsh noise, Guo et al. [17] incorporated the feature similarity of the multichannel filter into the NL-means filter. The experimental results indicate that this filtering method outperforms the more traditional NL-means and wavelet-based filtering methods in terms of filtering effect. On the other hand, Kanoun et al. [18] proposed the KS-NLM filtering algorithm, which combines the NL-Means filter with anisotropic weighting to handle the central pixels of the patch better. The filtering algorithms above based on NL-Means perform better than NL-mean at denoising. Nonetheless, it does not satisfactorily address the high computational complexity and lengthy procedure of the NL-mean algorithm. The INLM filtering algorithm performs admirably well in terms of filtering. Its computational complexity is significantly lower than NL-Means, and its convergence speed is considerably faster, resulting in a shorter filtering time.

Das et al. [19] used a more complex network architecture to boost classification accuracy to 95.91%, resolving the issue of neural networks having difficulty identifying features associated with heart diseases. Brahimi et al. [12] demonstrated that increasing the number of layers in neural networks improves model performance, and nine tomato disease regions were identified as a result. However, increasing the number of network layers allows for more accurate features extraction. Nonetheless, there are two disadvantages: (1) when the number of layers in the neural network exceeds a certain threshold, gradient explosion and disappearance occur. These factors jeopardize crop disease identification. (2) To extract more detailed features, a deep neural network must be designed. However, the number of parameters to compute increases when training a deep neural network, resulting in slow convergence. Based on the aforementioned issues, this paper proposes DCCAM-MRNet, which utilizes ResNeXt50 as the backbone network and its unique residual mechanism to avoid gradient explosion and disappearance problems. In STAGE 1 of the network, dilated convolution is introduced, as well as a coordinate attention mechanism is inserted between each 3×3 and 1×1 convolution to improve feature extraction for subtle diseases.

To address the difficulty of training deep models, Luo et al. [20] incorporated the highway network into a bidirectional gated recurrent unit. The attention mechanism is additionally utilized in an effort to assign the weights of key issues in the network structure. Peng et al. [21] proposed ResNet, which uses residual shortcut connection to combine

the output of residual components with the input and uses the residual learning mechanism to solve the trainability problem of deep neural networks. Lu et al. [22] proposed a multistep linear structure based on the numerical solution of differential equations in 2017. They built it to examine a more efficient deep neural network called LM-ResNet, based on ResNet. The preceding three examples show how to find appropriate residual shortcut connections to guide the structural design of deep neural networks and how to set the appropriate initialization conditions for network weights and training parameters, which can help solve network trainability and model efficiency problems. The mixed residual connection approach proposed in this study is based on the Adams method of numerical solution of differential equations and comes in two flavors: RS-Block and TRS-Block. The DCCAM-MRNet network is built in such a way that it can continue to update the weights, achieve high learning accuracy, and make the network effective even when the magnitude of the gradient value update is minimal during the later phases of network training, thanks to ResNeXt's residual mechanism.

The contributions of this paper are as follows:

- (1) To reduce the impact of complex tomato planting background and fuzzy features of tomato leaf diseases on recognition accuracy, the INLM filtering algorithm is proposed in this paper. The INLM filtering algorithm reduces computational complexity after integrating the images, and it effectively overcomes the disadvantage of slow NL-Means computation. As shown in Figure 1, the quality of the images processed by the INLM filtering algorithm is improved compared with the original images. As shown in Table 1, the INLM filtering algorithm is 10 times better than the NL-means algorithm in filtering speed.
- (2) To extract the scattered and narrow disease spot features of tomato, the DCCAM-MRNet is proposed in this paper. (a) In STAGE 1 of the network, dilated convolution is used to identify the scattered diseases of tomato leaf to capture multiscale contextual information without changing the number of parameters. As shown in Table 2, the use of dilated convolution improves the ability to extract the feature and detection accuracy by 1.7% to ResNeXt50. (b) A coordinate attention mechanism, cochannel correlation, and remote dependence are introduced between 3×3 and 1×1 convolution for modeling, which enhanced the extraction of tomato micro-disease features and increased recognition accuracy by 3.6% to ResNeXt50 (as shown in Table 3). (c) The residual block (RS-Block) and transformation block (TRS-Block) of the mixed residual connection method are used in this paper to improve the trainability of the DCCAM-MRNet structure. The similarity is that adjacent residual blocks are weighted and added to the current residual block, resulting in a more accurate extraction of features between adjacent layers. The distinction is that the

TRS-Block utilizes the channel conversion function to match the input channel to the input channel. As shown in Tables 4 and 5, The recognition accuracy of DCCAM-MRNet is increased by 0.9% to ResNeXt50, and 0.11 M reduces its parameter count compared to the backbone network of ResNeXt50

- (3) Compared to conventional deep neural networks, the DCCAM-MRNet proposed in this paper accurately recognizes tomato leaf diseases (as shown in Table 6). Additionally, this network has fewer parameters than the backbone network ResNeXt50 and is more trainable.

As a result, this paper proposes a method for identifying tomato diseases that combines the INLM filtering algorithm and the DCCAM-MRNet. The identification principle is depicted in Figure 2. Firstly, the expanded dataset is passed through the INLM filtering algorithm, which reduces the influence of complex background and blurred features on the image, laying the groundwork for recognition and classification using the model. The processed dataset is then used to train and test the DCCAM-MRNet. To improve the ability of the model to extract features from the image set, the DCCAM-MRNet is enhanced with dilation convolution and coordinate attention mechanisms. The mixed residual connection method is used to increase the trainability of the network.

2. Materials and Methods

2.1. Data Acquisition. Datasets have been critical components of tomato leaf disease identification methods. The tomato leaf disease dataset in the Hunan Academy of Agricultural Sciences demonstration base was compiled using data from tomato greenhouse and the internet. We used a Nikon camera with a resolution of 4460×3740 in the tomato greenhouse. Leaf mold, Septoria leaf spot, yellow leaf curl virus, tomato mosaic virus, target spot, and two-spotted spider mite were included in the dataset. As illustrated in Table 7, these diseases cause irregular, colorful, scattered spots with indistinct margins, and disease features differ significantly between the early and late stages.

To alleviate the strain on the computer system caused by these images, each image in the tomato leaf disease dataset was compressed at resolution using the Matlab 2020b software. The compression specification for these images was 224×224 , and they were imported into the computer in the jpg format. To avoid model overfitting and poor generalization performance because of its small number of training samples, we expanded the dataset using MATLAB by flipping, cropping, scaling, highlighting the images, and saving them in jpg format.

The expanded tomato leaf disease dataset contains 10,923 images of tomato leaf disease. This experiment separated the dataset into the training set of 7646 tomato leaf disease images, the validation set of 2185 tomato leaf disease images, and the test set of 1092 tomato leaf disease images in a 7:2:1 ratio to train and test the network. Table 8 displays the comparison of

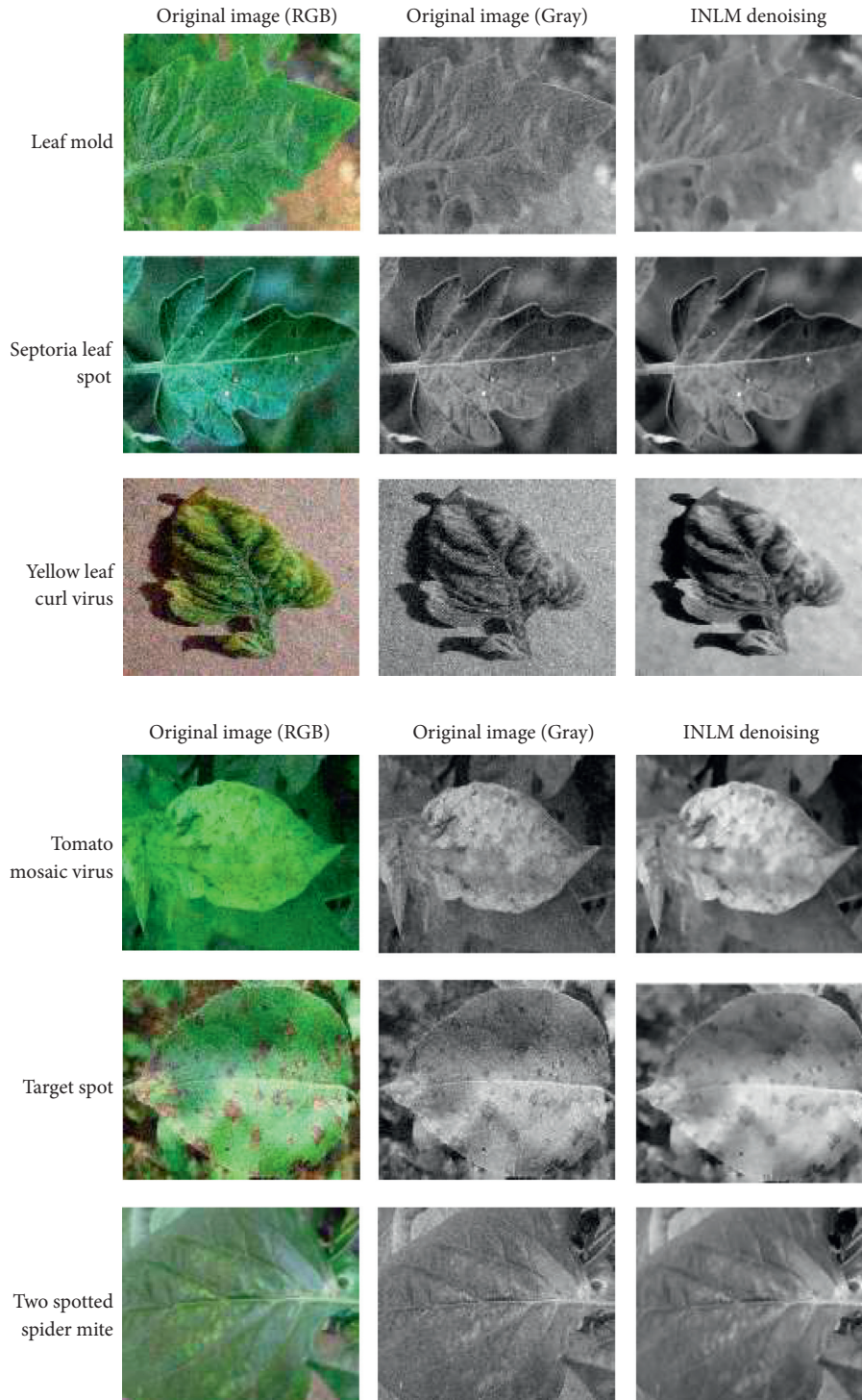


FIGURE 1: Comparison of original images and filtered images.

TABLE 1: Time spent by NL-Means and INLM filtering method.

Denoising algorithm	NL-means (s)	INLM (s)
Average time	45.63	4.32

TABLE 2: The accuracy of three networks.

Network model	Parameters (M)	Accuracy (%)
ResNeXt50	23.00	86.6
ResNeXt50-DC	23.00	88.3
DCCAM-MRNet	22.89	94.3

TABLE 3: The influence of coordinate attention on network accuracy.

Network model	Accuracy (%)
ResNeXt50	86.6
ResNeXt50-SE	87.5
ResNeXt50-CMBA	88.9
ResNeXt50-CA	90.2
DCCAM-MRNet	94.3

TABLE 4: Comparison of model parameters.

Network model	ResNeXt50 (M)	ResNeXt50-MRC (M)	ResNeXt50-LM (M)	ResNeXt50-CA (M)	DCCAM-MRNet (M)
Parameters	23.00	22.33	22.68	23.94	22.89

TABLE 5: Comparison of recognition accuracy and parameters of different networks.

Network model	Parameters (M)	Accuracy (%)
ResNeXt50	23.00	85.6
ResNeXt50-DC	23.00	88.3
ResNeXt50-CA	23.94	90.2
ResNeXt50-MRC	22.33	87.5
ResNeXt50-DC-CA	23.94	93.1
ResNeXt50-DC-MRC	22.33	89.6
ResNeXt50-CA-MRC	23.17	92.5
DCCAM-MRNet	22.89	94.3

TABLE 6: Evaluation indexes of the networks.

Network model	Recall (%)	F1-score (%)	Precision (%)	mAP (%)
MobileNet	78	74	77	71
ResNet50	83	81	80	74
ResNeXt50	87	85	83	77
LM-ResNet	86	86	87	82
InceptionResNetV2	84	80	85	80
EM-ERNet	82	83	85	81
B-ARNet	84	82	86	81
SENet	85	84	86	82
CMBA-ResNet	86	85	88	84
DCCAM-MRNet	94	93	94	90

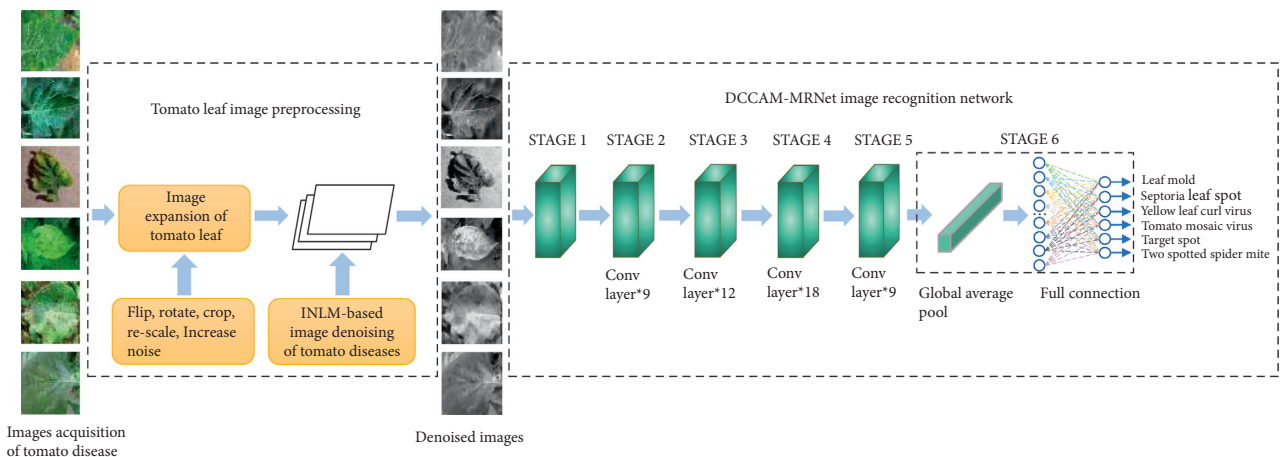


FIGURE 2: Principles of tomato disease identification.

TABLE 7: Symptoms and image sources of 6 tomato diseases.



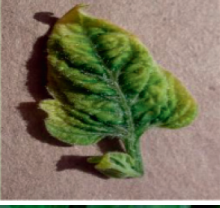



Disease type	Disease picture	Early symptoms of the disease	Advanced symptoms of the disease	Data sources
Leaf mold		Irregular or elliptical yellowish spots appear on the leaf blade, with indistinct margins of the spots.	The disease spot breeds gray or black irregular-shaped mold layer.	Tomato greenhouse
Septoria leaf spot		Round or nearly round spots appear on the front and back of the leaf with dark brown margins and many small ink-colored grain spots scattered.	The leaves are covered with spots, and the leaves turn yellow, causing early abscission.	Internet
Yellow leaf curl virus		The upper leaves are slightly yellowed and irregularly spotted. Purple veins frequently appear on the abaxial leaf.	The upper leaves and new shoots show symptoms, with small and red opaque	Internet
Tomato mosaic virus		Unevenly mottled shades of green, the leaves do not become smaller, and they do not produce deformities.	Leaf-blade shows yellow-green, flowering leaves are uneven.	Tomato greenhouse
Target spot		Subround, irregular brown spots on the leaf blade.	The color of the spot deepens, the area of the spot becomes larger, and it leads to leaf perforation.	Tomato greenhouse
Two-spotted spider mite		Many tiny greenish spots are scattered in the leaves' middle and lower parts.	The leaves fade to grayish-yellow and fall off.	Internet

TABLE 8: The recognition accuracy of the original dataset and the preprocessed tomato leaf disease dataset in the three models.

Network model	Original data set (%)	Preprocessed data set (%)
ResNeXt50	78.4	85.6
ResNeXt50-CA	84.7	90.2
DCCAM-MRNet	89.1	94.3

the accuracy before and after pre-processing. Table 9 shows the distribution of the six disease images in the dataset.

2.2. INLM Image Filtering. As shown in Table 1, the images of tomatoes taken in greenhouses have complex growing environments, while diseases such as leaf mold, yellow leaf

curl virus, and target spot exhibit blurred features. Both issues will inevitably introduce noise into the collected images, degrading their quality. As a result, it is worthwhile to investigate measures to eliminate noise while retaining the images' essential features. The INLM algorithm fully exploits redundant information in the images, significantly preserves the details and textures of the original images during

TABLE 9: Details of six tomato diseases.

Disease type	Size of the data set			Division of the data set		
	Original number	Expanded number	Percentage	Training set (70%)	Validation set (20%)	Test set (10%)
Leaf mold	465	1858	17.00	1300	372	186
Septoria leaf spot	436	1745	15.98	1222	349	174
Yellow leaf curl virus	490	1961	17.95	1373	392	196
Tomato mosaic virus	448	1790	16.40	1253	358	179
Target spot	457	1827	16.73	1279	365	183
Two-spotted spider mite	435	1741	15.94	1219	348	174

filtering, and accelerates the filtering process by integrating the images, thereby reducing the complexity of operations and filtering time. Because of its efficiency and simplicity, the INLM algorithm is used in this study.

2.2.1. NL-Means Filtering Algorithm. The core idea of the NL-Means algorithm is to take a rectangular window of each pixel point domain and calculate the weighted sum of the pixel values of all the pixel points within the window, with the weights obeying a Gaussian distribution. It is similar to Gaussian filtering, but unlike Gaussian filtering [23], NL-Means use the similarity between the domain block of the current filtered point and the domain blocks of other points in the rectangular window to calculate the weights, and the greater the similarity, the greater the weights.

For an image, suppose $v(x) = u(x) + n(x)$, where $v(x)$ is the observed image with noise, $u(x)$ is the real image without noise, and $n(x)$ is the noise perturbation of pixel x . A noisy image $v = \{v(x)|x \in I\}$ is given, and x represents the position of the pixel in the image, I represents the set of individual pixels in the image, and $v(x)$ represents the value corresponding to the position of the pixel x . The specific formula of the NL-Means filtering algorithm is as follows:

$$NL[v(x)] = \sum_{j \in I} w(x, y)v(y), \quad (1)$$

$NL[v(x)]$ represents the image filtered by the NL-Means filtering algorithm. $w(x, y)$ is the similarity between pixel x and pixel y , and its value is the Gaussian kernel of the pixel value between each point in a domain. The specific formula is as follows:

$$w(x, y) = \frac{1}{z(x)} e^{-\left(\|v(N_x) - v(N_y)\|_{2,a}^2 / h^2\right)}. \quad (2)$$

Among them,

$$Z(x) = \sum_y e^{-\left(\|v(N_x) - v(N_y)\|_{2,a}^2 / h^2\right)}, \quad (3)$$

$\|v(N_x) - v(N_y)\|_{2,a}^2$ represents the Gaussian kernel. $v(N_x)$ refers to a domain in the image centered on x . h is the attenuation factor. The smaller the value of h , the lesser the influence of the weighted point on the current point, and the edge is maintained well, however, the noise is serious. On the contrary, the edge is maintained poorly, however, the image is smoother, and the filtering level is high.

2.2.2. Tomato Disease Image Filtering Algorithm Based on INLM Algorithm. The size of the search area must be defined in the NL-Means algorithm, and the larger the search area, the greater the possibility of discovering similar pixels, however, the quantity of computation also increases exponentially. Assume the image is $N \times N$ pixels in size. With a color channel number of 3, a neighborhood window size of $k \times k$, and a search box size of $n \times n$, the complexity of the algorithm is $o(3N^2k^2n^2)$. Even in the original paper, the author defined the search area as a whole image, resulting in a few minutes of waste during the process of an image of 512×512 in size.

Based on the above description, it can be seen that the calculation complexity of the NL-Means algorithm is too high, and the program is very time-consuming, which is not conducive to practical application. After analyzing the formula, it was found that changing the similarity calculation between domains can reduce the time consumed.

If we first build an integral image for pixel differences, the equation is as follows:

$$S_t(x) = \sum_{\substack{Z_1 \leq x_1 \\ Z_2 \leq x_2}} S_t(Z), x(x_1, x_2), S_t(x) = \|u(x) - u(x+t)\|^2. \quad (4)$$

Using this method to calculate the distance between the two domains $v(N_x)$ and $v(N_y)$ takes only a small amount of time, and the calculation equation is as follows:

$$\begin{aligned} \|V(x) - V(y)\|^2 &= \frac{1}{d^2} (S_t(x_1 + ds, x_2 + ds) + S_t(x_1 - ds \\ &\quad - 1x_2 - ds - 1) - S_t(x_1 + ds, x_2 - ds - 1) \\ &\quad - S_t(x_1 - ds - 1, x_2 + ds)). \end{aligned} \quad (5)$$

When compared to the NL-Means approach, the overall complexity of the algorithm has been greatly lowered. At the same time, the offset is considered a cyclic determination condition to reduce space complexity. Rather than computing all of the integral pictures at once, each computation just has to get an integral image of the offset in one direction of the offset. We must, firstly, extend the image before filtering because each filter point in the original image requires a whole search window and

many field blocks. The search window is typically half the size of the neighborhood block plus half the size of the expansion.

The INLM filtering algorithm is used in tomato disease images, and its specific steps are as follows:

- Step 1: enter the tomato disease image to be filtered and convert it to a grayscale value
- Step 2: determine the domain window, search box size, and expand the image
- Step 3: take a point in the search block, and take the search block with the point y as the center and the search block with the first x in the image as the center to obtain $w(x, y)$
- Step 4: repeatedly take the next point y of the search block and repeat the c operation until the point of the search block is traversed
- Step 5: assign the maximum weight to point x , normalize the weight, and pass $NL[v(x)]$ to get the pixel value of the first point of the new image
- Step 6: take the second point x of the original image and repeat the c operation until the entire image is traversed
- Step 7: obtain an image of tomato disease after INLM filtering

As shown in Figure 1, the filtering results reveal that the images filtered using the INLM algorithm retain several features of the original image while lowering the noise. In 3.3.2, the testing results indicate that the INLM algorithm operates ten times faster than the NL-Means algorithm.

2.3. DCCAM-MRNet. Traditional convolutional neural networks cannot quickly identify diseased spots on tomato leaves because of their small size, lack of feature information, and relatively dispersed feature distribution. Deep neural networks must be utilized to extract more detailed features. As a result, the neural network we choose must have a sufficient number of layers to avoid the problem of gradient disappearance. Simultaneously, it must have the advantages of portability and quick training speed. In 2017, Pant et al. [24] proposed ResNeXt that incorporated the repetition strategy of ResNet and coupled it with the split-transform-merge strategy of the inception family. The residual error mechanism of ResNeXt can solve the problem of gradient disappearance, and when the number of parameters is the same, the recognition effect of ResNeXt is better than that of ResNet. All of the inception modules in the inception family have been meticulously designed [25]. Although the recognition result is satisfactory, several hyperparameters must be manually modified, and portability is lacking. Because of the topological structure of ResNeXt submodules, ResNeXt requires fewer manual modification parameters. After careful consideration, we proposed the DCCAM-MRNet, and Figure 3 depicts its network architecture. DCCAM-MRNet employs ResNeXt50 as its backbone network and replaces the original 7×7 convolution kernel with dilated convolution. It has a broader receptive field and improves extracting features without changing the parameters.

Between each 3×3 and 1×1 convolution, a coordinate attention mechanism is introduced. The coordinate channel attention mechanism can evaluate the relationship between channels and position information simultaneously and target the diseased area more precisely, giving it more high weight. Lastly, the DCCAM-MRNet is formed by the RS-Block and TRS-Block of the mixed residual connection method while retaining the RES-Block of the ResNeXt residuals. This combination makes the extraction of features between adjacent layers tighter and achieves higher learning accuracy while compressing the network.

2.3.1. ResNeXt. The main advantage of the ResNeXt is that it does not require deliberate construction of each portion of the network structure details, and it finishes complex classification jobs by the simple stacking of modules, which is relatively concise and easy to transplant. ResNeXt absorbs the advantages of group convolution [26]. The structure of ResNeXt is similar to that of the inception network [27, 28], which links contextual information spatially. It enhances network accuracy without increasing parameter complexity and minimizes the number of hyperparameters employed in the network. However, unlike the inception network, ResNeXt employs the same topology of parallel stacking, which is the same modules as ResNet, to extract features before merging the modules to limit the danger of overfitting.

2.3.2. Dilated Convolution. The disease features of tomatoes are not obvious, and disease spots are dispersed. As a result, expanding the receptive field is critical. The distinction between dilated and typical convolution is that dilated convolution introduces a new parameter known as expansion rate [29]. The receptive field is enlarged without affecting the size of the feature map by injecting holes into the ordinary convolution. We introduce dilated convolution on STAGE 1 of the DCCAM-MRNet to replace the original 7×7 convolution, as shown in Figure 4. Assuming that the size of dilated convolution kernel is $k \times k$ and the expansion rate is r , then the actual size of the convolution kernel is as follows:

$$K = r \times (k - 1) + 1. \quad (6)$$

After dilated convolution, the relationship between the size of the input and output feature maps is as follows:

$$W_2 = \frac{W_1 + 2p - r \times (k - 1) - 1}{s} + 1. \quad (7)$$

Among them, W_1 and W_2 represent the size of the input and output feature maps, respectively, the step-size is s , and p represents the patch.

2.3.3. Coordinate Attention. Attention mechanisms used in deep neural networks can provide good performance improvements [30, 31]. The SENet model builds a network model from the perspective of the correlation of feature channels, which enhances the directivity of the features

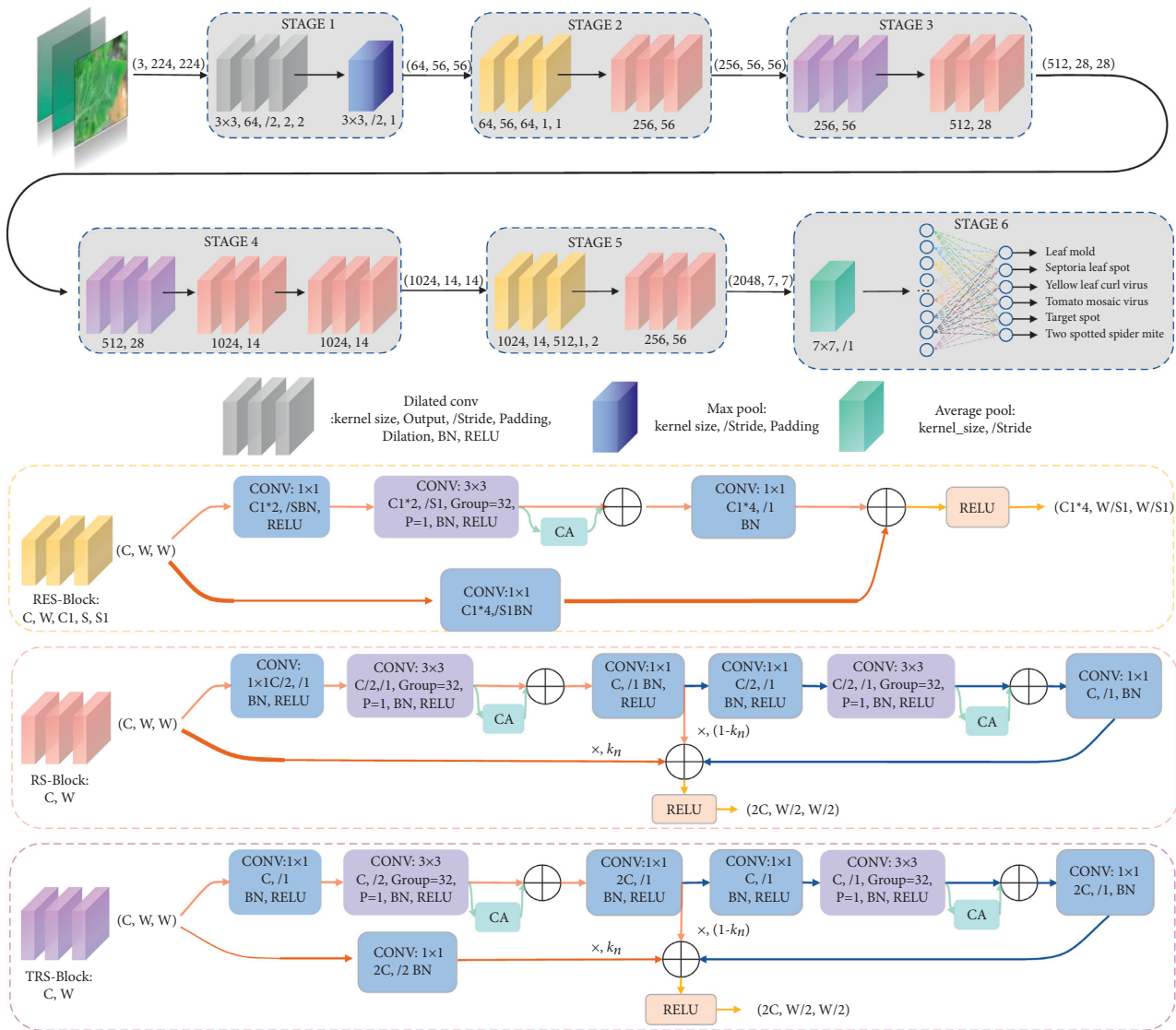


FIGURE 3: Architecture of the DCCAM-MRNet.

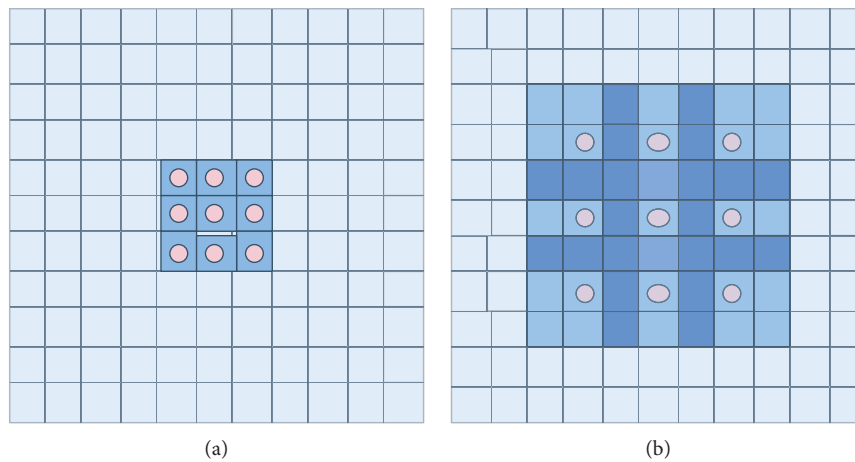


FIGURE 4: Dilated convolution with different rate. (a) Ordinary convolution ($r=1$). (b) Dilated convolution ($r=2$).

extracted by the convolution layer by strengthening the features of essential channels in feature mapping and weakening the features of unimportant channels. Wen et al. [32] embedded SENet into the ResNet-50 network [33], and on this basis, they identified five tomato diseases and achieved 89% detection accuracy. However, the limitation of SE is that only internal channel information is considered, and the importance of location information is ignored. Therefore, CBAM proposed by Woo et al. [34] tried to introduce location information by global pooling on the channel. Still, this method only captures local information and does not pay much attention to location information. To take account of the location relationship based on channel attention, Hou et al. [35] proposed coordinate attention, which is structured as shown in Figure 5. It decomposes channel attention into two feature coding processes, namely vertical and horizontal directions, integrating features with two spatial directions. With this processing, remote correlation can be captured in a spatial direction, while accurate location information can be maintained in another spatial direction.

The specific implementation method is as follows:

For a given input feature $X = [x_1, x_2, \dots, x_c] \in \mathbb{R}^{C \times H \times W}$, two spatial extents of pooling kernels $(H, 1)$ and $(1, W)$ are used to code channels along with the horizontal and vertical directions, respectively. The output of Channel C at height H can be formulated as follows:

$$z_C^h(h) = \frac{1}{W} \sum_{0 \leq i < w} x_C(h, i). \quad (8)$$

Similarly, the output of channel C with a width of W can be written as follows:

$$z_C^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_C(j, w). \quad (9)$$

After generating a pair of direction-aware feature maps, the concatenation connection operation is performed on the spatial dimension, and then the shared 1×1 convolution transformation function F_1 is used to get the following:

$$f = \delta(F_1([\![z^h, z^w]\!])), \quad (10)$$

$[\cdot, \cdot]$ denotes a concatenation operation along the spatial dimension, δ is a nonlinear activation function, and $f \in \mathbb{R}^{c/r \times (H+W)}$ is an intermediate feature mapping that encodes spatial information in the horizontal and vertical directions. Here, r is the reduction rate that controls the block size.

Splitting f into two independent tensors $f^h \in \mathbb{R}^{c/r \times H}$ and $f^w \in \mathbb{R}^{c/r \times W}$ along the spatial dimension, two 1×1 convolutions are used to transform f^h and f^w , respectively, so that they remain tensor with the same number of channels as the input X .

$$\begin{aligned} g^h &= \sigma(F_h(f^h)), \\ g^w &= \sigma(F_w(f^w)), \end{aligned} \quad (11)$$

σ is the sigmoid function. F_h and F_w are two 1×1 convolutions. g^h and g^w are the weights in two dimensions.

Finally, the weights g^h and g^w in the two dimensions are fused with the input X to obtain the output of the coordinate attention block Y , which is expressed as follows:

$$y_C(i, j) = x_C(i, j) \times g_C^h(i) \times g_C^w(j). \quad (12)$$

2.3.4. Mixed Residual Connection Method. The trainability of deep neural networks has always been a significant issue. RoyChowdhury et al. [36] were among the first to apply the mature numerical solution of differential dynamical systems to neural network learning. The mixed residual connection method, based on the Adams method of the numerical solution of differential equations, is used for network design in this paper. Figure 6 depicts two forms of the mixed residual connection method used in this paper: RS-Block and TRS-Block. This method enhances network performance and increases the tightness between adjacent layers for feature extraction by weighted summing the adjacent residual blocks with the current residual block.

In the DCCAM-MRNet, RS-Block is used in the latter part of STAGE 2, STAGE 3, STAGE 4, and STAGE 5. TRS-Block is used at the beginning of STAGE 3 and STAGE 4. The difference between them is that the input channel of TRS-Block is C and the output channel is $2C$. Hence, the number of channels is inconsistent and cannot be directly added. Therefore, the input channel needs to be convolved with 1×1 to change its channel number to $2C$.

Based on Adam's method, the specific steps are as follows:

$$h_{n+m} = h_{n+m-1} + \Delta h \sum_{i=0}^m \beta_i f_{n+i}. \quad (13)$$

In formula (13), Δh is the step-size. $h_t \in R^D$ is the output at time t . D represents the dimension of the output. β_i is the corresponding weight of f_{n+i} , and it satisfies the condition of $\sum_{i=0}^m \beta_i = 1$. f_{n+i} is the value entered in the layer of $n+i$.

In this paper, the mixed residual connection method sets $m = 0$ and $\beta_m = 0$ in formula (13). Then, formula (13) becomes as follows:

$$h_{n+2} = h_{n+1} + k_n f_n + (1 - k_n) f_{n+1}. \quad (14)$$

In (14), $k_n \in R$ is the weight coefficient corresponding to the information content in the hidden layer. Especially when $m = 1$, it is the Euler method in the differential numerical solution. In this paper, we let $k_n = 0.5$, which means that the importance of information in all hidden layers is the same.

3. Results

3.1. Experimental Environment. The hardware environment of this experiment is Windows (64bit) operating system, Intel Core i7-9700U CPU, and NVIDIA RTX 2080Ti GPU. The programming environment of the INLM filtering algorithm is MATLAB 2020b. The programming environment of the DCCAM-MRNet is Python 3.8.12, Pytorch 1.8.2, and

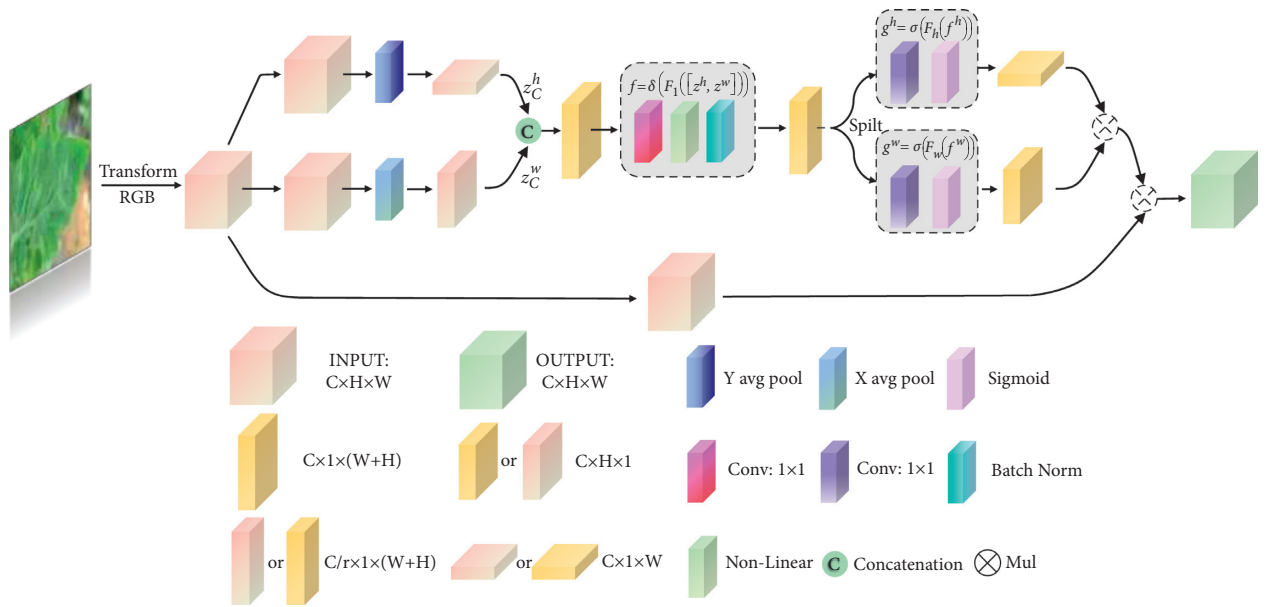


FIGURE 5: Coordinate attention structure.

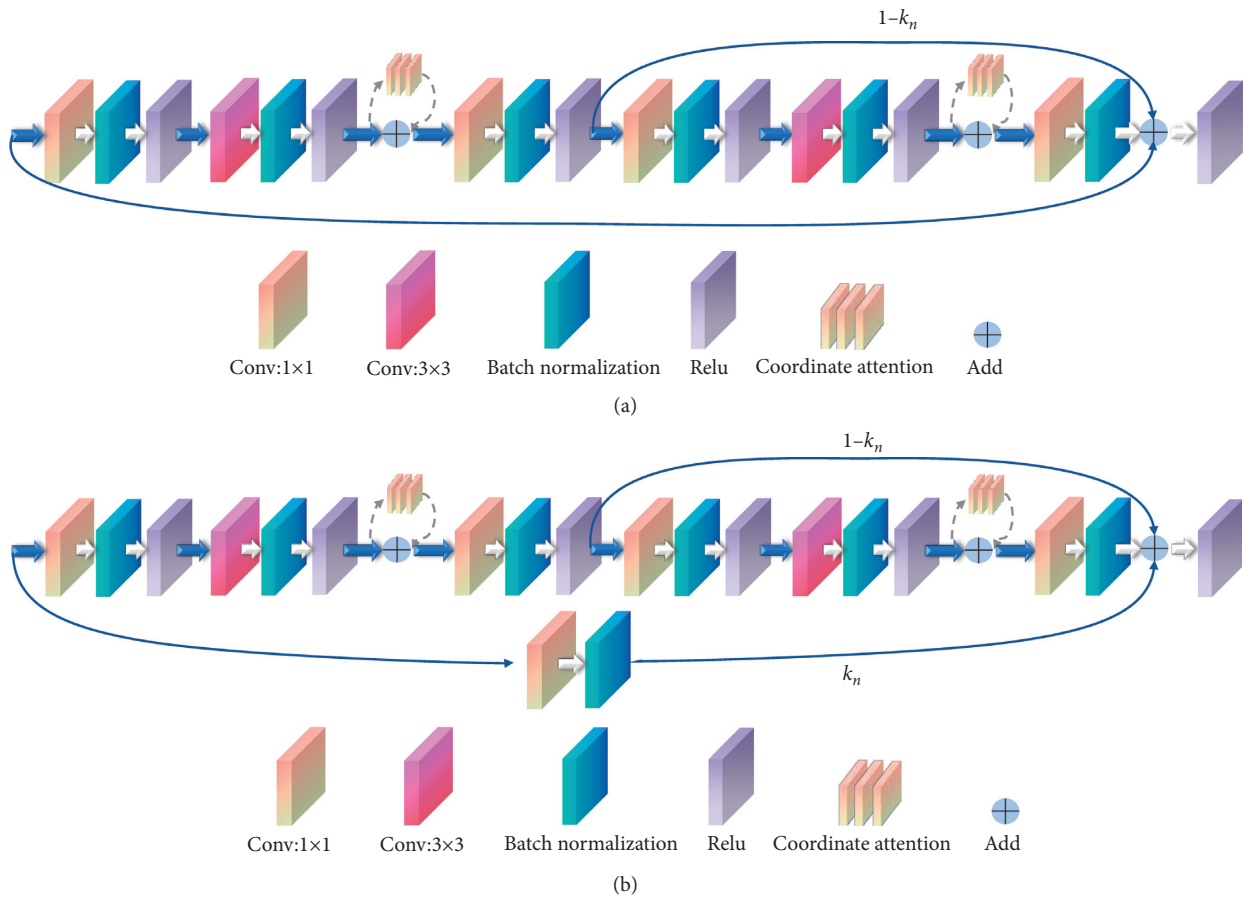


FIGURE 6: The method of mixed residual connection. (a) RS-Block. (b) TRS-Block.

CUDA 10.2. In this experiment, stochastic gradient descent was used to train the DCCAM-MRNet. The batch size of training samples was set to 32, and 8 for test samples. The learning rate lr was set to 10^{-3} , and the *epochs* was set to 140. The Adam optimizer was used during training, and the cross-entropy loss was used as the loss function.

3.2. Effectiveness Experiment of the Module

3.2.1. Effectiveness Experiment of Preprocessing. To test if the preprocessing of tomato leaf disease datasets can increase recognition accuracy, we fed the original dataset and the preprocessed data set, including the dataset expansion and filtering process, into ResNeXt50, ResNeXt50-CA, and DCCAM-MRNet, respectively, to conduct the experiments. Table 8 displays the recognition accuracy of the original dataset and the preprocessed dataset in three different types of networks. The results reveal that the three networks' recognition accuracy in the preprocessed dataset is greater than that in the original data set. It is because the data set is extended by cropping, flipping, zooming, and brightening, which increases the diversity of the dataset while avoiding the network coverage. The INLM filtering algorithm efficiently reduces the complicated background and removes fuzzy features, resulting in more apparent image features. As a result, following preprocessing, the accuracy of the dataset has increased in all three models.

3.2.2. Effectiveness Experiment of INLM Filtering Algorithm. To demonstrate that the INLM filtering algorithm has a faster convergence speed than the NL-means filtering algorithm, we randomly select 100 disease images from the dataset for filtering in MATLAB and calculate the average time spent by 100 images in the NL-means filtering algorithm and the INLM filtering algorithm. Table 1 shows that the convergence speed of INLM is ten times faster than that of NL-means.

3.2.3. Effectiveness Experiment of Dilated Convolution. In the DCCAM-MRNet, we used dilated convolution at STAGE 1. In the same test environment, we conducted experiments on ResNeXt50, ResNeXt50-Dilated Conv, and DCCAM-MRNet to validate their impact on classification performance. Table 2 demonstrated that utilizing dilated convolution in the ResNeXt50 could improve the accuracy of the network.

3.2.4. The Effectiveness Experiment of Coordinate Attention. To more intuitively understand the improvement in accuracy induced by coordinate attention, we trained and tested the preprocessed dataset using ResNeXt50, ResNeXt50-SE, ResNeXt50-CMBA, ResNeXt50-CA, and DCCAM-MRNet. Table 3 displays the accuracy of different networks on the test set. The experimental results showed that the three networks using the attention mechanism improved 0.9%, 2.3%, and 3.6%, respectively, in terms of accuracy compared to ResNeXt. The CA attention mechanism outperformed the

other attention mechanisms in terms of improving accuracy. The accuracy of the DCCAM-MRNet proposed in this paper is 94.3%, which indicates that the tomato leaf disease features are deeply extracted, and the network is effective in identifying.

3.2.5. Effectiveness Experiment of Mixed Residual Connection Method. We measured the number of parameters of ResNeXt50, ResNeXt50-MRC, ResNeXt50-LM, ResNeXt50-CA, and DCCAM-MRNet in terms of model compression. The results are displayed in Table 4, indicating that the number of parameters in the ResNeXt50-MRC is 0.35 M less than that 22.68 M of ResNeXt50-LM. The experiment result shows that MRC is superior to LM in model compression. The number of parameters in the DCCAM-MRNet is 0.11 M less than that of ResNeXt50.

3.3. Ablation Experiment. To thoroughly validate the effectiveness of the method proposed in this paper, we employed the same dataset and experimental environment in each experiment, only changing the components that needed to be compared. The backbone network in the ablation experiment is ResNeXt50, and the performance of several schemes is compared by adding one or more of the three methods of DC, CA, and MRC. Table 5 displays the comparing results.

Table 10 shows that the DCCAM-MRNet has higher accuracy than other networks, reaching 94.3 percent. When the coordinate attention is given to ResNeXt50, it enhances its accuracy by 3.6 percent when compared to the initial ResNeXt50. Similarly, the ResNeXt50 with Dilated Conv or Mixed Residual Connection outperforms the original ResNeXt50 by 1.7 percent or 0.9 percent, respectively. According to the evidence shown above, all three strategies are successful at increasing accuracy.

The number of parameters in the Dilated Conv network is the same as in the single variable network, which is consistent with the premise that Dilated Conv does not change the number of parameters. ResNeXt50 with mixed residual connection technique has 0.67 M fewer parameters than ResNeXt50 without mixed residual connection method, suggesting that the mixed residual connection approach aids in network compression.

3.4. Overall Evaluation of the DCCAM-MRNet. In the same test scenario, the DCCAM-MRNet outperforms its backbone network ResNeXt50 in terms of learning stability in the learning process and recognition accuracy. Figure 7 depicts the performance of DCCAM-MRNet in each category. The numbers 0, 1, 2, 3, 4, and 5 in the confusion matrix represent the six diseases, namely the leaf mold, Septoria leaf spot, yellow leaf curl virus, tomato mosaic virus, target spot, and two-spotted spider mite, respectively. The test set contains 1092 images in total, however, only 1090 of them are used in this test and displayed in the confusion matrix. The number of accurately predicted images, 1028 in total, is shown in the diagonal of the confusion matrix. The overall recognition

TABLE 10: Performance evaluation of each disease.

Disease type	Recall (%)	F1-score (%)	Precision (%)
Leaf mold	99	98	97
Septoria leaf spot	97	98	93
Yellow leaf curl virus	82	87	96
Tomato mosaic virus	93	89	99
Target spot	96	97	84
Two-spotted spider mite	98	97	97

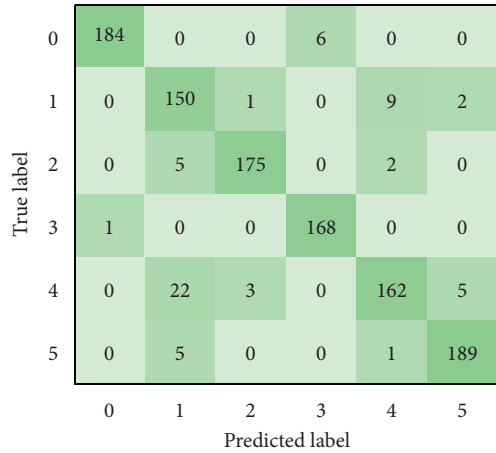


FIGURE 7: Confusion matrix of the DCCAM-MRNet.

accuracy of DCCAM-MRNet is 94.3%. Table 10 displays the accuracy of disease recognition in the DCCAM-MRNet for six different diseases. It can be seen that the highest recognition accuracy for the network is the tomato mosaic virus, reaching 99%, while only 84% for the target spot.

3.5. Comparison with Other Networks. We employ four indexes to evaluate the performance of the network: recall, F1-score, precision, and mAP. Table 6 displays the results. The indexes of DCCAM-MRNet surpass 90%, which is higher than those of other networks, showing that this network has more advantages and a more robust recognition effect for tomato leaf diseases than other networks.

3.6. Performance on the Plant Village Public Dataset. As experimental data, 1000 images of tomato leaf diseases from the plant village public dataset [37] are used. For recognition, the disease dataset is sent into the DCCAM-MRNet. Consequently, the recognition accuracy on the plant village public dataset is 97.0%, while the recall and F1-score are 97.6% and 97.1%, respectively. Figure 8 depicts the confusion matrix. The recognition accuracy for tomato mosaic virus using the public dataset is 92%, which is 8% higher than that obtained using our dataset. The recall and F1-score are also higher than those obtained using our dataset.

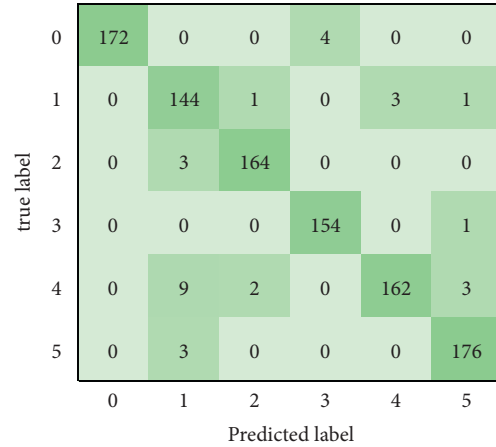


FIGURE 8: Confusion matrix of the DCCAM-MRNet.

4. Discussion

To identify tomato leaf diseases, we build a DCCAM-MRNet in this article. In experiment 3.7, we use the DCCAM-MRNet to identify the public dataset plant village, and the accuracy is 97.0 percent. When we use our dataset, we get an 8% better recognition accuracy for the tomato mosaic virus. The tests indicate that integrating the INLM and DCCAM-MRNet for tomato leaf disease identification is effective and capable of tackling the problem of low accuracy in tomato leaf disease identification to some extent, while more research is needed. (1) In this research, the DCCAM-MRNet is connected using the mixed residual connection method, with the weight coefficient of information amount in the hidden layer equal to 0.5. However, because the importance of information quantity varies in different hidden layers, it should be a floating number, and the mixed residual connection method should be adjusted to accommodate for floating, resulting in a superior network compression impact. (2) Current research focuses on identifying a single disease type on a single leaf, with less emphasis on identifying many illnesses on the same leaf, which has limits. Extracting traits and identifying mixed illnesses on tomato leaves will require more research. (3) This paper's data on leaf disease is insufficient. To improve the model's generalization capacity, the image data of tomato leaf diseases should be gradually added in the future.

5. Conclusions

For the complex tomato planting background, inconspicuous tomato leaf disease features, and distributed disease spots, a method for tomato leaf disease identification based on the INLM and the DCCAM-MRNet model is proposed. Firstly, a tomato leaf disease classification dataset with 10,923 tomato leaf images is generated. Secondly, the INLM filtering algorithm filters the tomato leaf disease dataset to reduce the influence of complex tomato planting background and blurred disease features on the images and improve image quality. The INLM filtering method is ten times faster than the traditional NL-Means filtering

algorithm in terms of filtering speed. Then, for tomato leaf diseases with obscure disease features and scattered spots, a novel neural network DCCAM-MRNet is developed using the ResNeXt50 as the backbone network. Dilated convolution and coordinate attention methods are used in the DCCAM-MRNet to improve the extraction of subtle and scattered feature points. The mixed residual connection method is used to enhance the tightness between adjacent layers for feature extraction, which reduces the number of network parameters and improves the learning accuracy of the network. The final experimental results demonstrate that the DCCAM-MRNet has an accuracy of 94.3% in identifying tomato leaf diseases. In addition, the number of parameters decreased by 0.11 M compared to the ResNeXt50 backbone network, which aids in network compression.

Tomato and other crop leaf disease recognition is still a hot research area in image recognition technology. The DCCAM-MRNet may be used for disease recognition after capturing tomato leaf pictures, which is critical for preventing and controlling tomato leaf diseases and ensuring tomato productivity and quality. The next step in this paper's research will be to see how the network can handle more types of tomato leaf illnesses and how to increase the network's disease recognition accuracy by enhancing the extraction of small, scattered data. In addition, we must investigate how to reduce the network's size to improve the detection of tomato leaf diseases and ensure agricultural productivity.

Data Availability

The data used to support the findings of this study are obtained from [37].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Yujian Liu contributed to methodology, prepared the original draft preparation, conceptualized the study, and performed data curation. Yaowen Hu was responsible for software, data acquisition, and investigation. Weiwei Cai provided model guidance. Guoxiong Zhou was involved in validation and project administration. Jialei Zhan performed formal analysis. Liujun Li visualized the study and reviewed and edited the manuscript.

Acknowledgments

The authors are grateful to all members of the Food College of Central South University of forestry and technology for their advice and assistance in the course of this research.

References

[1] M. A. Pervez, C. M. Ayub, H. A. Khan, M. A. Shahid, and I. Ashraf, "Effect of drought stress on growth, yield and seed

quality of tomato (*Lycopersicon esculentum* L.)," *Pakistan Journal of Agricultural Sciences*, vol. 46, no. 3, pp. 174–178, 2009.

- [2] M. Nouri, N. Gorretta, P. Vaysse et al., "Near infrared hyperspectral dataset of healthy and infected apple tree leaves images for the early detection of apple scab disease," *Data in Brief*, vol. 16, pp. 967–971, 2018.
- [3] S. Huang, G. Zhou, M. He, A. Chen, W. Zhang, and Y. Hu, "Detection of peach disease image based on asymptotic non-local means and PCNN-IPELM," *IEEE Access*, vol. 8, pp. 136421–136433, 2020.
- [4] J. Chaki, R. Parekh, and S. Bhattacharya, "Plant leaf recognition using texture and shape features with neural classifiers," *Pattern Recognition Letters*, vol. 58, pp. 61–68, 2015.
- [5] H. Lin, G. Zhou, A. Chen et al., "EM-ERNet for image-based banana disease recognition," *Journal of Food Measurement and Characterization*, vol. 15, no. 5, pp. 4696–4710, 2021.
- [6] W. Zhang, J. Hu, G. Zhou, and M. He, "Detection of apple defects based on the FCM-NPGA and a multivariate image analysis," *IEEE Access*, vol. 8, pp. 38833–38845, 2020.
- [7] A. Tan, G. Zhou, and M. He, "Surface defect identification of Citrus based on KF-2D-Renyi and ABC-SVM," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 9109–9136, 2021.
- [8] M. Mandal, P. K. Singh, M. F. Ijaz, J. Shafi, and R. Sarkar, "A tri-stage wrapper-filter feature selection framework for disease classification," *Sensors*, vol. 21, no. 16, p. 5571, 2021.
- [9] P. Xiao, G. Zhou, M. He, and A. Chen, "Apple image recognition based on deep and migration learning," *International Agricultural Engineering Journal*, vol. 27, no. 3, pp. 371–379, 2018.
- [10] D. Oppenheim, G. Shani, O. Erlich, and L. Tsrur, "Using deep learning for image-based potato tuber disease detection," *Phytopathology*, vol. 109, no. 6, pp. 1083–1087, 2019.
- [11] P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," *Artificial Intelligence in Agriculture*, vol. 5, pp. 90–101, 2021.
- [12] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299–315, 2017.
- [13] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, "Tomato plant disease detection using transfer learning with C-GAN synthetic images," *Computers and Electronics in Agriculture*, vol. 187, Article ID 106279, 2021.
- [14] Z. Su, C. Zhang, and T. Yan, "Application of hyperspectral imaging for maturity and soluble solids content determination of strawberry with deep learning approaches," *Frontiers of Plant Science*, p. 1897, 2021.
- [15] A. Buades, B. Coll, and J. -. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 60–65, San Diego, CA, USA, June 2005.
- [16] V. Dore and M. Cheriet, "Robust NL-means filter with optimal pixel-wise smoothing parameter for statistical image denoising," *IEEE Transactions on Signal Processing*, vol. 57, no. 5, pp. 1703–1716, 2009.
- [17] T. Guo, Q. Liu, and J. Luo, "Filter bank based nonlocal means for denoising magnetic resonance images," *Journal of Shanghai Jiao Tong University*, vol. 19, pp. 72–78, 2014.
- [18] B. Kanoun, M. Ambrosiano, and F. Baselice, "Anisotropic weighted KS-NLM filter for noise reduction in MRI," *IEEE Access*, vol. 8, pp. 184866–184884, 2020.

- [19] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7675–7680, 2009.
- [20] X. Luo, W. Zhou, W. Wang, Y. Zhu, and J. Deng, "Attention-based relation extraction with bidirectional gated recurrent unit and highway network in the analysis of geological data," *IEEE Access*, vol. 6, pp. 5705–5715, 2017.
- [21] S. Peng, H. Huang, W. Chen, L. Zhang, and W. Fang, "More trainable inception-ResNet for face recognition," *Neuro-computing*, vol. 411, pp. 9–19, 2020.
- [22] Y. Lu, A. Zhong, Q. Li, and B. Dong, "Beyond finite layer neural networks: bridging deep architectures and numerical differential equations," in *Proceedings of the International Conference on Machine Learning*, pp. 3276–3285, Atlanta GA USA, June 2018.
- [23] D. H. Shin, R. H. Park, S. Yang, and J. H. Jung, "Block-based noise estimation using adaptive Gaussian filtering," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 218–226, 2005.
- [24] G. Pant, D. P. Yadav, and A. Gaur, "ResNeXt convolution neural network topology-based deep learning model for identification and classification of *Pediastrum*," *Algal Research*, vol. 48, Article ID 101932, 2020.
- [25] S. Ahmed and S. H. Cho, "Hand gesture recognition using an IR-UWB radar with an inception module-based classifier," *Sensors*, vol. 20, no. 2, p. 564, 2020.
- [26] 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.
- [27] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4: inception-ResNet and the impact of residual connections on learning," *AI Magazine*, vol. 38, no. 3, pp. 72–82, 2017.
- [28] A. Kumthekar and G. R. Reddy, "An integrated deep learning framework of U-Net and inception module for cloud detection of remote sensing images," *Arabian Journal of Geosciences*, vol. 14, no. 18, pp. 1–13, 2021.
- [29] Y. Wang, G. Wang, C. Chen, and Z. Pan, "Multi-scale dilated convolution of convolutional neural network for image denoising," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19945–19960, 2019.
- [30] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2020.
- [31] 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, 2021.
- [32] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Computing & Applications*, vol. 32, no. 10, pp. 6111–6124, 2020.
- [33] X. Chen, G. Zhou, A. Chen, J. Yi, W. Zhang, and Y. Hu, "Identification of tomato leaf diseases based on combination of ABCK-BWTR and B-ARNet," *Computers and Electronics in Agriculture*, vol. 178, Article ID 105730, 2020.
- [34] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, Munich, Germany, September 2018.
- [35] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13713–13722, Nashville, TN, USA, June 2021.
- [36] P. RoyChowdhury, Y. P. Singh, and R. A. Chansarkar, "Dynamic tunneling technique for efficient training of multilayer perceptrons," *IEEE Transactions on Neural Networks*, vol. 10, no. 1, pp. 48–55, 1999.
- [37] A. Ali, "PlantVillage Dataset. Kaggle," 2018, <https://www.kaggle.com/abdallahalidev/plantvillage-dataset>.

Research Article

Customer Relationship Management Based on SPRINT Classification Algorithm under Data Mining Technology

Yazhou Sun¹ and Xueqing Tan² 

¹Department of Economic Management, Pingdingshan Polytechnic College, Pingdingshan 467000, Henan, China

²School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454003, Henan, China

Correspondence should be addressed to Xueqing Tan; tanxq@hpu.edu.cn

Received 23 February 2022; Revised 18 March 2022; Accepted 24 March 2022; Published 14 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Yazhou Sun and Xueqing Tan. 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 advance of computational intelligence, customer relationship management system based on data mining technology can not only bring more economic benefits to an enterprise but also improve the management and decision-making level of Chinese enterprises. In this paper, the application of data mining technology in customer relationship management (CRM) is analyzed, and four data mining modes are realized: customer classification, cross-marketing, customer acquisition, and customer retention. In the data mining module, SPRINT classification algorithm is used in customer classification. At the same time, FP-growth, an association rule algorithm without candidate set, is applied in cross-marketing, which enhances the practicability of the system. The algorithm of optimal customer retention strategy under digital intelligence technology is adopted in customer retention, which makes up for the shortcomings of traditional CRM system and helps enterprises to better operate and adjust marketing strategies.

1. Introduction

With the rapid development of technology, more and more enterprises find that products tend to be homogeneous in performance, price, appearance, and even advertising. In this environment, the value of customers has become more important, which will also directly affect the value of enterprises. Enterprises in China are facing more and more competition. In the early days, enterprises began to use databases to store customer information and analyze them by computer [1]. To achieve truly scientific customer relationship management, data mining technology is indispensable. The emerging data mining technology in recent years can provide technical support for customer relationship management, better maintain customer relationship, and give full play to the role of customer management system [2].

Customer relationship management system based on data mining technology has been applied to many industries, such as banks, large-scale retail industries, and e-commerce

companies [3]. Through the application of data mining technology in customer relationship management system, credit evaluation, customer retention, price setting, financial analysis, and marketing planning can be carried out in these industries, thereby reducing costs and increasing profits for enterprises. Customer relationship management system based on data mining technology can not only bring more economic benefits to an enterprise but also improve the management and decision-making level of Chinese enterprises.

2. Overview of Customer Information Management and Data Mining Technology

2.1. Customer Information Management

2.1.1. Concept and Origin. The idea of customer relationship management (CRM) comes from the marketing theory, management science, behavioral relationship research, and other disciplines of western enterprises. With the arrival of

the industrialized society, the productive forces have been greatly developed. After automation has gradually entered the industrial field, there are more types of products, and the demands of consumers become diversified. In addition, the rapid progress of technology has also broadened consumers' choices. The technology-oriented or product-oriented seller's market in the past gradually became a customer-oriented buyer's market. In order to cope with the new changes, maximize the profits of enterprises, and occupy more markets, the marketing methods of enterprises have also changed. Especially after entering the postindustrial era, knowledge will become the dominant force [4]. With the development of modern information technology, enterprises can survive in the fierce market competition by capturing customers' needs through technical means combined with traditional theories and providing targeted products and services to meet customers' real needs [5]. Therefore, the theory and practice of CRM have been greatly developed.

Customer relationship management can be divided into three phrases: customer, relationship, and management.

- (1) *Customers.* The customers here do not just refer to "people" who use products or services in a narrow sense. They can be an organization, a person, or a group. Nor is it just the user of the final product or service. Generalized customers can be any organization or individual, as long as it is related to the operation of the enterprise. For example, employees in an enterprise are internal customers of the enterprise, and the department of the next process is the customer of the last process. Externally, customers can also be dealers, trade associations, and so on.
- (2) *Relationship.* Relationship refers to the behavior and feeling state between enterprises and customers. Behavior mode is the performance of the degree of relationship in behavior, such as repeated purchase or reduced purchase frequency. Feeling refers to the expression of the degree of relationship in attitude, which can include preference or aversion, spreading good, or bad word of mouth [6]. In customer relationship management, relationship is the link between enterprise customers and the core purpose of the whole customer relationship management.
- (3) *Management.* Management is more of a means. Although customer relationship management was put forward only in the 1980s, its customer-centered concept has been widely used in business activities for a long time. With the advent of the computer era, customer relationship management can be realized more easily. Through various modern tools, the resources of enterprises can be conveniently organized. Adjust, organize, or allocate resources according to the state of customer relationship [7].

2.1.2. Classification and Composition. Customer relationship management (CRM) is a systematic concept, which includes CRM strategy, CRM system, and CRM marketing

strategy, all of which are indispensable. Its main goal is to optimize the management of customer data and expect more customization for the ever-changing customer demand [8]. Customers who request customer service through the website should be recognized and provide personalized responses to their requests. CRM and cloud computing are technologies aimed at centralized customer data management because it allows companies to respond effectively to customer requests [9]. At the same time, big data depends on using the massive data generated by CRM channels to improve knowledge. Therefore, CRM solutions will mainly involve business teams (e.g., marketing) as well as marketing teams, IT teams, or customer service teams (Figure 1). Therefore, the centralized information in CRM will be shared by all relevant teams to ensure the best customer relationship management.

According to the activities and nature of work, CRM systems are generally divided into three categories: operational CRM, cooperative CRM, and analytical CRM [10]. CRM in operation makes marketing automation, thus generating clues of contact in CRM and interdepartmental collaboration to share data to achieve strategic goals and objectives. However, analysts are responsible for enhancing business status according to decisions provided by customer data. Figure 2 shows the integration of operational and analytical CRM with data mining and business planning. In today's business world, new technologies enable organizations to more accurately target individual customers or market segments [11]. At the same time, advanced marketing technology enables people to pay more attention to the customer-centered point of view.

2.2. Data Mining Technology

2.2.1. Process of Data Mining. Data mining technology itself is the core of knowledge discovery. It needs to sort out hidden and valuable knowledge from a large amount of data. In the aspect of marketing decision-making, it is necessary to analyze the data of enterprises, discover the hidden rules, and then further model them. Generally speaking, data mining has four processes [12–14], problem definition, data sorting, selection of the model/algorithm, and analysis results, as shown in Figure 3.

The process of data mining is not one-way, but a cycle. After the results are formed, there may be errors or irrelevant parts in them. Irrelevant parts represent the abnormal data, which can't be used to the following algorithms, reselect the algorithm/model, or adjust the parameters to exclude irrelevant parts. If that result is found to deviate greatly from the target, it may even need to go back to the first step and reestablish goals. Therefore, data mining is a dynamic and cyclic process. Complete data mining is a process of continuous feedback.

2.2.2. General Algorithms. According to the different objectives of data mining or different data types, there are many algorithms currently used in the field of data mining [15–17]: Bayesian algorithm, decision tree, clustering algorithm,

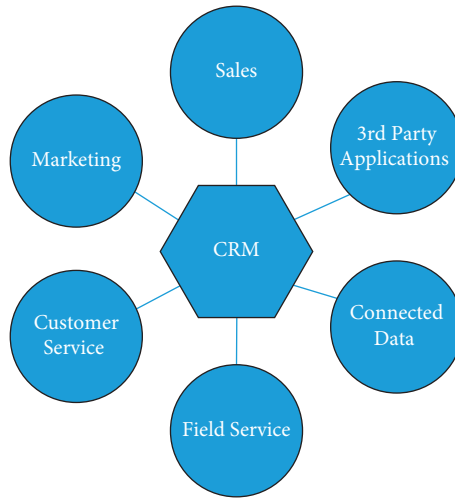


FIGURE 1: Organizational structure of CRM.

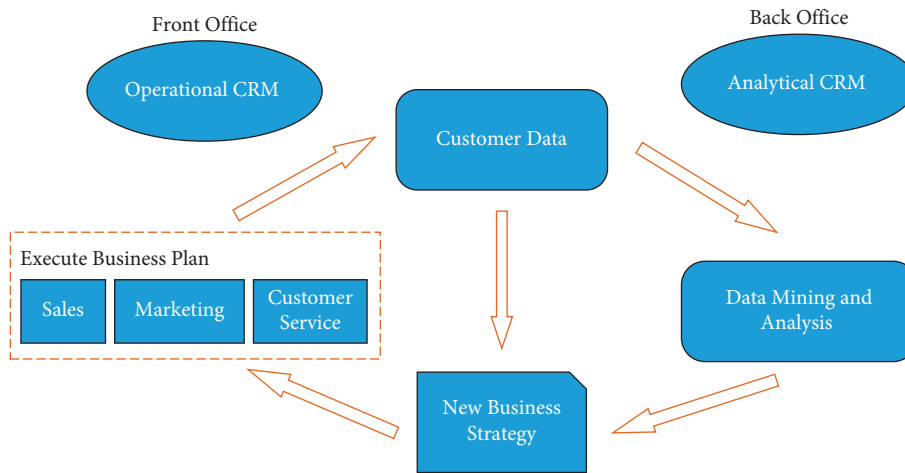


FIGURE 2: Relationship between different types of CRM.

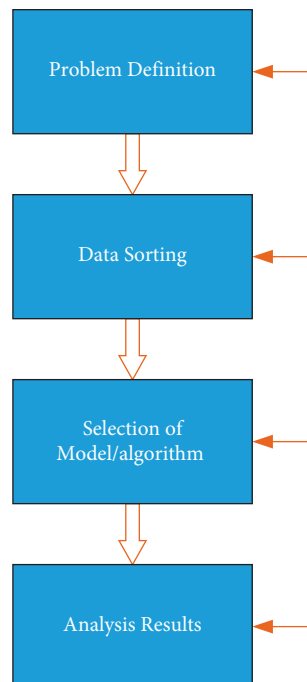


FIGURE 3: Data mining process.

association algorithm, neural network algorithm, regression algorithm, association rule algorithm, and so on:

- (1) Bayesian algorithm is an algorithm commonly used for classification prediction, which is based on Bayesian theory and uses statistics to classify samples. The Bayesian network is constructed by statistical data, and samples of unknown categories are classified and predicted according to the Bayesian network. Then, the possibility that samples belong to each classification is predicted and the most probable classification is selected [17]. Compared with other algorithms, Bayesian has the advantages of less computation and higher accuracy. Even the simplest Naive Bayes algorithm can get an excellent result which is especially suitable for the task of sample classification and prediction.
- (2) Decision tree is a regression algorithm, which uses recursion to establish a tree-like structure of classification rule. Firstly, the algorithm finds the input attribute that has the greatest influence on the target variable from the sample data and establishes the root node. According to the different values of input attributes, the sample data are divided into different subsets, and then the subsets are gradually divided according to the degree of influence between them and the input attributes until all attributes are included in the tree structure or the splitting is stopped because of insufficient subset samples. Finally, a tree structure is formed. Decision tree can be used to examine the influence of input attributes on target variables. Decision tree is commonly used to classify samples according to attributes. It can also be used to predict unknown classified samples [18]. The advantage of decision tree is that users do not need to have a deep understanding of the attributes of samples, and it can learn by itself and discover rules according to the sample data.
- (3) Clustering algorithm is also an algorithm for sample classification. Multidimensional space is established according to various attributes of samples, and classes are sorted according to the geometric distance between samples in the multidimensional space. Unlike other classification algorithms, clustering algorithms do not need to know the classes to be divided beforehand, and the formation of the class is completely generated automatically. The advantage of the clustering algorithm is that it hardly needs any prior knowledge, which is an unsupervised algorithm.
- (4) Neural network algorithm is an algorithm that imitates the characteristics of connections in brain neuron and is often used for classification. Generally speaking, the neural network has three levels: input, optional hiding, and output. Each neuron receives one or more inputs and then produces one or more identical outputs according to a simple nonlinear function. The sample is input into the hidden layer

from the input layer and finally to the output layer. Neurons at the same level are not connected [19]. The advantages of the neural network are the ability to self-learn; realization of the association function [20]; multichannel parallel computing, which can be used for particularly complex problems.

- (5) The commonly used linear regression algorithm is to find the linear function between dependent variable and independent variable according to statistical data. Regression algorithm can find the correlation between independent variables and dependent variables very directly [21], which is often used to forecast marketing, and so on. Time series algorithm is also a kind of regression algorithm, which is to find the relationship between dependent variables, time series, and other possible independent variables, which is often used to predict marketing.
- (6) Association rule algorithm is used to find the connection or correlation between different sets in data samples. For example, if the customer has a higher probability of buying commodity B after purchasing commodity A, then there is a certain degree of connection between commodity A and commodity B. Association rules allow users to analyze the behavior of customers, which is beneficial to formulate marketing strategies.

2.3. Data Mining-Based CMR. Data mining technology has been widely used in all aspects of customer relationship management, such as analyzing the factors that affect customer satisfaction, subdividing customer market, predicting customer behavior, predicting marketing trends, and cross-marketing. Data mining has also brought CMR to a new level.

2.3.1. Architecture Mode

(i) Centralized mode

The earliest centralized system consists of main-frame and many computer terminals.

(ii) C/S mode

That is, the client-server model consists of two-tier architecture, in which the server is responsible for data processing and users can obtain graphical interfaces. Servers generally use high-configuration personal computers, workstations, or minicomputers and use large database systems, such as Oracle and SQL Server. Clients need to install dedicated software [22].

(iii) B/S mode

That is, the browser server mode is an improvement of the C/S architecture where the client only needs a single computer with Internet access. B/S mode is divided into two types [23]. One is that customers need to download special controls and then operate all programs through browsers, which are usually

used on intranet [24]. There is also a control that does not need to be downloaded and completely run on the Internet and intranet [25].

Based on the framework above, close study and management of customer relationships and their interactions will help to identify, attract, and retain effective customers in this field. In the next stage of data preparation or preprocessing, data is prepared for further establishment and evaluation through cleaning, attribute selection, data conversion, and other processes. The model built in the CRM framework is an important step to establish an effective model to meet business needs. These models help to predict customer behavior and evaluate and visualize the effectiveness of measurement models to improve their performance which is shown in Figure 4.

2.3.2. Application Status. At present, in the process of marketing goods and services, enterprises cannot continue to treat them equally as before. Because different types of customers have different preferences, they must provide different marketing strategies to improve customers' satisfaction and loyalty and finally achieve the purpose of profitability by realizing customer value:

- (1) Customer classification
 - (i) Customer segmentation is the basis of targeted marketing strategy. The category of the customer will be expressed by the customer's own attributes and the customer's purchasing behavior pattern. After mastering a certain amount of customer data, managers can classify existing customers by analyzing their behavior patterns and attributes [26]. Although the loss of customers has caused losses to enterprises, the loss of customer data is also a very valuable asset for enterprises. Through data mining, customers who are easy to lose can be classified. Different types of plans for customer retention can be introduced for different types of easy-to-drain customers, thereby reducing the customer churn rate of enterprises.
- (2) Forecast of marketing amount

By mining the historical data of marketing, the regular behavior of sales can be obtained. For example, by introducing time series, the trend of marketing or whether there is a seasonal change rule can be discovered. Through predictive marketing, managers can better serve customers by adjusting inventory and production capacity, preparing raw materials, and reducing delivery time. Regression analysis can also be used in the management of products' life cycle. With more accurate help to enterprises to determine where products are located and through product strategies of different life cycles, customer loyalty can be improved.
- (3) Customer churn was found

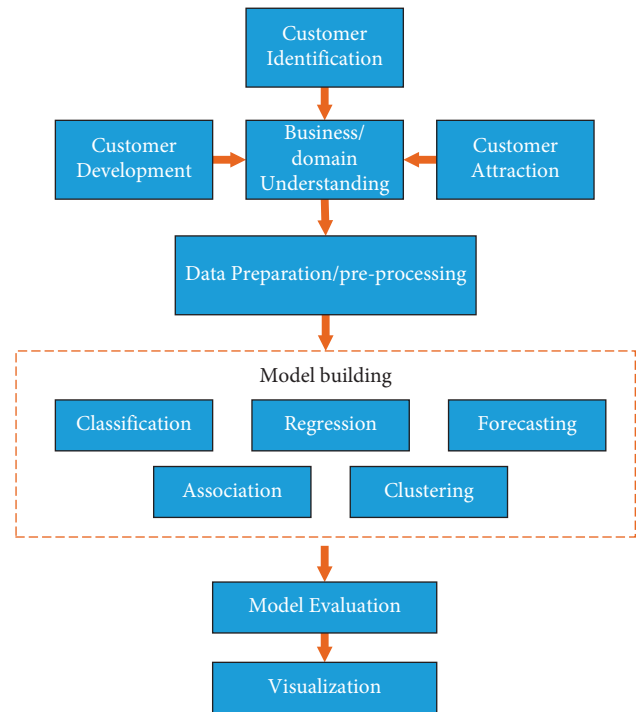


FIGURE 4: Architecture of the CMR system based on data mining.

Real customer churn generally does not show clear notice from customers. It is just that customers gradually disappear and do not come back to continue purchasing [27]. However, every purchase of customers will leave its mark of characteristics. Therefore, by analyzing these marks, managers can get the rules of customer churn and find out which customers are going to be lost so that they can modify customer strategies and retain valuable customers as much as possible.

3. Design of the CMR System in Marketing Strategy Based on Data Mining

3.1. Module Design. In this paper, the CMR system in marketing strategy consists of the following three subsystems [28]: business operation system, customer cooperation system, and data analysis system. Among them, the customer operation subsystem and data analysis subsystem are the most important subsystems, including customer information management and information analysis and processing:

- (1) Business operation subsystem

Mainly with the help of computer technology, it manages all aspects of marketing, sales, and service. The operating system is also known as an invoicing system which can enable enterprises to adopt better methods to achieve optimal results.
- (2) Customer cooperation subsystem

This subsystem includes customer information entry, customer information processing, code management,

customer management, supplier management, department management, employee management, product management, and partner management. It mainly manages the interaction between enterprises and customers, including e-mail, customer service center, call center, and electronic community. Putting these together, it means all channels for enterprises interact with customers.

(3) Data analysis subsystem

By processing and analyzing all kinds of data, using data mining technology to realize customer relationship management, managers can get meaningful information from it. Data information obtained by different ways, such as customer cooperation system and business operating system, should be sorted and summarized so as to help enterprises understand the classification, satisfaction, demand information, and other pieces of useful information of customers.

4. Realization of the CMR System in the Marketing Strategy Based on Data Mining

4.1. Overall Architecture. The CRM system of shopping malls based on data mining adopts a modular design. Considering the need for modular development and maintenance, the system adopts J2EE architecture and B/S system architecture based on the browser. Its development tool is JBuilderX, the back-end database is Oracle9i, and the application server is implemented by WebLogic8.1.x. The overall architecture is shown in Figure 5.

Data mining-based CRM system in marketing strategy adopts EJB container for data mining. EJB is a stateful session bean, which mainly includes the following six aspects: data selection, customer classification, cross-selling, customer acquisition, customer retention, and result output. The design scheme is as follows:

(1) Data input ()

(i) In the whole database, make a reasonable judgment on the data submitted by the client to find out more complete and consistent data.

(2) Customer segmentation ()

(ii) According to the different attributes of each customer, customers are divided into different categories according to different classification standards.

(3) Cross buying ()

(iii) Through the buying behavior of customers in a certain period of time, the influence of one commodity on other commodities and whether it is suitable for bundling can be analyzed.

(4) Customer obtainer ()

By analyzing the customer's response to market activities, some attributes of potential customers can be found.

(5) Customer retention ()

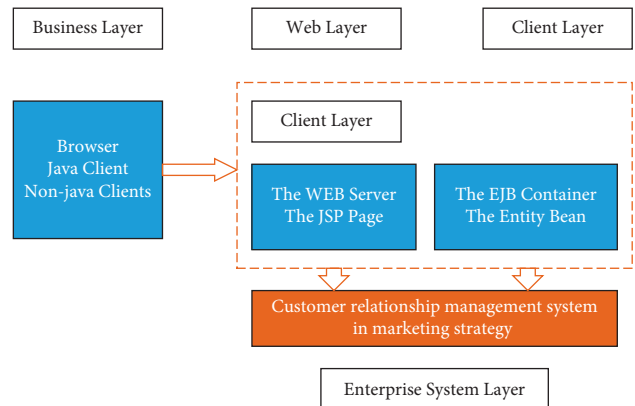


FIGURE 5: Overall architecture mode of CMR.

Establish a customer classification model with churn rate, analyze the churn of a customer, and determine how to keep the customer according to its attribute characteristics.

(6) Result output ()

Pass the results of data mining back to the client. Include customer information data, marketing campaign data, and customer transaction data. Customer EJB is used as the entity bean to describe the customer information of the system. Session Bean--DataminingEJB EJB is used to wrap entity Bean. CustomerEJB client calls entity beans by interacting with session beans.

The code that defines the remote interface is as follows:

```

/*Dataminigjava*/
import java.util.*
import javax.ejb.EJBObject.
import java.rmi.RemoteException.
public interface Datamining extends EJBObject.
{
public Boolean dataInput(String selectString) throws.
RemoteException.
public void customerSegmentation(String tablename)
throws.
RemoteException.
public void crossBuying(String tablename) throws.
RemoteException.
public void customerObtainer(String tablename)
throws.
RemoteException.
public void customerRetention(String tablename)
throws.
RemoteException.
public String resultOutput(String selectString) throws.
RemoteException.
}

```

The code for calling the entity Bean in the session Bean is as follows:

```

public Collection getAllData()
{
Vector vectUserInfos = new Vector();

```



```

Try.
{
Contextc tx = new InitialContext();
Object obj = ctx.lookup("CustomerHome");
CustomerHome customerHome=(CustomerHome).
PortableR-
emotcObject.narrow(obj, CustomerHome.class).
Collection collection = customerHome.findAll();
Return collection;
}
catch (Exception e).
{
e.printStackTrace();
}
}

```

4.2. Algorithm Design. Based on the above research conclusions, the decision tree algorithm is adopted for customer classification. Association rule algorithm is adopted for cross-selling. Classification algorithm is adopted for customer acquisition; The best customer retention strategy algorithm is adopted for customer retention. The algorithm design of these data mining technologies is analyzed in the following.

4.2.1. Data Mining Algorithm for Customer Classification. In the formulation of marketing strategy, the classification of customer information is relatively simple, and SPRINT algorithm in decision tree algorithm is adopted here. The hash table used in it does not need to reside in memory, so it can handle larger training data sets and classify them more accurately. Enterprises can formulate one-to-one marketing strategies according to different types of customers. The specific algorithm process is given in Algorithm 1.

Establish different attribute lists according to each field in the customer information table, and presort them. For example, in the list of age attributes, it is sorted by age from large to small. Then find out the best splitting point of each node in the decision tree. First, the root node is created, then the candidate split points are selected as 15,23,29, and their Gini values are calculated. Select the smallest as the best splitting point of the attribute list. By calculation, the best split point is 29, which divides the age attribute list into two parts. After that, the decision tree is divided into two branches: the left branch is the data record whose age is less than 29, and the right branch is the data record whose age is greater than or equal to 29. Then a hash table is established according to the root node, recording which child node each data belongs to, respectively. Finally, continue to take the above method to calculate the best splitting points for the two branches created and stop when the data in the attribute list are of the same class or the number of data is very small.

4.2.2. Data Mining Algorithm for Cross-Selling. According to the customers' purchasing, establish the database of customers' transaction data, and then conduct data

mining on these data to find out which commodities are usually purchased together. Choose the transaction records of purchasing more than 5 kinds of goods in the mall; each transaction record includes the transaction number, purchase time, and purchased products. The transaction data of these customers are shown in Table 2.

Among the association rule algorithms, Apriori classical algorithm is an association rule algorithm based on mining sets in customer transaction database [30]. The core method is recursion, that is, the method based on frequency set. However, there are some shortcomings when Apriori classical algorithm is introduced into the cross-selling of CRM system. When analyzing the transaction data, the candidate set is always very large, which will seriously affect the efficiency and may make the system abnormal.

FP-growth algorithm is an improvement of the Apriori algorithm, and it is an association rule algorithm [31] that does not need candidate sets. Therefore, the FP-growth algorithm is selected to solve the problem of cross-selling. In this way, through the data mining technology, when making the publicity plan, the managers can distribute the publicity materials of related product at the same time or match products for customers according to these association rules.

4.2.3. Data Mining Algorithm for Customer Retention. The target variable includes customer attributes, service attributes, and customer consumption data. According to the customer classification model with churn rate, which classification each customer should belong to is determined.

According to the customer classification model with churn rate, the relevant managers can discover the customers that may be lost. If the possibility of customer churn is high, the company should adopt various promotional measures to improve customer loyalty so as to reduce customer churn and customer churn rate. To establish such a customer classification model, it is necessary to calculate the loss rate of each classification. Churn rate of classified customers is calculated by dividing the number of customers churned by the total number of classified customers.

The optimal customer retention strategy is to establish a simple decision tree of customer classification with churn rate. According to different customer attributes such as service type, gender, and credibility, each leaf node represents a customer category with different values. For example, the first leaf node represents a female customer, and the wastage rate is 0.1. The second leaf node represents male customers, and the wastage rate is 0.8. The third leaf node represents general customers, and the wastage rate is 0.2. The fourth leaf node represents loyal customers, and the customer churn rate is 0.5.

The algorithm of the best customer retention strategy can calculate the node that brings the highest net profit to the company. Finding the best node is the best strategy for customer retention. Its specific algorithm is given in Algorithm 4.

```

(i) {
(ii) Preprocess data set S, establish list C based on attributes in A, create root node R, and create junction
(iii) Point queue Q.
(iv) While (queue Q is not empty)
(v) {
(vi) Fetch node N from queue Q
(vii) If (records in node N's attribute list are of the same category or few)
(viii) then
(ix) {
(x) Mark node N as a leaf node.
(xi) The continue
(xii) }
(xiii) 33
(xiv) Each split point K in for N
(xv) {
(xvi) Create a class histogram
(xvii) The Gini index of the split point K is calculated to select the best K '.
(xviii) }
According to K ', node N is divided into two children, and N1 and N2 are put into the queue.
Divide the list where K ' is located, create a hash table according to the RIds of the list, and divide other attribute lists.
}
MDL algorithm is used to prune decision tree.
}

```

According to the theory of marketing, the purchasing power of customers is very important [29], which has a great influence on the formulation of marketing strategies by enterprises, so the target variable is customer purchasing power. Customer purchasing power is divided into three types: high purchasing power, average purchasing power, and low purchasing power. Customers are defined according to their age, gender, occupation, family status, and other attributes. A customer information table is arranged in order according to customer information, as shown in Table 1.

ALGORITHM 1: SPRINT algorithm

TABLE 1: Basic information of customers.

Id	Age	Gender	Income	Family status	Professional category	Purchasing power
ID902310	18	F	21378.2	General	1	Low
ID902311	26	M	12892.3	Good	3	High
ID902312	20	F	5656.2	General	1	Low
ID902313	24	F	7998.2	Good	3	High
ID902314	16	M	56569.6	Poor	2	High

TABLE 2: Customer transaction records.

Transaction number	Time	Product
0000001	20-6-18	F118,A003,C151,D027,G055,I328,M045,P147
0000002	20-6-18	F118,A150,B013,F051,F027,G055,H028,L025
0000003	20-6-18	A003,B003,F028,M102,G023
0000004	20-6-18	A003,B203,C151,F118,L122,M045,O057
0000005	20-6-18	B023,F118,H025,J015,O057
0000006	20-6-18	A003,B203,C151,F118,L122,M045,O057
0000007	20-6-18	F118,A150,B013,F051,F027,G055,H028,L025
.....		
00000500	20-6-18	F033,0018,B021,F006,L012,F145,E245

```

(i) Input:
(ii) A minimum support alpha, a transaction database DB
(iii) Output: FPtree
(iv) Procedure FPtree (DB, alpha)
(v) {
(vi) Create FPtree  $T$  and mark it as "NULL".
(vii) For the DB
(viii) {
(ix) For each  $I \in T_i$ 
(x) {
(xi) If  $I \in (S)$ 
(xii) Then joins the set  $S'$ 
(xiii) The else delete
(xiv) }
(xv) Rank the frequencies in  $S'$  according to the order in  $S$  - item for the  $[p | p]$ .
(xvi) Insert_tree ( $[p | p]$ ,  $T$ )
(xvii) }
(xviii) }
(xix) Procedure Insert_tree ( $[p | p]$ ,  $T$ )
(xx) {
    child  $N$  of For  $T$ 
    {
    If  $N.itemname = P.itemname$ 
    Then  $N.count++$ ; flag = true; Break;
    }
    If (! Flag)
    Then
    {
    Create a new child of  $T$   $N$   $ount = 1$ 
    Make  $n.node\_link$  point to a node with the same  $item\_name$  field.
    }
    If  $P$  is not empty
    Then Insert_tree ( $P$ ,  $N$ )
    }
}

```

ALGORITHM 2: FPtree based on database.

```

Input:
FPtree built above, a minimum support alpha
Output: Frequency set
Procedure FP-growth (FPtree, alpha)
{if FPtree has only one path  $P$ 
then
{
The frequency set produced by the combination of nodes  $\beta$  in  $P$   $\beta \cup \alpha$ 
Specify the frequency set support as the minimum support of the node in  $\beta$ .
}
The else
{for FPtree header table each entry  $c$  is processed from back to front
{generate a frequency set  $\beta = C \cup \alpha$ 
The support of the frequency set is that of entry  $C$ 
The conditional pattern library and conditional tree FPtree2 of  $\beta$  were established
If FPtree2 is not empty
Then FP-growth FPtree2, beta,
}
}
}

```

ALGORITHM 3: Generation of frequent sets.

```

Input: decision tree  $d\_tree$  with attrition rate. The revenue that customers bring to the company is  $P E$ 
The cost set  $S$  consumed by the movement between leaf nodes
Output: customer best leaf node
Procedure BestSelect (D-tree, C, P E)
{
  Int  $I=1, k$ ;
  Double  $d$ ; // Determine whether the customer belongs to the category  $k$ 
   $D = leaf\_node[k]$ .
  while( $i < n$ )
  {
    If  $d < leaf\_node [I]$ 
    Then continue.
    Else  $P[I] = P E * (d - leaf\_node [k].d)$  // Cost of moving between leaves
  }
  The maximum value of  $P[I]$  is found, and the optimal leaf node is  $J$ 
}

```

ALGORITHM 4: Optimal customer retention strategy algorithm.

5. Conclusion

Data mining technology has been applied in customer relationship management system, which can make enterprises better understand customers and make better business strategies, thus improving the quality of marketing decisions. In this paper, the technology of data mining with CRM in marketing strategy is combined, and data processing methods are provided. In addition, customer classification, cross-selling, customer acquisition, and customer retention are realized in the data mining module. The SPRINT classification algorithm is used in customer classification, which improves the accuracy of customer classification, thus making the relevant decisions of companies more credible. At the same time, FP-growth, an association rule algorithm without candidate set, is applied in cross-selling, which makes the system more practical. The algorithm of the optimal customer retention strategy is adopted in customer retention so as to help the companies to make decisions, thus better retaining customers and making the store get the maximum profit.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] X. Zhou, *Research on Classification Algorithm of Solar Wind Big Data Mining Based on Cloud Computing [D]*, Chengdu University of Technology, China, 2014.
- [2] D. Yao, *Research and Implementation of Hongta Group Database Marketing System Based on Data Mining [D]*, Zhejiang Sci-Tech University, China, 2014.
- [3] A. Berson, S. Smith, and K. Thearling, *Building Data Mining Applications for CRM*, McGraw-Hill, New York, NY, USA, 2000.
- [4] P. Cortez, "Data mining with neural networks and support vector machines using the R/rminer tool," in *Advances in Data Mining. Applications and Theoretical Aspects*, pp. 572–583, Springer, Germany, 2010.
- [5] X. Wang and Y. Liu, *Application of Data Mining in enterprise Customer Behavior Analysis [J]*, Journal of Information, 2004.
- [6] E. W. T. Ngai, "Customer relationship management research 1992-2002," *Marketing Intelligence & Planning*, vol. 23, no. 6, pp. 582–605, 2005.
- [7] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: a literature review and classification," *Expert Systems with Applications*, vol. 36, no. 2, 2602 pages, 2009.
- [8] Y. Liu, *Forecasting Analysis of Commodity Sales Based on Data Mining [J]*, Bulletin of Science and Technology, 2014.
- [9] R. Zhang, "On the role of Customer relationship management in enterprise marketing," *Science & Technology Economic Guide*, vol. 14, no. 31, p. 220, 2016.
- [10] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.
- [11] Xi Qian, *Research on the Application of Data Mining in Customer Satisfaction Management of Aviation Training Center [D]*, Beijing Jiaotong University, Beijing, China, 2013.
- [12] J. M. Bland and D. G. Altman, "Statistics notes: cronbach's alpha," *BMJ*, vol. 314, no. 7080, 572 pages, 1997.
- [13] R. Ling and D. Yen, "Customer relationship management: an analysis framework and implementation strategies," *Journal of Computer Information Systems*, vol. 41, pp. 82–97, 2001.
- [14] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: a survey," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 3–14, 2002.
- [15] M. J. A. Berry and G. S. Linoff, *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management*, Wiley, Indianapolis, 2004.
- [16] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems*, vol. 62, pp. 23–31, 2014.
- [17] S. Moro, R. Laureano, and P. Cortez, *Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology*, *Proceedings of the European Simulation and Modelling Conference*, pp. 117–121, Portugal, Portugal, 2011.

- [18] R. S. Swift, *Accelerating Customer Relationships: Using CRM and Relationship Technologies*, Prentice Hall PTR, N.J., 2001.
- [19] T. Munkata, *Fundamentals of New Artificial Intelligence*, Springer-Verlag, London, 2nd ed. edition, 2008.
- [20] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, NY, USA, 2nd ed. edition, 2010.
- [21] E. Turban, J. E. Aronson, T. P. Liang, and R. Sharda, *Decision Support and Business Intelligence Systems*, Pearson Educ, 8th ed. edition.
- [22] J. Zheng, *Analysis of the Role of Customer Relationship Management in enterprise Marketing [J]*, Shopping Mall Modernization, 2016.
- [23] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: a frequent-pattern tree approach," *Data Mining and Knowledge Discovery*, vol. 8, no. 1, pp. 53–87, 2004.
- [24] M. Mehta, R. Agrawal, and J. Rissanen, "MDL-based decision tree pruning," in *Proceedings of the 1995 Int. Conf. Knowledge Discovery and Data Mining(KDD'95)*, pp. 216–221, Montreal Canada, USA, August 1995.
- [25] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proceedings of the. 1996 Int. Conf. Very Large Data Bases (VLDB'94)*, Santiago, Chile, September.1994.
- [26] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [27] J. R. Quinlan, *C4.5 Programs for Machine Learning*, CAMorgan Kaufmann, SanMateo, 1993.
- [28] FriedmanN, GeigerD, and GoldszmidtM, "Bayesian network classifier[J]," *Machine Learning*, vol. 29, no. 1, p. 131163, 1997.
- [29] H. W. LiuB and MaY, "Integrating classification and association rulemining [A]. AgrawaR," . *Proc of the 4th Int. Conf. On Knowledge Discovery and Data Mining[C]*, AAAI Press, N Y, USA, 1998 8086.
- [30] M. Mehta, R. Agrawal, and J. Rissanen, "SLIQ: a fast scalable classifier for data mining," in *Proceedings of the SAVE Proceedings. 1996 Int. Conf. Extending Database Technology (EDBT'96)*, Avigon, France, March .1996.
- [31] J. Shafer, R. Agrawal, and M. Mehta, "SPRINT;A scalable parallel classifier for data mining," in *Proceedings of the. 1996 Int. Conf. Very Large Data Bases (VLDB'96)*, Bombay, India, 1996.

Research Article

Sentimental Analysis of Twitter Users from Turkish Content with Natural Language Processing

Cagla Balli ¹, Mehmet Serdar Guzel ¹, Erkan Bostanci ¹ and Alok Mishra ^{2,3}

¹Department of Computer Engineering, Ankara University, Ankara 06830, Turkey

²Faculty of Logistics, Molde University College-Specialized University in Logistics, Molde 6402, Norway

³Software Engineering Department, Atilim University, Ankara 06830, Turkey

Correspondence should be addressed to Alok Mishra; alok.mishra@himolde.no

Received 9 February 2022; Revised 14 March 2022; Accepted 17 March 2022; Published 13 April 2022

Academic Editor: Shengrong Gong

Copyright © 2022 Cagla Balli 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.

Artificial Intelligence has guided technological progress in recent years; it has shown significant development with increased academic studies on Machine Learning and the high demand for this field in the sector. In addition to the advancement of technology day by day, the pandemic, which has become a part of our lives since early 2020, has led to social media occupying a larger place in the lives of individuals. Therefore, social media posts have become an excellent data source for the field of sentiment analysis. The main contribution of this study is based on the Natural Language Processing method, which is one of the machine learning topics in the literature. Sentiment analysis classification is a solid example for machine learning tasks that belongs to human-machine interaction. It is essential to make the computer understand people emotional situation with classifiers. There are a limited number of Turkish language studies in the literature. Turkish language has different types of linguistic features from English. Since Turkish is an agglutinative language, it is challenging to make sentiment analysis with that language. This paper aims to perform sentiment analysis of several machine learning algorithms on Turkish language datasets that are collected from Twitter. In this research, besides using public dataset that belongs to Beyaz (2021) to get more general results, another dataset is created to understand the impact of the pandemic on people and to learn about public opinions. Therefore, a custom dataset, namely, SentimentSet (Balli 2021), was created, consisting of Turkish tweets that were filtered with words such as pandemic and corona by manually marking as positive, negative, or neutral. Besides, SentimentSet could be used in future researches as benchmark dataset. Results show classification accuracy of not only up to ~87% with test data from datasets of both datasets and trained models, but also up to ~84% with small “Sample Test Data” generated by the same methods as SentimentSet dataset. These research results contributed to indicating Turkish language specific sentiment analysis that is dependent on language specifications.

1. Introduction

Artificial intelligence is simply defined as machines trying to imitate human intelligence and behavior. Machine learning is computer models that learn big data and make inferences from them [1]. These models consist of a series of steps that are based on statistical algorithms, process big data, and make predictions with the mathematical results it draws from them [2]. These models can be iterated and used for different data, improved with different algorithms, or retrained with different hyperparameters to get better results. It can be reused without retraining with different test

data using the trained model. The goal is always to make a better guess and get better results.

Natural language processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence concerned with interactions between computer and human language, specifically how to program computers to process and analyze natural language data. Its aim is to obtain a computer that can “understand” the content, along with the idiosyncratic aspects of the natural language used in writing the texts. Afterward, it is possible to classify and edit the information in the text content and extract the information correctly [3].

Sentiment analysis [4] is a field that computer science and linguistics use together that aims to determine the sentiment contained in written data. In general, sentiment analysis algorithms are used to classify datasets by dividing them into different categories or classes [5]. For the studies in this field, working with machine learning algorithms, for these meanings extracted from the data, such as writing techniques, language tools, linguistic developments for different languages, interpretation of different meanings of the word in linguistics, and the change of the emotion expressed when words come together, has produced very successful results. Besides, there are text classification studies [6, 7] in the literature that also worked on text and documents; however, they are aimed to find useful information for business intelligence instead of emotions [8].

Today, one of the most common and most diverse data sources used for sentiment analysis is social media. Social media offer to be important data source [9] in current big data studies, where the data is constantly renewed instantly, information on current issues spreads rapidly between societies and people, and it is full of different perspectives on every subject such as politics, science, and history. Twitter is one of the best source websites and also a popular micro-blogging forum [10] for providing written data on current topics or big topics in social media.

Sentiment analysis is a very common topic in the literature and worldwide; many studies have been done on this subject. However, sentiment analysis has strong dependency on the linguistic features since it is based on the language of a text, and modelling is established by a text from the same language [11]. There are many successful studies done for texts in English language. As an alternative to the language dependency, there is a study that belongs to Denecke [12]; to translate a language to English before doing sentiment analysis regardless of language is used for text. However, in general, sentiment analysis studies are done individually per language such as [13, 14].

Machine learning algorithms [11, 15–31] were commonly used for Turkish sentiment analysis problems in previous studies. Bozyigit et al. [19] present a study that used LSTM and different CNN networks for Turkish sentiment analysis over Turkish user comments. SVM is used by Kaya et al. [15] in 2012 on a study of sentiment analysis for Turkish political news. In the study of Coban et al. [17] in 2015, Twitter data were used for sentiment analysis according to emojis using various machine learning algorithms such as SVM, Naive Bayes, Multinomial Naive Bayes, and KNN.

Although, in recent years, the number of Turkish studies has increased, there is still a need for more in the literature to provide variety. In this study, it is aimed to indicate accuracy of various machine learning algorithms on Turkish sentiment analysis with using different datasets and preprocessing steps over Twitter data. Unlike other similar studies, the study is conducted with not only a public dataset (Beyaz [32]), but also a custom dataset (SentimentSet [33]) that is created by using social media with custom topic, which is “pandemic.” This multicomination study could be useful in increasing the accuracy or validation of the research to compare results of the machine learning algorithms with

Turkish language sentiment analysis. Moreover, it will contribute with other Turkish studies that already exist and give insights for the next researchers about compatibilities of machine learning algorithms with the sentiment analysis in Turkish language.

In this study, which is conducted in Turkish, two datasets that consist of twitter data are labeled as positive, negative, and neutral, while the training models with these datasets neutral tweets are ignored. The marked (labeled) data was preprocessed with various libraries according to the language characteristics specific to Turkish. Afterwards, datasets are used to train Machine Learning models with various algorithms, and predictions were made on the test data with these models. The results are compared between different combinations of the datasets, algorithms, and different preprocessing libraries.

The remainder of this paper consists of the following parts. In Section 2, the literature review is briefly represented. Section 3 introduces the methodology that is used for the sentiment analysis models and datasets in detail. Section 4 explains the experimental studies, as well as the results and metrics of the different machine learning models per datasets. Section 5 compares the results of similar studies from the literature and gives authors insights about the study. Section 6 shows the conclusion and prospects.

2. Literature Review

Although the beginning of Artificial Intelligence dates back to very old times, the beginning of Natural Language Processing is a subject that dates back to ancient times and is now a subbranch of Artificial Intelligence in the field of Machine Learning. Natural Language Processing was first published by Alan Turing [34] in 1950, a seminal paper on Artificial Intelligence known as the Turing Test. Turing had set the machine’s task and intelligence criterion to be the automatic interpretation and generation of natural language. But it was not yet studied under Natural Language Processing at that time. Afterward, John Searle’s [35] paper titled *Minds, Brains, and Programs*, in which he put forward the Chinese Room Experiment, published in 1980, studied the imitation of NLP tasks by the computer when a set of rules, such as a Chinese learning guide, was given to the computer.

Due to the continuous increase in computational power and the emergence of Machine Learning algorithms, there has been a great development in the field of Natural Language Processing. Systems created with the use of these algorithms since the late 1980s were included under the heading Statistical NLP. In automatic speech recognition using statistical methods, they made a great impact with the article “A Maximum Likelihood Approach to Continuous Speech Recognition,” published by Bahl et al. [36] in 1983. According to the article, a maximum likelihood decoding formulation was created for the speech recognition task. In the study, a number of statistical models are explained for use in the speech recognition task. Another statistical study of the 1990s, *A Statistical Approach to Machine Translation* published by Brown et al. [37], is one of the important

researches in this field. In the article, a statistical translation approach from French to English is presented using Bayes' theorem.

Machine Learning algorithms, fast computers, and artificial intelligence networks that are developing day by day can be used in this field. Models created using Deep Neural Networks together with the developing Machine Learning algorithms have begun to become widespread and used in the field of NLP. The Natural Language Processor with Neural Networks (NALPRONN) model, developed by Martinez [38] in 1995 using artificial neural networks in the field of NLP, is one of the milestones in the field of Neural NLP. The NALPRONN system is a multimodal multilingual computer interface. It is a system of artificial neuron networks. This system has processing modules that include the backpropagation network such as I/O displays, input subsystem, output subsystem, dictionary subsystem, and monitor subsystem. There are also memory modules with feature mapping networks. According to the system, tasks are performed by modules. These modules are trained independently of each other with the same data. With NALPRONN, a generalized NLP system has been introduced using artificial neural networks.

In this study, NLP tasks are performed using supervised learning. For this reason, the study focuses on examining the studies conducted with supervised learning in the literature. One of the studies dealing with the sentiment analysis task of NLP using supervised learning is the article "Sentiment Analyzer: extracting sentiments about a given topic using natural language processing techniques" published by Yi et al. [39] in 2003. In the article, two subjects, namely, digital camera and music review, were defined while creating data. Documents collected under these two data groups are randomly selected documents from web pages collected from web scans. These documents were mixed and randomly placed under the subject headings. Two different labels were made for each subject. Subject-related documents are marked as (D+), while off-topic documents are marked as (D-). Various feature selection algorithms such as the Mixture model and Likelihood test were used with this data system. One of the studies carried out using supervised learning for text classification within the scope of the study is the comparison of SVM with kNN Decision Tree and Naive Bayesian methods by Liu et al. [40] in 2010. In the study, an SVM-based classification model is proposed. As a result of these experiments with other given algorithms, it was revealed that the F1 value of the SVM classifier exceeds 86.26%. Another sentiment analysis study using supervised learning is the article titled "A Sentiment Analysis Model For Hotel Reviews Based on Supervised Learning," published by Shi and Li [41] in 2011. By considering the hotel reviews of the users, it was tried to classify the emotions with a machine learning approach. SVM, one of the supervised learning algorithms, was used in the study.

There are several machine learning algorithms that can be applied to sentiment analysis. Besides, neural networks are also commonly used lately under the sentiment analysis topic. Yao and Guan [42], in 2020, proposed an advanced NLP method. This method was based on the LSTM

structure. In the study, compared to Basic and other LSTM, the improved method has better F1 score results in the Wall Street Journal dataset; it is concluded that the revealed method is more suitable for NLP when there are limited computing resources and a large amount of data.

Data collecting is a very important part of the sentiment analysis. Feldman [43] said that Twitter and Facebook are focal points of many sentiment analysis applications in 2013. Since then, Twitter became even the most important data source. Twitter provides user data as anonymous, which is eligible for the sentiment analysis. There are many sentiment analysis studies [44–46] that used Twitter in the literature with several algorithms. In the study on Twitter sentiment analysis conducted by Tam et al. [46] in 2021, an accuracy rate of 91.13% was obtained in the classification performed by using CNN and bidirectional LSTM (Bi-LSTM) models together on tweets in English. However, topic selection is the first step of the collecting data for sentiment analysis. COVID-19 has been a hot topic all over the world since its start. Hence, it took its place in the literature as well. There are COVID-19 related sentiment analysis studies [47–51] that are conducted with Twitter API.

Although there are many articles on sentiment analysis in the literature with COVID-19 topic or in general, there are not as much as studies in Turkish. However, the number of studies in the Turkish language has increased in recent years. In the study of Kaya et al. [15] in 2012 on Sentiment analysis on Turkish political news, political news from different sites of Turkish news were collected. Four algorithms within the scope of supervised learning were compared for emotion classification. These are Naive Bayes, Maximum Entropy, SVM, and the character-based N-Gram Language Model. From the empirical findings, it was observed that the Maximum Entropy and N-Gram Language Model outperformed SVM and Naive Bayes. By using different features, it has been demonstrated that all approaches reach 65% to 77% accuracy rates.

In the study conducted by Akba [16] in 2014, an F1 score of up to 83.9% was obtained over the models trained with film evaluations. Results were measured with the Information Gain and Chi-Square metrics. Zemberek was used as a preprocess while creating the data in the study, and SVM was used for the classification of the data. In the study of Coban et al. [17] in 2015, the accuracy of up to 66% was obtained by tagging tweets received on Twitter according to emojis and classifying them with various machine learning algorithms using two different feature extraction methods, Bag of words and N-gram model. In another study conducted by Karamollaoglu et al. [18] in 2018, sentiment analysis processes were applied to user comments collected from various websites using the Lexicon-Based method. The classification and sentiment analysis process were carried out with an average success rate of 80%.

Another study conducted in Turkish is the Turkish cyberbullying detection published by Bozyigit et al. [19] in 2019. Artificial neural networks were used in the study. Existing libraries were not used for Turkish Natural Language Processing. In the study, a list named "Harmful Terms" was created, and the wrong spellings between the

term and the input were tried to be corrected with the Levenshtein algorithm. For the text mining section, the TF-IDF method that is mentioned in this study was used. Two hidden layers are predicted for the Neural network. As a result of the study, an F1 score of 91% was obtained.

In the study carried out by Pervan [20] in 2019 with LSTM and different CNN networks, using the word2vec model on Turkish user comments collected from the websites, an accuracy value of up to 94% was obtained in LSTM.

One of the NLP studies on Turkish is a sentiment analysis study conducted with Machine Learning algorithms, published by Rumelli et al. [21] in 2019. During the study, open-source libraries such as Zemberek [22] prepared on Turkish Natural Language Processing were used. The data received from an e-commerce website are marked as negative, neutral, and positive according to the scores given by the users and trained with Machine Learning algorithms such as Naive Bayesian, Random Forest, and SVM. As a result of the study, a score of 73.8% was obtained.

A Corpus of Turkish Offensive Language on Social Media, published by Coltekin [23] in 2020, is a study to detect Turkish offensive language in social media. In the study, a dataset consisting of 19% of messages labeled as offensive language was used. A 77.3% F1 score was obtained with the linear SVM. Another Turkish study conducted in recent years is the sentiment analysis of people on global warming and climate change, conducted by Kirelli and Arslankaya [11] in 2020.

Although there are a limited number of Turkish language sentiment analysis studies, it increased in recent years [24, 25]. Aydogan and Kocaman [26] offered a new dataset since there are limited Turkish datasets to work on. Lately, some COVID-19 related studies [28–31] can be found in the literature.

The accuracy values of some of the sentiment analysis studies on Turkish in the literature are shown in Table 1.

According the literature review, our main contribution is providing a Turkish sentiment analysis, which is limited in number in the literature over COVID-19 topic. Besides, in our study, a pandemic topic based dataset was created as benchmark dataset to be used also by not only us, but also future researchers. Moreover, another public dataset that belongs to Beyaz [32] was also used for the sentiment analysis. There is no other study published in the corresponding literature using this public dataset. Finally, several machine learning algorithms such as SVM, Logistic Regression, and LSTM with applying different preprocessing techniques were used together, and the results were compared.

3. Methodology

Two different datasets were used for working with text data as a natural language in order to conduct sentiment analysis. The Public dataset has big amount of data that is already tagged and ready to use, which can provide more accurate results with variety of the data. This dataset also has no data that is relevant to the pandemic. It is compared with inhomogeneous SentimentSet dataset, which is created in the

TABLE 1: Accuracy table of given studies that are conducted in the Turkish language.

Study	Accuracy
Kaya et al. [15]	%77
Akba [16]	%83.9 F1 score
Coban et al. [17]	%66.06
Karamollaoglu et al. [18]	%80
Bozyigit et al. [19]	%91 F1 score
Pervan [20]	%94.21
Rumelli et al. [21]	%73.8
Coltekin [23]	%77.3 F1 score
Kirelli and Arslankaya [11]	%74.63
Shehu et al. [24]	%88.8
Hayran and Sert [27]	%80.05
Kabakus [28]	%97.895

scope of this study and has mostly negative data. These datasets consist of pretagged positive and negative tweets that are gathered from Twitter. The datasets are trained with various Machine Learning algorithms. The emotional states of social media users were tried to be classified as positive or negative using that trained models. The architecture of model generation that includes dataset preparation, training, and classification process is shown in Figure 1. After creating these models, the success rates of the algorithms used and the results obtained in this research were compared, and the accuracy rates were revealed.

The public dataset used within the scope of the study is an open dataset developed by Beyaz [32] in 2020 for a project on the detection of bullying in social media and was put into use as a public dataset that underwent various preprocesses. This dataset, which contains approximately 11,000 data, consists of Turkish tweets marked as positive and negative. It is seen that there are already some preprocesses in the dataset used, but in this study, the data was preprocessed again while creating the models. An example of the public dataset is shown in Figure 2.

In addition to the public dataset used for training the models in this study, the second dataset used is the SentimentSet developed within the scope of the study. The dataset consists of approximately 2600 Turkish tweets. Each tweet is marked as positive, negative, or neutral. Two methods were used while collecting the tweets that make up the dataset. The first of these methods is the use of the stream method of the Twitter API, and the second is the use of the open-source snsrape [52] library. Tweets are randomly selected without following any order. The first method used for collecting the tweets is the Stream method, which belongs to the Tweepy library and provides data in accordance with the given parameters. The dataset was created from real-time tweets selected by searching the words “aşı”, “aşılınmak”, “aşı olmak”, “vaccine”, “vaccinated”, “vaccination”, “stayhome”, “stay home”, “covid”, “corona”, “coronavirus”, “korona”, “covid-19”, “Covid19”, “Covid-19”, “Corona Virus”, “pandemic”, “pandemi”, “COVID-19”. Another parameter is “languages=“tr”,” which is used to extract only Turkish tweets. Received tweets were manually transferred to a table and manually marked there. In the second method used, when collecting tweets, they were drawn by searching for the

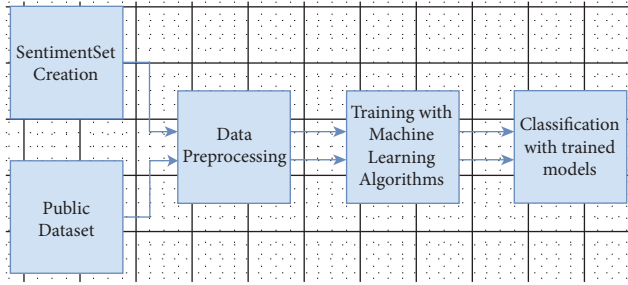


FIGURE 1: Model creation architecture general schematic.

sentiment	tweets
positive	uçur beni rüzgar
negative	sana küfür ederim

FIGURE 2: Sample data that belongs to the public dataset.

word “pandemic” from random months from March 2020 to March 2021 with the sncscrape library, and tweets collected at random time intervals were purified from those written as retweets, links, and replies.

The tweets created by the specified two methods were brought together. After collecting the tweets, they were subjected to various data preprocessing. First, collected tweets underwent the Noisy Data Cleaning process. The purpose of this process is to clear the data from unnecessary and nonsignificant data under sentiment analysis. Emojis links starting with HTTP, various symbols such as punctuation marks, and usernames starting with @ sign in tweets belonging to Twitter have been removed from tweets. After this process, the content of the tweet is completely converted to lowercase in order to edit the tweet. Afterward, the whitespaces at the beginning and end of the tweets were removed.

The next data preprocessing is the deletion of stopwords. These words, which show language-specific variability, may vary in studies conducted in the literature. Within the scope of this study, stopwords are taken from the Turkish language stopword list created by Son [53] by combining the LUCENE-559 Turkish stopwords list and the “Information Retrieval on Turkish Texts” list. An example for preprocessing a tweet is shown in Figure 3.

After combined tweets were preprocessed, they were marked manually, and SentimentSet was created. Tweets are marked as positive, negative, or neutral, which is shown in Figure 4. Tweets that express emotion but are not fully qualified and understood or that do not express positive or negative emotion are marked as neutral. Within the scope of the study, since the sentiment analysis was based on positive or negative classification, tweets marked as neutral were not processed.

Before using these two datasets to train the models under the study, various data preprocessing processes are done to improve the data quality for the algorithms. At this stage, the preprocessing steps of Zeyrek [54] were used.

In this preprocess, firstly, noisy data cleaning and pause words are removed. These two operations do not affect the

<i>Before</i>	AyLar sonra harika geçen bi haftasonu ..
<i>Data</i>	BoL boL sohbet muhabbet, bol kahkaha ..
<i>Preprocess</i>	Ahhh ulenn corona neLer çaldin bizden ... https://t.co/r18xUomkGk
<hr/>	
<i>After</i>	aylar sonra harika geçen haftasonu bol bol sohbet
<i>Preprocess</i>	muhabbet bol kahkaha ahhh ulenn corona neler çaldin

FIGURE 3: Data preprocessing example of cleaning noisy data and removing pause words.

SentimentSet already prepared in this way. However, the dataset prepared by Beyaz [32] was thus passed through these processes.

Afterward, data stemming/lemmatization was performed. This process was performed using two different libraries. One of them is Zemberek [22] library, and the other is the SnowBall library belonging to NLTK. Zemberek [22] is a library written in java. With the help of the library’s TurkishMorphology class, the lemmas of words were found with lemmatization. An example of word roots found with the help of Zemberek is shown in Figure 5.

The roots of the words were found by stemming using the class named TurkishStemmer belonging to the Snowball library of NLTK, another root-finding library used. Figure 6 shows the rooting process by stemming with the Snowball library. With these libraries, root finding is carried out using two different methods, lemmatization and stemming. While creating the models, both methods were tried, and the results were reported.

After this process, another data preprocess, Text Vectorization, was applied to the data with roots. Before starting this process, the data must be separated as training and test data. In all models created within the scope of the study, 20% of the data was reserved for testing. The remainder was used for training. After the training and test data were separated, the Text Vectorization process was performed. The purpose of this process is simply to translate the data in human language, which underwent various preprocessing steps, into a language that the machine can understand. The data obtained as a result of this process are given as input to machine learning algorithms.

In this study, the TF-IDF technique was used for Text Vectorization. Figure 7 shows how the TF-idf technique is used. Since tf-idf is a bag of words technique, the “ngram_range = (1, 2)” parameter indicates that the unigram and bigram approaches of the word bag method are used together. The “max_df = 0.9” parameter shown in the figure means ignores terms that appear in more than 90% of the documents. Likewise, the “min_df = 5” parameter means ignore terms that appear in fewer than five documents. As shown in the figure, after the vectorizer object is created with the specified parameters, the fit_transform method is called to scale according to the “x_train” data reserved for training and to learn these scaling parameters. The mean and variance of the features of the training set are learned. Then, “x_test” is scaled according to these learned parameters by calling the transform() method. [55]

sentiment	tweets
positive	pandemi bitse karakoçan spor fkmizin maçına gitsek
neutral	kanıtlanmayan sebep sorusu aşı konusuna girmeden ilk sorulacak soru
negative	abi korona maskeyi cikariyorsun tukuruyorsun sonra baska birisi gelip aliyor gecdim cocuksu hareket nefretinde bogul

FIGURE 4: Sample data that belongs to the SentimentSet.

```
Tweet before preprocess : teketekbilim fatih bey sebepden olmamız gereken doz
aşı temin edilemez farklı ülkenin aşısı olunabilir
Tweet after preprocess : UNK fatih bey UNK ol gerek doz aşı temin et fark ülke
aşı ol
```

FIGURE 5: Example of word roots found using the Zemberek library.

```
Tweet before preprocess: toplu gidilen özledim pandemi bitsin
kalabalığa giricem
Tweet after preprocess: toplu gidile özledi pande bit kalabalık
girice
```

FIGURE 6: Example of word roots found using the snowball library.

```
tfidf_vectorizer = TfidfVectorizer(ngram_range=(1,2), max_df=0.9, min_df=5)
x_train_tfidf = tfidf_vectorizer.fit_transform(x_train)
x_test_tfidf = tfidf_vectorizer.transform(x_test)
```

FIGURE 7: Tf-idf method representation used for Text Vectorization.

After the Text Vectorization process, the data was used in Machine Learning algorithms. However, these data preprocesses for LSTM are different from the others.

In LSTM, Tokenization preprocessing is performed instead of rooting (stemming or lemmatization) and text vectorization. For this process, the Tokenizer class of the Keras library is used. [56] The Tokenizer class and parameters used for LSTM are shown in Figure 8. The value of “num_words=2000” from the parameters shown in the figure determines how many words will be processed. Word separation operations were made according to the space with the given “split=“ ”” parameter. As seen in the figure, the dataset is given as a parameter to the fit_on_text method. Thus, the tokenizer has frequency information about the data. This method creates a word index based on frequency. Each word has its own integer value. The text_to_sequences in the figure replaces each word with the corresponding integer in the word_index dictionary. Pad_sequences in the figure is used for ensuring all sequences in a list have the same length. By default, this is done by adding 0 to the beginning of each sequence until each sequence has the same length as the longest sequence.

The data quality has been increased, and the data has been preprocessed and brought into a form that machine learning algorithms can use. Logistic Regression, SGD, Random Forest, Bayesian, SVM, and LSTM were used to train the models. The models created within the scope of the research are written in python language [57]. Google Colab was chosen as the working environment. The data prepared

```
max_features = 2000
tokenizer = Tokenizer(num_words=max_features, split=' ')
tokenizer.fit_on_texts(data['tweets'].values)
X = tokenizer.texts_to_sequences(data['tweets'].values)
X = pad_sequences(X)
```

FIGURE 8: Tokenizer method notation used for LSTM.

and preprocessed in the study carried out on Google Colab were used in many machine learning algorithms.

The Logistic Regression Model, which is one of the very common models in classification and regression problems, is used in the study. Model hyperparameters are used by default. The SGDClassifier model based on the Stochastic Gradient Descent algorithm is used in the study. This model’s hyperparameters are used by default, except for the “max_iter=5” parameter. The “max_iter” parameter, with a default value of 1000, indicates the maximum number of times to go over the data during training. The RandomForestClassifier model created with Random Forest Algorithm is used as one of the hyperparameters “n_estimators=20”. This parameter, with a default value of 100, indicates how many trees are in the forest. If the other parameter used is “random_state=0”, it is then added to ensure that the same result is obtained in every study. Other hyperparameters are used by default. The BernoulliNB model is used since the binary classification was made within the scope of the study. The model based on the Bayesian algorithm was used with the default hyperparameters.

The SVC model, which is a classification model of the Support Vector Machine algorithm, has been used as default with parameters other than “kernel=linear” one of the hyperparameters in the study. Since the data can be separated linearly, and the number of features extracted from the data is high, this parameter value is generally used in text classification problems [58]. The training dataset and the training vector were added to the models with the fit method. Prediction or classification was made on the test data by using the predict method on these models.

While creating the LSTM model, root finding and text vectorization were not applied in the data preprocessing step. After the noisy data cleaning and the removal of pause words, the previously mentioned tokenization data preprocessing method was used. The parameters of the Sequential model used are shown in Figure 9. An embedding layer, a dropout layer, an LSTM layer, and a Dense layer have been added to the model with the hyperparameters max_features, embed_dim, and lstm_out [59].

The Sequential model provides a sequential and layered structure. Each layer has an input and an output value. Layers are added to the model with the add method [60]

Embedding layer added that is frequently used for Keras text data to the model. Of the parameters in this method, “max_features” refers to the size of the vocabulary, and “input_length” refers to the length of the input strings. “embed_dim” defines the size of the vector space in which the words will be embedded. Furthermore, the “embed_dim” parameter expresses the size of the output vectors in the relevant layer for each word [61].

After the Dropout layer had been added to avoid the overfitting problem, the LSTM layer was added to the model. The lstm_out parameter used in this method represents the size of the output space. In the Dense layer, “2” is added as the first parameter because of the binary classification within the scope of the study. The second parameter is added as “sigmoid” that is used for the activation function.

Finally, the “categorical_crossentropy” Adam algorithm and the accuracy metric have been added as parameters to the model compile method. Training data is added to this model with the fit method, as shown in Figure 10. The training data, the “epoch” number that shows the number of times to go over the whole dataset, the “batch_size” that represents the number of training samples used in each iteration, and the verbose parameter used to see a detailed output, have been added to the fit_method.

Afterward, the test data is classified by using the evaluate method with the model. The method that gives the predicted classes by the model is shown in Figure 11. The evaluate method used in the model gives the loss function, while the predict method gives the predicted values.

4. Experimental Studies

There are approximately 11 thousand tweets in the public dataset [32]. The negative and positive category ratios of this dataset are shown in Figure 12. The results of the models created with the machine learning algorithms mentioned earlier on this dataset on the test data created by separating

20% of this dataset are shown in Figure 13. It shows the accuracy rates of these visual models and the comparison of results obtained when Zemberek or Snowball library is used as root-finding algorithms. It was observed that the success rates of the models are increased when the Snowball library belonging to NLTK was used as the root-finding algorithm with this dataset.

The correct and incorrect predictions obtained when the Bayesian model is tested with the test data separated from the tilted models using the Zemberek library with the ready dataset are given in Figure 14. The correct and incorrect predictions are obtained when the Logistic Regression model, which is one of the models trained using the NLTK Snowball library with the ready dataset, is tested with the test data separated from the dataset. These are shown in Figure 15.

SentimentSet dataset, which has approximately 2600 tweets, is created within the scope of the study. When the neutral category was removed from this dataset, a dataset containing 2551 tweets with positive negative category ratios shown in Figure 16 was created. 20% of this dataset was reserved for test data, and 80% was used for training.

The results of the models created with the machine learning algorithms mentioned earlier on this dataset on the test data created by separating 20% of this dataset are shown in Figure 17. It shows the accuracy rates of these visual models and the comparison of results obtained when Zemberek or Snowball library is used as root-finding algorithms. It was observed that the success rate of the models increased when the Zemberek library was used as the root-finding algorithm with this dataset. The correct and incorrect predictions obtained when the SVM model is tested with the test data separated from the dataset, which is one of the best models using the Zemberek library with the SentimentSet, are given in Figure 18. The correct and incorrect predictions obtained when testing with the test data separated from the Random Forest model dataset, which is one of the models trained using the NLTK Snowball library with SentimentSet, are given in Figure 19.

Models trained with the SentimentSet were also tested with a small sample test data consisting of 20 randomly picked and marked nine positive and ten negative tweets that did not belong to the dataset but were generated by the same way of SentimentSet. The results obtained are shown in Figure 20 with the comparison of root-finding algorithms. The correct and incorrect predictions obtained when the Bayesian model is tested with sample test data, one of the best models using the Zemberek library with the SentimentSet, are given in Figure 21. The correct and incorrect predictions obtained when the SVM model, which is one of the models trained using the NLTK Snowball library with SentimentSet, is tested with the sample test data are given in Figure 22.

Since the aforementioned sample test data was created with the tweet collection method of SentimentSet described earlier within the scope of the study, the tests performed with this sample test data on the models trained with the ready dataset did not result in high accuracy rates, as shown in Figure 23.


```

max_features = 2000
embed_dim = 128
lstm_out = 196
model = Sequential()
model.add(Embedding(max_features, embed_dim, input_length = X.shape[1]))
model.add(Dropout(0.8))
model.add(LSTM(lstm_out))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics = ['accuracy'])

```

FIGURE 9: LSTM model and hyperparameters.

```

batch_size = 16
model.fit(X_train, Y_train, epochs = 6, batch_size=batch_size, verbose = 2)
score, acc = model.evaluate(X_test, Y_test, verbose = 2, batch_size = batch_size)

```

FIGURE 10: LSTM model fit and evaluate methods and parameters.

```

data_list = pd.DataFrame (model.predict_classes(X), columns = ['sentiment'])

```

FIGURE 11: LSTM model predict_classes method.

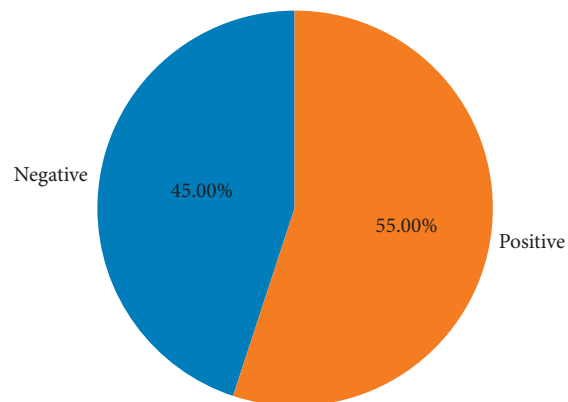


FIGURE 12: Visualization of the category ratios of the public dataset.

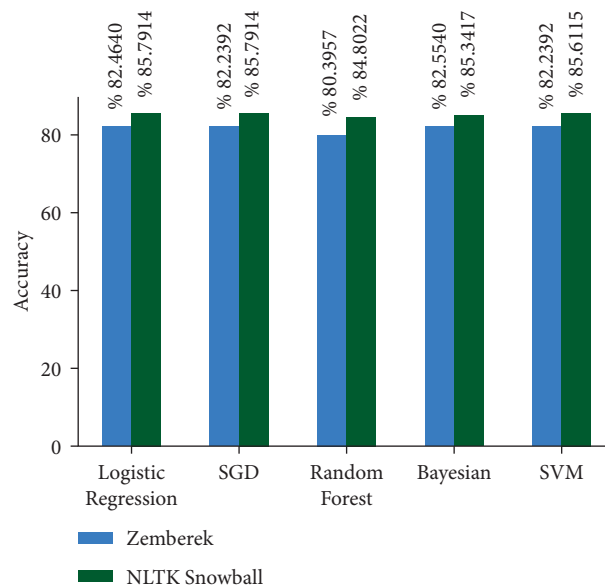


FIGURE 13: Comparison of accuracy rates of public dataset machine learning algorithms according to root-finding libraries.

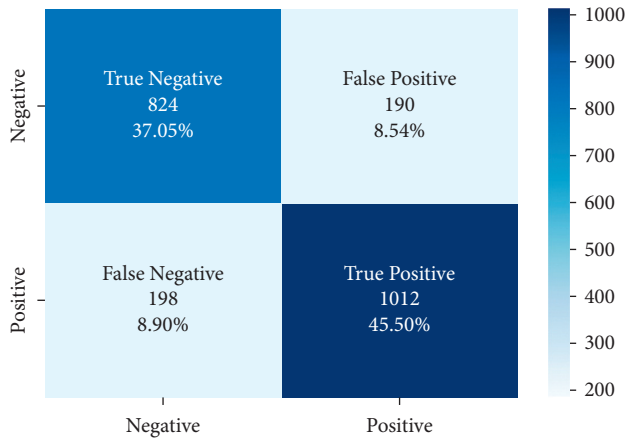


FIGURE 14: Distribution of test results and test data belonging to the dataset of Bayesian model trained with public dataset and Zemberek root-finding library.

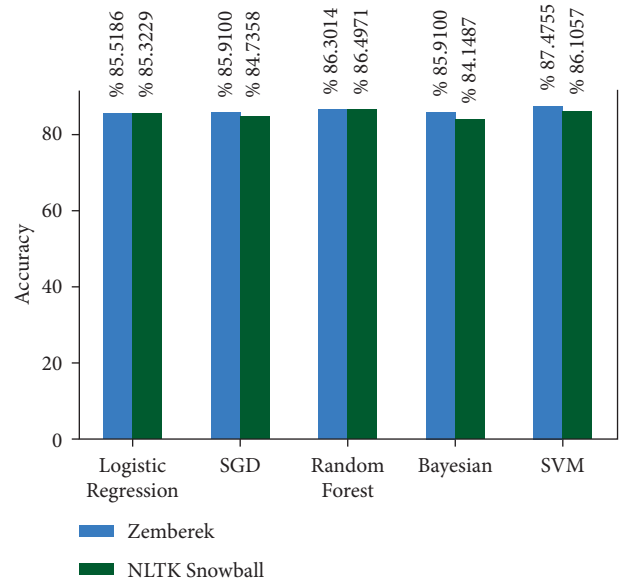


FIGURE 17: Comparison of the accuracy rates of SentimentSet and machine learning algorithms according to root-finding methods.

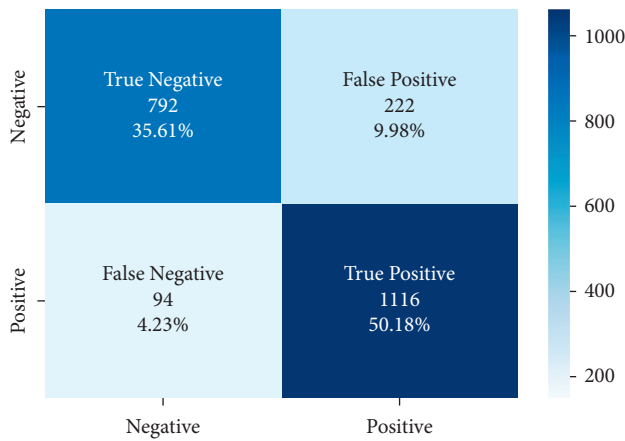


FIGURE 15: Distribution of test results and test data belonging to the dataset of logistic regression model trained with public dataset and Zemberek root-finding library.

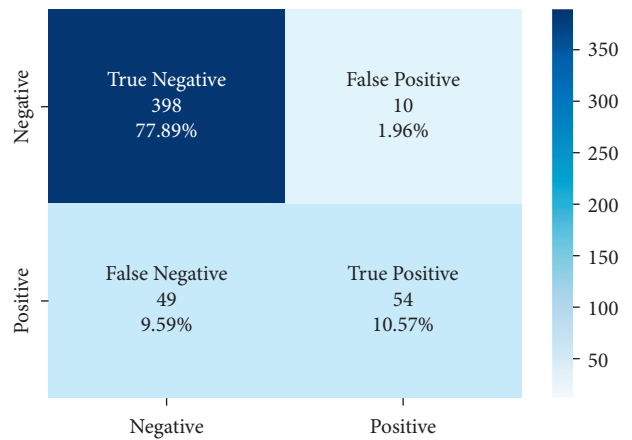


FIGURE 18: Distribution of test results and test data of SVM model trained with SentimentSet and Zemberek root-finding library.

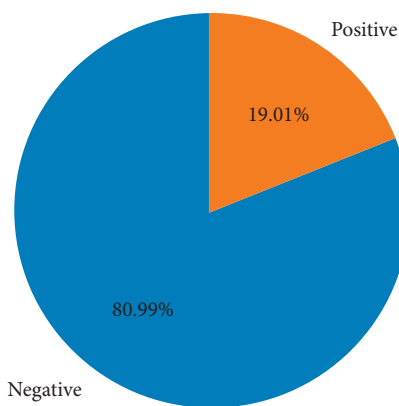


FIGURE 16: Visualization of the category ratios of the SentimentSet that is created in this study.

In the studies conducted with LSTM, root finding preprocess was not performed. While the model was trained, 20% of the dataset was reserved for the test data, and 80% of

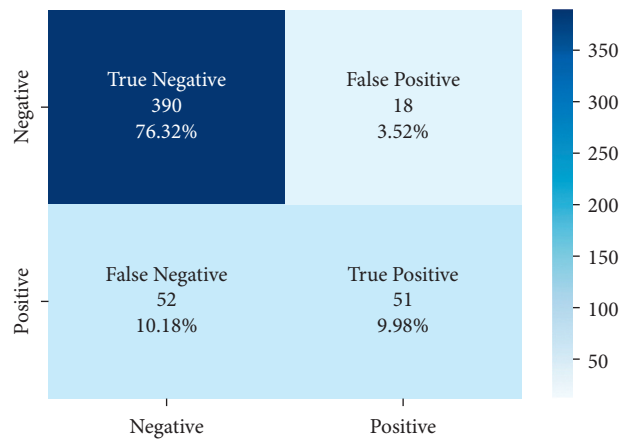


FIGURE 19: Distribution of test results and test data belonging to the dataset of random forest model trained with SentimentSet and NLTK snowball rooting library.

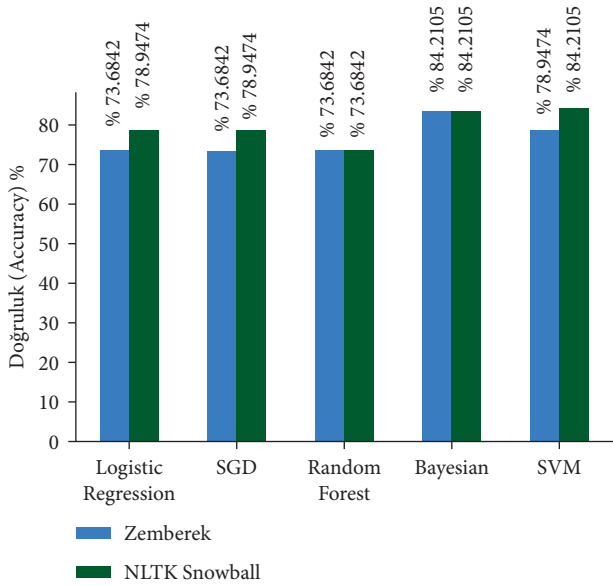


FIGURE 20: Test results with sample test data on models trained with SentimentSet.

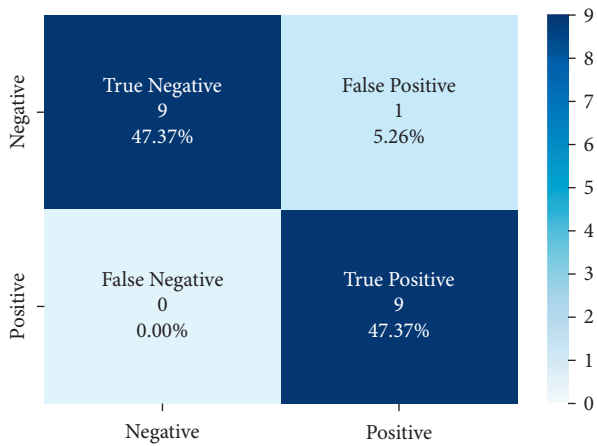


FIGURE 21: Distribution of test results with sample test data of Bayesian model trained with SentimentSet and Zemberek root-finding library.

the dataset was used for training in the models created for both the public dataset and SentimentSet. The results obtained when the models trained using the public dataset or SentimentSet are tested with the test data of the dataset or with the sample test data consisting of 20 tweets are shown in Figure 24.

The training and validation accuracy values obtained while training the LSTM model on the SentimentSet are shown in Figure 25, and the training and validation error rate values obtained during the training are shown in Figure 26.

The correct and incorrect predictions obtained when the LSTM model trained with SentimentSet is tested with the test data that belongs to the dataset are given in Figure 27. The correct and incorrect predictions obtained when the

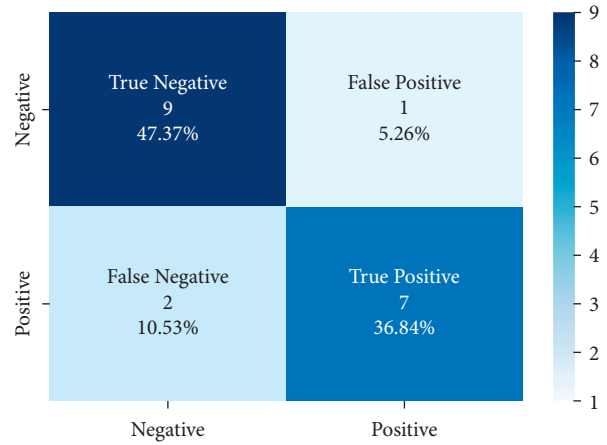


FIGURE 22: Distribution of test results with sample test data of SVM model trained with SentimentSet and NLTK snowball root-finding library.

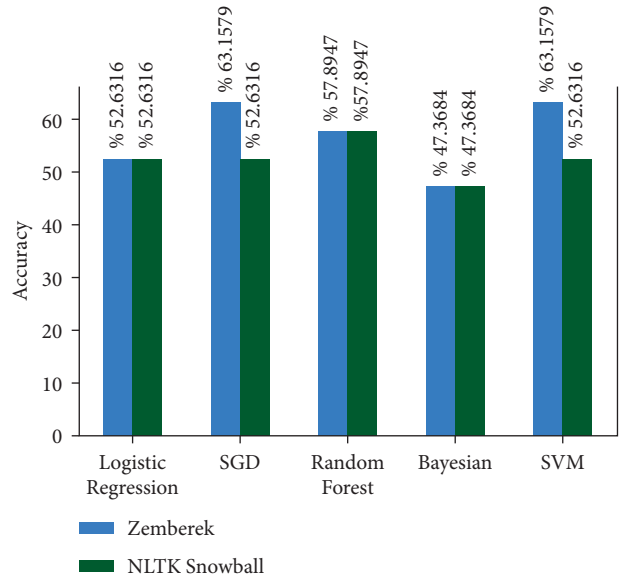


FIGURE 23: Test results with sample test data on models trained with the public dataset.

LSTM model trained with SentimentSet is tested with the sample test data are given in Figure 28. Training and validation accuracy values during LSTM model training with the public dataset are shown in Figure 29. Training and validation error rate values during LSTM model training with the public dataset are shown in Figure 30.

The correct and incorrect predictions obtained when the LSTM model trained with the public dataset is tested with the test data that belongs to the dataset are given in Figure 31. The results obtained when the LSTM model trained with the public dataset is tested with the sample test data are shown in Figure 32.

In this study, various machine learning algorithms are used with different datasets and root-finding algorithms. All accuracy ratios of the combinations are shown in Table 2.

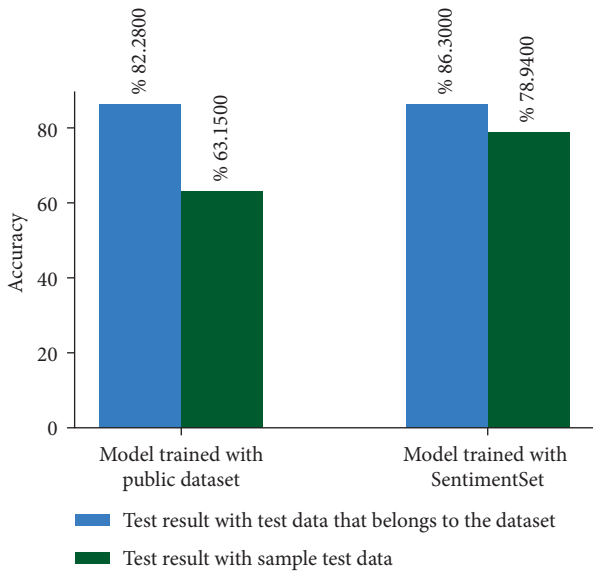


FIGURE 24: LSTM model training and test results with different datasets.

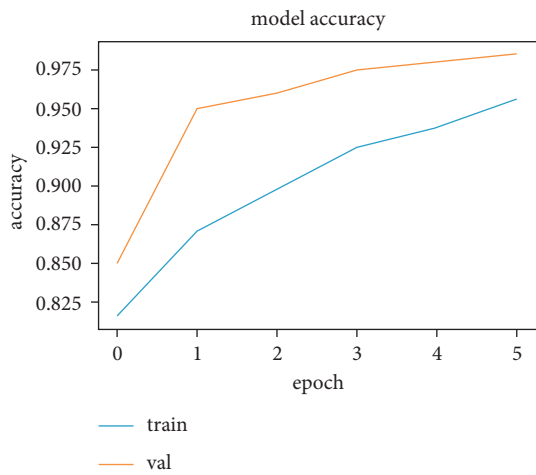


FIGURE 25: Training and validation accuracy values during LSTM model training with SentimentSet.

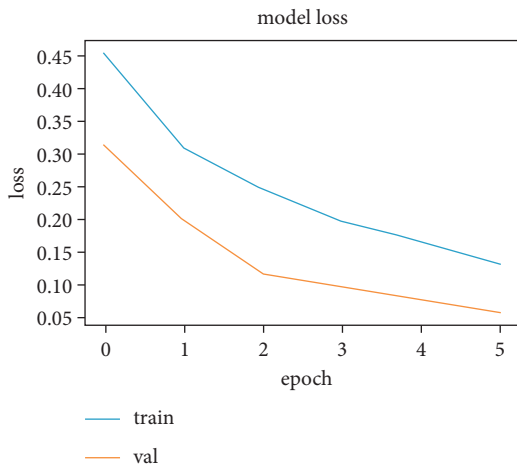


FIGURE 26: Training and validation error rate values during LSTM model training with SentimentSet.

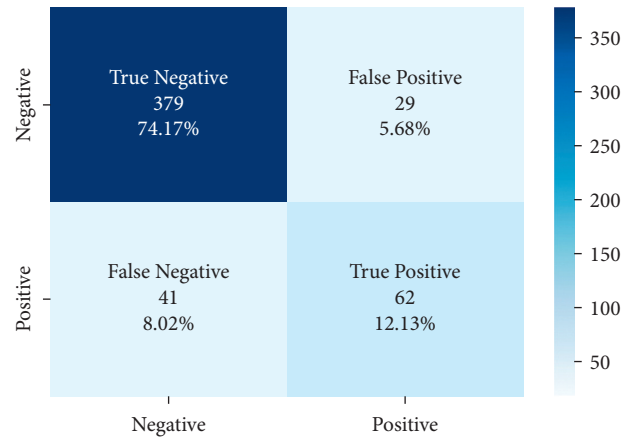


FIGURE 27: Distribution of test results and test data belonging to the dataset of the LSTM model trained with SentimentSet.

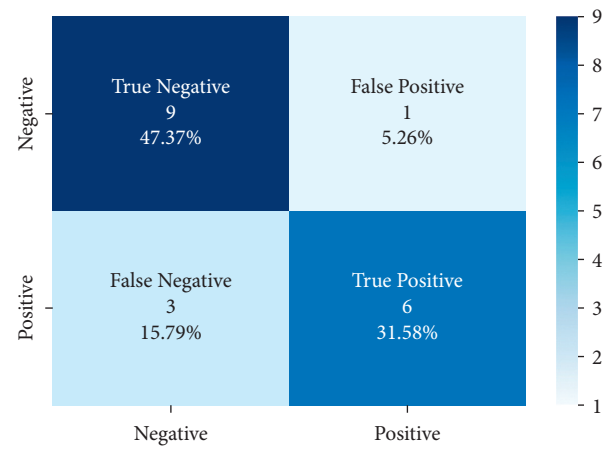


FIGURE 28: Distribution of test results with sample test data of the LSTM model trained with SentimentSet.

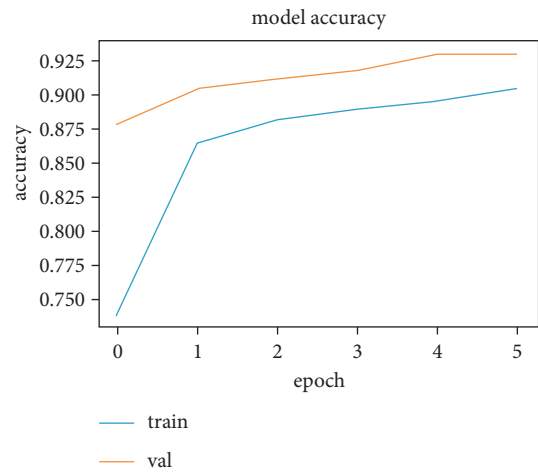


FIGURE 29: Training and validation accuracy values during LSTM model training with the public dataset.

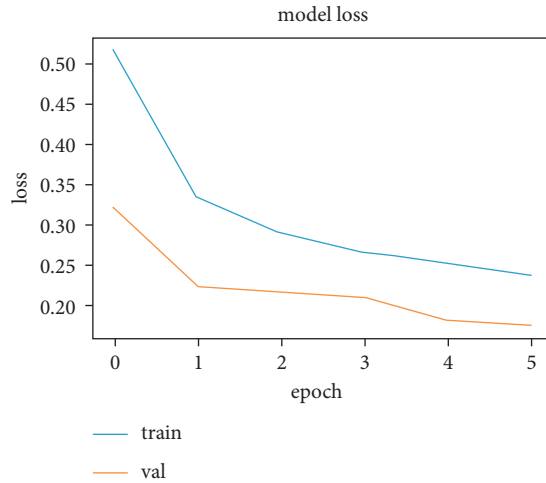


FIGURE 30: Training and validation error rate values during LSTM model training with the public dataset.

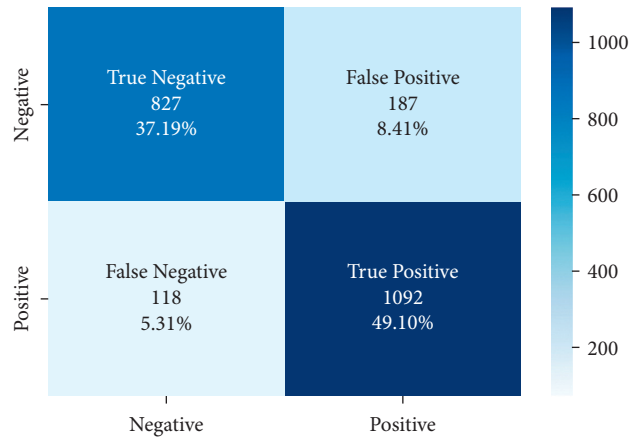


FIGURE 31: Distribution of test results and test data belonging to the dataset of the LSTM model trained with the public dataset.

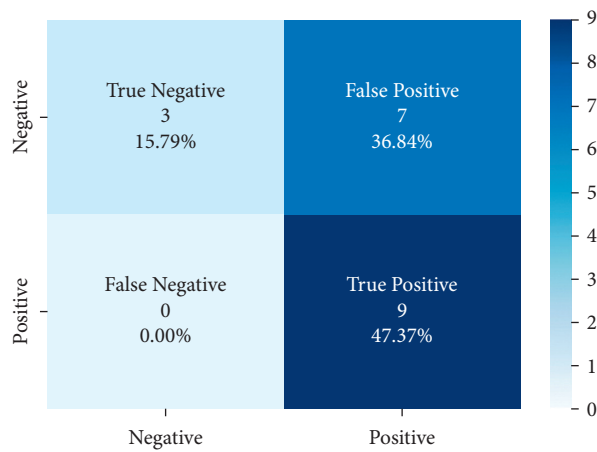


FIGURE 32: Distribution of test results with sample test data of the LSTM model trained with the public dataset.

TABLE 2: Accuracy table of the algorithms with given combinations.

Accuracy table (%)	Zemberek				NLTK snowball			
	Train with public dataset		Train with SentimentSet		Train with public dataset		Train with SentimentSet	
	Test with public dataset	Test with sample test data	Test with SentimentSet	Test with sample test data	Test with public dataset	Test with sample test data	Test with SentimentSet	Test with sample test data
Logistic regression	%82.46	%52.63	%85.51	%73.68	%85.79	%52.63	%85.32	%78.94
SGD	%82.23	%63.15	%85.91	%73.68	%85.79	%52.63	%84.73	%78.94
Random forest	%80.39	%57.89	%86.30	%73.68	%84.80	%57.89	%86.49	%73.68
Bayesian	%82.55	%47.36	%85.91	%84.21	%85.34	%47.36	%84.14	%84.21
SVM	%82.23	%63.15	%87.47	%78.94	%85.61	%52.63	%86.10	%84.21
LSTM accuracy without using any root finding library (Zemberek or NLTK snowball)								
LSTM accuracy table (%)	Train with public dataset				Train with SentimentSet			
	Test with public dataset		Test with sample test data		Test with SentimentSet		Test with sample test data	
LSTM	%86.28		%63.15		%86.30		%78.94	

5. Discussion

In this study, we examined sentiment analysis for Turkish or different language in Twitter or in general in the literature over different topics. We observed that the dataset is one of the most challenging parts of sentiment analysis in Turkish. The public dataset [32] that we used is a general dataset that could be worked with different topics. Besides, we also aimed to contribute to adding benchmark dataset for “pandemic,” which is the hot topic recently. We made our study over these datasets, with using several machine learning algorithms. The results obtained by using different preprocessing techniques, different datasets for training and testing, and different machine learning algorithm combinations can be found in Table 2.

Furthermore, there are limited number of studies existing in the literature. Although there are not an enormous number of Turkish sentiment analysis studies, there are limited valuable studies in the literature. The accuracy results that belong to some of these studies from literature are given in Table 1.

Kirelli and Arslankaya [11] carried out a study that is Turkish sentiment analysis on global warming topic over 30000 random tweets from Twitter with using SVM, K-NN, and Bayesian. In that study, Hayran [27] used a dataset to train their models. The dataset was created with the labelling data based on emoticons; therefore, the accuracy could be little less than that of the manually labeled datasets. If we compare our study with some recent pandemic related Turkish language studies, we have nearly the same accuracy as the study [24] that used their own manually labeled benchmark dataset and worked with RNN, CNN, and HAN, which are deep learning models. However, our study differentiates with using different algorithms and datasets with different topics. On the other hand, another study [28] has very high accuracy over 15k tweets and CNN and bidirectional LSTM, which used a lexicon to label big number of data. As for us, in our study, we did not use lexicons; however, we used manually labeled dataset and SentimentSet that is created in the scope of the study. Our study

contributes to the literature in many ways, but mainly it offers a general view of Turkish sentiment analysis accuracy on the several machine learning algorithms with SentimentSet, which could be used for future studies as benchmark dataset.

6. Conclusion

6.1. Theoretical Conclusion. With the development of technology day by day and the acceleration of artificial intelligence in the sector and academic studies, the diversity and number in this field in the literature are quite high today. Natural language processing, which is growing in use, is one of the topics of interest in this field. Social media are a very good data source for natural language processing studies because a wide variety of data can be accessed very quickly, and they are open source. Twitter API enables the anonymous use of Twitter users’ tweets for academic studies or research after obtaining the necessary permissions. While the number of studies conducted in Turkish was very few in the past years, it is observed that it is increasing day by day.

Although Liu [62] says that it is also possible to make sentiment classification based on unsupervised learning, in this study, we preferred to make classification that is based on supervised learning with manually labeled datasets. Due to create SentimentSet, some words such as corona and pandemic were searched in the tweets. Likewise, when the sample data created with filters are tested with both models trained with the public dataset [32] and models trained with SentimentSet, it has been seen that the models created with SentimentSet are more successful. This shows that, in NLP tasks such as sentiment analysis, the similarity of the training data content and the test data content yields more successful results. The positive and negative category weights of the datasets used in the study are given. In the comparisons conducted on the same libraries and the same algorithms in the training phase carried with these datasets, it was seen that the negative prediction accuracy rate of SentimentSet with high negative data weight was higher than the positive prediction accuracy rate. It was observed that the positive

prediction accuracy rate in the models trained with the ready dataset was higher than the SentimentSet. This confirms what is known that the dataset to have equal weights in classification problems is very important for machine learning algorithms to function in their best. However, it is more difficult to achieve these equal weights in real-life problems. Tweets with words such as “Pandemic” and “Corona” generally have a negative meaning. Hence, it is challenging to make manually labeled and equal weights in each class dataset with big amount of data on the other hand.

High negative data ratio of SentimentSet shows that there are some limitations choosing negative words like “pandemic” to collect tweets. While creating a custom dataset, it is ideal to have homogeneous labeled data with working on machine learning tasks. The more balanced dataset could give more accurate results in addition to SentimentSet dataset that was created for a periodic time that is approximately early of COVID-19 with limited data. Opinions about the words may vary because we are still in pandemic globally, and lots of changes have occurred since the beginning. Time period and data could be improved to get more valid and accurate results.

6.2. Practical Conclusion. Within the scope of the study, sentiment analysis studies were carried out on Turkish tweets with natural language processing, and the results obtained by using various machine learning algorithms in this field were compared. Two different datasets, one public dataset [32] and the other one being SentimentSet [33] dataset that was created within the scope of the study with manually labelling, were used. These datasets were preprocessed before being used with algorithms. In the root-finding preprocess, which is one of these preprocesses, two different libraries were used and compared. In order to do sentiment analysis, these datasets were trained with Logistic Regression, SVM, Bayesian, Random Forest, and SGD algorithms, and models were produced. Apart from these algorithms, models were produced by training with datasets with LSTM, a deep learning network, and the results were compared in Table 1. This study is aimed to contribute to the natural language processing studies in the Turkish language in the literature. The trained models were tested with “Sample Test Data.” This test data was created separated from the datasets. “Sample Test Data” consists of 20 tweets collected with using the methods of creating the SentimentSet created within the scope of the study.

As seen in the study, the quality of the data is as important as the creation of models. On the other hand, the success of root-finding algorithms differs according to the dataset and the tested data. It is seen that the models trained with the SentimentSet have higher success rates with the test data separated from the dataset within itself compared to the models trained with the public dataset [32]. However, this may be a result of the negative category weight being too high in the SentimentSet. The models produced within the scope of the study can be improved, and models with better results can be produced. One of the ways to be followed for

this is to increase the quality of the data. Solutions such as better filtering, detailing, and diversification of preprocesses can be produced.

Better training can be provided by changing the hyperparameters of the algorithms used in the study. The hyperparameters of the LSTM algorithm used in this study were determined with using trial and error method. In future studies, the most appropriate values can be determined by using methods such as genetic algorithms while selecting the hyperparameters of this model.

As a result, it is challenging to work with Turkish language because of the language specifications; for example, Turkish is an agglutinative language, so it requires different techniques from English to work with, and there are a limited number of studies in the literature. This study aimed to contribute as a Turkish language study to generating good overview to compare performances of machine learning algorithms with the different libraries and datasets. To make supervised learning, manually labeled datasets were used. One of them is SentimentSet dataset that could be used as benchmark dataset by future studies. Totally, with 2 datasets and 2 preprocessing techniques and different test data combinations, there are significant results that up to %87 are taken from the models. The results of the models are shown in Table 2.

The models and results created in this study show that machine learning algorithms in the Turkish language and sentiment analysis are promising and can be better in the future. It is, of course, possible to expand this success, to produce larger data and better models. Turkish studies can be developed by eliminating the weaknesses of the models and increasing the data quality. Thus, models that perform sentiment analysis tasks in Turkish can continue to influence our lives with much higher success rates and to develop with technology as it progresses.

7. Managerial Implication

The managerial implication of our research is that organizations can apply the proposed social analytics methodology to understand people’s sentiment either pandemic or another topic that depends on people’s opinion and hence improve their approach about their services to the people.

8. Practical/Social Implications

The findings could be used to understand how the pandemic affected people’s sentiments from the tweets about this topic with given time duration. If this model is used for different timelines during the pandemic, it can be seen how people react with significant changes like COVID tests, vaccination, etc. Moreover, in general, with these public dataset [32] models, if they are used to make classification in another topic like world peace, women, or human rights, and if these results are shared with people, they could create solidarity about the event or topic and make their voices heard.

9. Limitations and Future Research

Although many machine learning algorithms including LSTM are used in this Turkish language study and taken satisfying results, better preprocessing and more balanced dataset are still needed to get better ones. In the results of algorithms, there is also a gap between accuracy of “sample test data,” which is independent from datasets that are used for training models, and “test data from datasets”; this can be seen in Table 2. This shows that there should be more qualified data to learn features more effectively.

Furthermore, while SentimentSet has limited amount of data, these custom datasets may contain more data in the future with applying lexicon-based approach to prevent manually labelling. In addition to data and preprocessing, studies [28, 46] that are conducted with LSTM and CNN have higher accuracy results; these techniques provide good results on sentiment analysis in English language and may also be used for Turkish language studies to get higher accuracy results.

As future work, new algorithm approaches including artificial neural networks with highly preprocessed balanced datasets could be done on Turkish social media sentiment analysis to maximize the accuracy.

Data Availability

The datasets used to support the findings of this study are available from the direct link in the dataset citations. The public dataset used within the scope of the study is an open dataset that can be accessed via <https://www.kaggle.com/mrtbeyz/trke-sosyal-medya-paylam-veri-seti>. The second dataset used is the SentimentSet, developed within the scope of this study, which can be accessed via <https://www.kaggle.com/caglaballi/sentimentset>.

Consent

Not applicable.

Disclosure

Part of this study was published in the MSc. thesis of the first author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] H. I. Safak, “Makine ogrenmesi nedir?” *medium.com*, 2021, <https://web.archive.org/web/20190127071042/https://medium.com/t%C3%BCrkiye/makine-%C3%B6%C4%9Frenmesi-nedir-20dee450b56e>.
- [2] Anonymus, “Yapay Zeka ve Makine Ogrenimi arasindaki fark nedir?” *exastax.com*, 2021, <https://web.archive.org/web/20210302201905/https://www.exastax.com.tr/makine-ogrenimi/yapay-zeka-ve-makine-ogrenimi-arasindaki-fark-nedir/>.
- [3] Anonymous, “Natural language processing,” 2021, https://web.archive.org/web/20210506094022/https://en.wikipedia.org/wiki/Natural_language_processing.
- [4] M. Taboada, “Sentiment analysis: an overview from linguistics,” *Annual Review of Linguistics*, vol. 2, pp. 325–347, 2016.
- [5] J. R. Saura, D. P. Marques, and D. R. Soriano, *Exploring the Boundaries of Open Innovation: Evidence from Social Media Mining*, Article ID 102447, Technovation in press, UK, 2022.
- [6] A. M. Aubaid and A. Mishra, “A rule-based approach to embedding techniques for text document classification,” *Applied Sciences*, vol. 10, no. 11, 2020.
- [7] A. M. Aubaid and A. Mishra, “Text classification using word embedding in rule-based methodologies: a systematic mapping,” *TEM JOURNAL-TECHNOLOGY EDUCATION MANAGEMENT INFORMATICS*, vol. 7, no. 4, pp. 902–914, 2018.
- [8] L. Sigler, “Text analytics vs. sentiment analysis qualtrics.com,” 2022, <https://www.qualtrics.com/blog/text-analytics-vs-sentiment-analysis/>.
- [9] G. Beigi, X. Hu, R. Maciejewski, and H. Liu, “An overview of sentiment analysis in social media and its applications in disaster relief,” *Sentiment analysis and ontology engineering*, pp. 313–340, 2016.
- [10] Y. Ruan, A. Durrezi, and L. Alfantoukh, “Using Twitter trust network for stock market analysis,” *Knowledge-Based Systems*, vol. 145, pp. 207–218, 2018.
- [11] Y. Kirelli and S. Arslankaya, “Sentiment analysis of shared tweets on global warming onTwitter with data mining methods: a case study on Turkish language,” *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 1904172, 9 pages, 2020.
- [12] K. Denecke, “Using SentiWordNet for multilingual sentiment analysis,” in *Proceedings of the IEEE 24th International Conference on Data Engineering Workshop*, pp. 507–512, Cancun, Mexico, April 2008.
- [13] H. Ghorbel and D. Jacot, “Sentiment analysis of French movie reviews,” in *Advances in Distributed Agent-Based Retrieval Tools*, Springer, Berlin, Germany, 2011.
- [14] H. Peng, E. Cambria, and A. Hussain, “A review of sentiment analysis research in Chinese language,” *Cognitive Computation*, vol. 9, no. 4, pp. 423–435, 2017.
- [15] M. Kaya, G. Fidan, and I. H. Toroslu, “Sentiment analysis of Turkish political news,” in *Proceedings of the IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, pp. 174–180, Macau, China, December 2012.
- [16] F. Akba, “Assessment of feature selection metrics for sentiment analysis: Turkish movie reviews,” MSc Thesis, Hacettepe University, Graduate School of Science and Engineering, Computer Engineering Department, Ankara, 2014.
- [17] O. Coban, B. Ozyer, and G. T. Ozyer, “Sentiment analysis for Turkish Twitter feeds,” in *Proceedings of the 23rd Signal Processing and Communications Applications Conference (SIU)*, pp. 2388–2391, Malatya, Turkey, May 2015.
- [18] H. Karamollaoglu, I. A. Dogru, M. Dorterler, A. Utku, and O. Yildiz, “SentimentAnalysis on Turkish social media shares through lexicon based approach,” in *Proceedings of the 3rd International Conference on Computer Science and Engineering (UBMK)*, pp. 45–49, Sarajevo, Bosnia and Herzegovina, September 2018.
- [19] A. Bozyigit, S. Utku, and E. Nasiboglu, “Cyberbullying detection by using artificial neural network models,” in *Proceedings of the 4th International Conference on Computer*

- Science and Engineering (UBMK)*, pp. 520–524, Samsun, Turkey, September 2019.
- [20] N. Pervan, “Derin ogrenme yaklasimlari kullanarak Turkce metinlerden anlamsal cikarim yapma,” MSc Thesis, Ankara University, Graduate School of Science and Engineering, Computer Engineering Department, Ankara, 2019.
- [21] M. Rumelli, D. Akkus, O. Kart, and Z. Isik, “Sentiment analysis in Turkish text with machine learning algorithms,” in *Proceedings of the Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1–5, Izmir, Turkey, October 2019.
- [22] A. A. Akin and M. D. Akin, “Zemberek, NLP tools for Turkish,” 2021, <https://web.archive.org/web/20210415054058/https://github.com/ahmetaa/zemberek-nlp>.
- [23] C. Coltekin, “A Corpus of Turkish offensive language on social media,” in *Proceedings of the 12th Language Resources and Evaluation Conference (LREC)*, pp. 6174–6184, May 2020.
- [24] H. A. Shehu, M. H. Sharif, M. H. U Sharif et al., “Deep sentiment analysis: a case study on stemmed Turkish twitter data,” *IEEE Access*, vol. 9, Article ID 56836, 2021.
- [25] A. Koksal and A. Ozgur, “Twitter dataset and evaluation of transformers for Turkish sentiment analysis,” in *Proceedings of the 29th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, Istanbul, Turkey, June 2021.
- [26] M. Aydogan and V. Kocaman, “TRSAv1: a new benchmark dataset for classifying user reviews on Turkish e-commerce websites,” *Journal of Information Science*, Article ID 01655515221074328, 2022.
- [27] A. Hayran and M. Sert, “Sentiment analysis on microblog data based on word embedding and fusion techniques,” in *Proceedings of the 25th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, Antalya, Turkey, May 2017.
- [28] A. T. Kabakus, “Novel COVID-19 sentiment analysis in Turkish based on the combination of convolutional neural network and bidirectional long-short term memory on Twitter,” *Concurrency and Computation: Practice and Experience*, Article ID e6883, 2022.
- [29] D. Kucuk and N. Arici, “Sentiment analysis and stance detection in Turkish tweets about COVID-19 vaccination,” in *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media*, IGI Global, Pennsylvania, PA, USA, 2022.
- [30] I. A. Acar and V. Altintas, “Analysis of selected twitter headers during the pandemic using big data method,” in *Pandemnomics: The Pandemic’s Lasting Economic Effects*, pp. 257–273, Springer, Singapore, 2022.
- [31] E. O. Akcan and N. Sututemiz, “Analysis of brand perceptions of covid-19 vaccines with sentiment analysis on social media,” *Pamukkale Universitesi Sosyal Bilimler Enstitusu Dergisi*, vol. 49, pp. 145–162, 2022.
- [32] T. Balli and C. Balli, “SentimentSet,” 2021, <https://www.kaggle.com/caglaballi/sentimentset>.
- [33] M. Beyaz, “Turkce sosyal medya paylasim seti,” 2021, <https://www.kaggle.com/mrtbeyz/trke-sosyal-medya-paylam-veri-seti>.
- [34] A. M. Turing, “Computing machinery and intelligence,” *Mind*, vol. 59, no. 236, pp. 433–460, 1950.
- [35] J. Searle, “Minds, brains, and programs,” *Behavioral and Brain Sciences*, vol. 3, no. 3, Harvard University Press, Cambridge, MA, USA, 1980.
- [36] L. R. Bahl, F. Jelinek, and R. L. Mercer, “A maximum likelihood approach to continuous speech recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 2, pp. 179–190, 1983.
- [37] P. F. Brown, J. Cocke, S. A. Della Pietra et al., “A statistical approach to machine translation,” *Computational Linguistics*, vol. 16, no. 2, pp. 79–85, 1990.
- [38] W. M. Martinez, “A natural language processor with neural networks,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, pp. 3156–3161, Vancouver, BC, Canada, October 1995.
- [39] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, “Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques,” in *Proceedings of the 3rd IEEE International Conference on Data Mining*, pp. 427–434, Melbourne, FL, USA, November 2003.
- [40] Z. Liu, X. Lv, K. Liu, and S. Shi, “Study on SVM compared with the other text classification methods,” in *Proceedings of the 2nd International Workshop on Education Technology and Computer Science*, pp. 219–222, Wuhan, China, March 2010.
- [41] H. Shi and X. Li, “A sentiment analysis model for hotel reviews based on supervised learning,” in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 950–954, Guilin, China, July 2011.
- [42] L. Yao and Y. Guan, “An improved LSTM structure for natural language processing,” in *Proceedings of the IEEE International Conference of Safety Produce Informatization (IICSPI)*, pp. 565–569, Chongqing, China, December 2018.
- [43] R. Feldman, “Techniques and applications for sentiment analysis,” *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [44] J. R. Saura, P. Palos-Sanchez, and A. Grilo, “Detecting indicators for startup business success: sentiment analysis using text data mining,” *Sustainability*, vol. 11, no. 3, 2019.
- [45] J. R. Saura, P. Palos-Sanchez, and M. A. Rios Martin, “Attitudes expressed in online comments about environmental factors in the tourism sector: an exploratory study,” *International Journal of Environmental Research and Public Health*, vol. 15, no. 3, 2018.
- [46] S. Tam, R. B. Said, and O. Tanriover, “A ConvBiLSTM deep learning model-based approach for twitter sentiment classification,” *IEEE Access*, vol. 9, Article ID 41283, 2021.
- [47] K. Garcia and L. Berton, “Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA,” *Applied Soft Computing*, vol. 101, Article ID 107057, 2021.
- [48] J. Samuel, G. G. Ali, M. Rahman, E. Esawi, and Y. Samuel, “Covid-19 public sentiment insights and machine learning for tweets classification,” *Information*, vol. 11, no. 6, 2020.
- [49] S. Boon-Itt and Y. Skunkan, “Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study,” *JMIR Public Health and Surveillance*, vol. 6, no. 4, Article ID e21978, 2020.
- [50] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, “COVIDSenti: a large-scale benchmark Twitter data set for COVID-19 sentiment analysis,” *IEEE Transactions on Computational Social Systems*, vol. 8, no. 4, pp. 1003–1015, 2021.
- [51] J. R. Saura, D. Ribeiro-Soriano, and P. Z. Saldaña, “Exploring the challenges of remote work on Twitter users’ sentiments: from digital technology development to a post-pandemic era,” *Journal of Business Research*, vol. 142, pp. 242–254, 2022.
- [52] JustAnotherArchivist, “A library for getting tweets,” 2021, <https://web.archive.org/web/20210507030818/https://github.com/JustAnotherArchivist/snsrape/blob/master/snsrape/modules/twitter.py>.

- [53] J. Son, “Stopwords for Turkish,” 2021, <https://github.com/Samsung/KnowledgeSharingPlatform/blob/master/sameas/lib/lucene-analyzers-common-5.0.0/org/apache/lucene/analysis/tr/stopwords.txt>.
- [54] B. Zeyrek, “Makine öğrenmesi ile türkçe tweetlerde duygu analizi,” 2021, <https://medium.com/@bernazeyrekk/makine-ogrenmesi-ile-covid-19-türkçe-tweetlerde-duygu-analizi-79f1acd8cbdf>.
- [55] C. Khanna, “What and why behind fit_transform() and transform() in scikit-learn!” towardsdatascience.com,” 2021, <https://towardsdatascience.com/what-and-why-behind-fit-transform-vs-transform-in-scikit-learn-78f915cf96fe>.
- [56] I. R. Hallac, “Metin on isleme adimlari icin Keras tokenizer sinifi kullanimi,” 2021, <https://web.archive.org/web/20200718020415/http://buyukveri.firat.edu.tr/2018/06/04/metin-on-isleme-adimlari-icin-keras-tokenizer-sinifi-kullanimi/>.
- [57] B. Zeyrek, “Covid-19 sentimental analysis,” 2021, <https://github.com/BernaGulZeyrek/Covid-19/blob/master/covid-19.ipynb>.
- [58] Anonymous, “Creating linear kernel SVM in Python,” 2021, <https://web.archive.org/web/20201028004630/https://www.geeksforgeeks.org/creating-linear-kernel-svm-in-python/>.
- [59] P. Nagy, “LSTM sentiment analysis | Keras,” 2021, <https://www.kaggle.com/ngyptr/lstm-sentiment-analysis-keras>.
- [60] fchollet, “The sequential model,” 2021, https://web.archive.org/web/20210507003609/https://keras.io/guides/sequential_model/.
- [61] J. Brownlee, “How to use word embedding layers for deep learning with Keras,” 2021, <https://web.archive.org/web/20210116064720/https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/>.
- [62] B. Liu, “Sentiment analysis and subjectivity,” *Handbook of natural language processing*, vol. 2, pp. 627–666, 2010.

Research Article

Research on the Design of Intelligent Music Teaching System Based on Virtual Reality Technology

Wei Chen 

Music and Dance Department, Zhengzhou Normal University, Zhengzhou 450000, Henan, China

Correspondence should be addressed to Wei Chen; chenwei@zznu.edu.cn

Received 23 February 2022; Revised 9 March 2022; Accepted 21 March 2022; Published 7 April 2022

Academic Editor: Tongguang Ni

Copyright © 2022 Wei 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.

With the continuous development and innovation of artificial intelligence technology, its application in the field of music education is also increasing, music classroom has accepted and applied a more efficient and intelligent teaching system. In the reform of teaching, virtual reality (VR) technology has gradually become a new means which occupies a place in the field of education and scientific research. The teaching system based on virtual reality has been focused in all kinds of teaching. Therefore, in this paper, VR is used to build a music teaching system based on model embedding, bread capture, packing capture and camera establishment, so as to implement the music teaching platform based on VR. Through the construction of different virtual elements, it can better achieve the goals of public participation and can effectively stimulate the singer's sensory organs.

1. Introduction

Virtual Reality (VR) technology and augmented reality (AR) technology attract more and more people to engage in the research and development of related theories and technologies [1]. At present, VR technology and AR technology are widely used in entertainment [2], tourism, medical [3], games [4], education [5], etc. Many VR and AR companies and related talents [6] have emerged in the society. Many colleges and universities are naturally unwilling to fall behind. They have built a series of virtual laboratories and virtual courses [7] related to virtual reality and augmented reality. Moreover, not only higher education institutions, but also many training institutions have seized the opportunity, a lot of courses about virtual reality are set up to continuously import talents for the society [8].

In the reform of teaching, the education mode has gradually changed from the traditional blackboard-writing and PPT teaching to the combination of informatization and traditional education [9]. Especially in the field of practice teaching in higher education, According to “The Implementation Plan for Accelerating the Modernization of Education (2018–2022)” issued by the general office of the CPC Central Committee and the general office of the State

Council, it is mentioned that “efforts should be made to build new education and teaching mode based on information technology,” “it is necessary to develop the construction of national teaching project about virtual simulation experiment,” which points out that we should make full use of computer simulation technology to carry out the construction of virtual simulation experiment teaching, promote the sharing of advantageous educational resources in Colleges and universities, solve the problem of unreasonable allocation of educational resources, and improve the overall level of education in China [10, 11].

2. Analysis of Virtual Reality Technology

2.1. Virtual-Real Integration Technology. As an important basis to distinguish VR and AR technology, Virtual-real integration technology is one of the three characteristics of augmented reality technology. The “real” here refers to the real objects captured by the camera, and the “virtual” here is composed of models, sounds and words created by computer software, which is a supplement to the objects in real world. Then, virtual real integration technology is used to stack the virtual thing and the reality to ensure the consistency of illumination, geometry and motion, so as to

realize seamless superposition and achieve augmented reality [12, 13].

Virtual-real integration technology is realized by three methods: 3D model reading and rendering, pixel operation and texture rendering. The main principle of texture rendering is to take the real image as the texture data, and then regard the texture as the unit to draw the real image on the surface perpendicular to the virtual camera, afterwards, a virtual object is generated between its plane and the camera [14]; the basic principle of pixel operation is to use its function to read the data of the real image into the cache and draw the virtual object; In addition, 3D model reading and drawing is mainly through reading the format of the 3D model, and then reading the data (vertex and face of the model); Furthermore, open GL can draw objects through these data and materials [15]. The way of reading the data of 3DS model is employed to realize the integration of virtual information and landmarks.

An experiment in VR course on music genre recognition was carried out in a primary school, which immerses students in different musical styles (such as classical, country, jazz and swing) through mobile VR devices. The results show that compared with the traditional courses of printed materials and passive listening, the combination of mobile VR technology and traditional teaching methods can improve experience of music learning in the aspects of actively listen, attention focusing and others [16]. The combination of VR technology and these related courses can better solve the disadvantages in the traditional teaching process. Through VR teaching, the scene can be vividly presented in front of the students. While students will not be limited by time, frequency, distance, safety and other factors. VR equipment can be used to listen to the sound repeatedly, so as to achieve emotional integration, which can not only enhance the interest in class but also solve the practical problems in music teaching.

2.2. Unity3d Technology. Unity3d is a multi-platform game tool designed to be easy to use from the beginning [17]. As a fully integrated professional application, it is also a powerful game engine with multi-million dollar, as well as a fully integrated editor [17].

Unity3d and integrated development environment are perfectly combined [18]. This joint integration allows the editor to do whatever it takes to publish a game [18]. Simple, visual, intuitive, these features of editor make the construction of games more interesting. Originally a game development kit for Mac, windows, and Linux, it was developed to be deployed on iPhone and Wii, or on the web [19]. However, this is not common in game engines that Unity3d is a scripting language. Another example is that Second Life also uses mono as the script engine and C# as the scripting language [20]. Its application in the game engine promoted the progress of Mono itself, including Mono.Simd, which makes Mono or managed code more suitable for the development of game [21].

3. Classifications of Music Teaching System Based on Virtual Reality Technology

3.1. Model Embedding System. In virtual reality, model embedding system is an important factor to express the effect of experience, and its function of geometric segmentation plays a decisive role in the fluency of experience in scene interaction. The same model is divided into three levels of LOD (levels of detail) accuracy specifications, so as to flexibly switch the model accuracy in different Line-of-sight range [22, 23]. The principle that this switching mode must follow is to ensure that the model accuracy of the foreground is relatively high, while the accuracy of the model of the middle scene is moderate, otherwise, the model of the long-range is relatively low, and even can be expressed by the way of map mask or image substitution, so that the virtual reality can be optimized in real time according to the different perspectives of the scene.

3.2. System of Face Tracker. The main task of face tracker is to determine the size, position, distance and other attributes of facial features such as iris, nose wing, mouth corner and so on, and then their geometric features are calculated to form a feature vector to describe the face as a whole [24]. The core principle of the technology is to follow the analysis of local human feature and algorithm of neural recognition. The main purpose of this paper is to compare, judge and confirm all the original parameters in the recognition database based on the features of human facial activity.

3.3. System of Gesture Tracker. The system of gesture tracker is based on Oculus quest2 hand positioning and tracking technology, which can capture the spatial coordinates of the joints of human hands, and transmit them to the animation of virtual reality in real time [25]. The main principle is to collect the bending posture of each finger, and make all fingers form a data format of unified single byte through data normalization algorithm, so as to reduce redundant data. At the same time, smoothing algorithm is used to process the spatial and temporal parameters between fingers to make the skeleton and muscle of gesture form a natural and soft state [26]. In addition, in the process of gesture capture, segmentation methods based on obvious features will be formed, including skin color segmentation and hand shape segmentation.

- (1) Skin-color segmentation: it is a method of using cluster skin color to establish skin color model in precise coordinates, which comprehensively confirms skin color with the help of RGB color gamut
- (2) Handshape segmentation: it is a method based on multi-mode integration, which is mainly to overcome the limitations of segmentation conditions of the main structure in complex environment, and improve the apparent characteristics and motion information of hand. The strategies that commonly used include segmentation of geometric features, deformable features and spatial coverage features.

3.4. Camera System. The camera system in virtual reality means the input of the first view angle, which is also a form of active vision calibration that can effectively record the dynamic scene observed by human eyes [27, 28]. Different from the traditional virtual camera, it does not need to use a calibrated object of known size, but establishes the coordinate points and image points on the calibration of object. If the access of stable camera function needs to be get, it is necessary that program optimization work must be completed at the bottom of the program.

4. Design of Music Teaching System Based on Virtual Reality Technology

4.1. Overall Design. Based on the analysis of the system principles above, it can be concluded that the ideas of virtual reality in music teaching should be studied from three-dimensional modeling, facial capture, gesture capture, camera processing and other aspects. Firstly, it is necessary to use 3dsmax as the initial tool to complete the modeling of scene and role. After all models are improved, the scene model and character model should be imported into platform of Unity3d virtual reality, and the model embedding system in the platform should be used to make appropriate geometric segmentation of the model, and determine the relationship between LOD of different levels and the scene [29]. Secondly, the system of face tracker is used to identify and bind the faces of singer, and manages the positioning points of the main structure. Thirdly, the nodes of hand bone are confirmed by system of gesture tracker, and the corresponding data are calculated simultaneously with the interface of virtual reality engine. Finally, by optimizing the camera system, the singer's performance can be recorded in real time, which is convenient for the follow-up to analyze and evaluate the changes before and after the application of virtual reality. Generally speaking, the process above is based on an efficient and concise idea, as shown in Figure 1.

4.2. Creation and Optimization of Model. Model is very important in the process of production and experience. The structure accuracy and patch distribution of the model will directly affect the degree of simulation and interaction of virtual reality.

4.2.1. Creation of Model

(1) *Scene Model.* Taking a scene of T-shaped stage as an example, the main methods are as follows:

- (1) First, use spline in 3dsMax to create a T-shaped stage of 3000 cm (length) × 2200 cm (width) × 1000 cm (height), convert it to polygon edit, and weld each vertex into a whole, so as to facilitate the subsequent connections of each boundary.
- (2) Secondly, use grid wiring to process the details of the overall model, with functions such as connection, extrusion, chamfering, and insertion to refine the

local structure of the stage model. After independent modeling, the overall bridge is carried out.

- (3) Finally, use geometric lofting and polygon editing to create auxiliary models such as auditoriums and top light stands around the stage. During the modeling process, mirroring and copying of simple models can be used to enrich the overall scene, such as Figure 2.

(2) *Role Model.* The role model should be created under box elements, while the face and body of the role should be wired as a whole with polygon editing. It is necessary to ensure the wiring of facial features, body joints and other areas that need movement under virtual reality animation, and refine the structural relationship. In some areas that do not participate in animation motion, the number of model faces can be effectively controlled by means of collapse, patch merge, etc. which is shown in Figure 3. The unavoidable triangular wiring in the model is placed in the hidden area where the character does not participate in the animation calculation, so as to avoid the unfavorable phenomena such as patch folds in the animation of virtual reality.

4.2.2. Optimization of the Model. In this process, the model is imported as a whole into Unity3D, the frame rate of the preview virtual reality is set to 70~90 FPS, and the vertex closures of the model patches is deleted in the engine. In addition, the code is implanted at the blueprint interface where secondary optimization of the subtle parts of the scene is carried out. In order to meet the optimization of geometric segmentation, spatial coordinates, patch processing, rendering baking and other aspects of the model, the program design is as follows:

```
int main (int, char**) {
    osg Producer:: Viewer viewer; &par; Create a scene
    viewer. set UPViewer ( ) :
    II Load the osga terrain model into the node variable
    osg:: Node * node = osg DB::read Nodefile ("Wutai.
    osga");
    viewer.set Scene Data(node); II Load the model into the
    scene
    &par; Enter the rendering loop
    viewer. realize ( );
    while (! viewer. done ( ) ) {
        viewer.sync( ); II Wait for the completion of all cull and
        draw threads
        viewer.update( ); II Update the scene by traversing the
        nodes
        viewer.frame( );&par;Render the updated result
    }
    viewer.sync( ); II Wait for the completion of all cull and
    draw threads before exiting the program
    return;
}
```

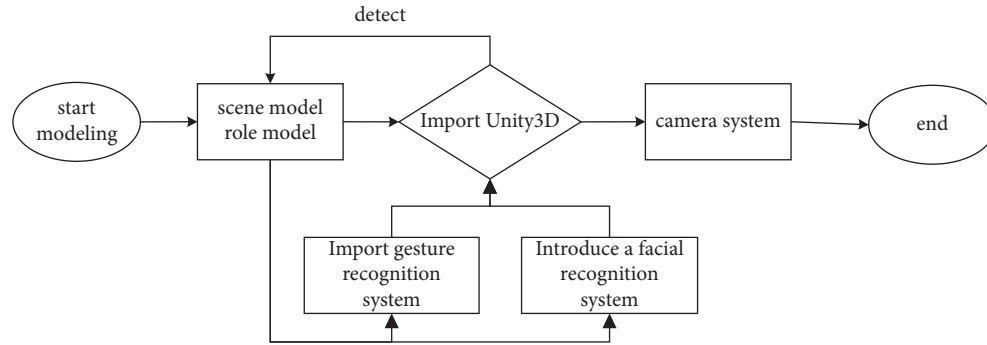


FIGURE 1: Design of work flow.

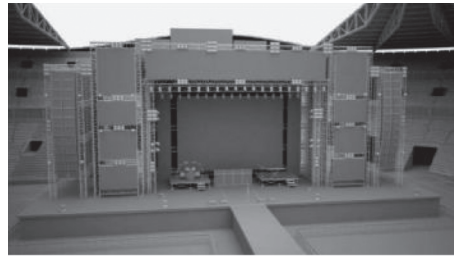


FIGURE 2: Model of Stage scene.

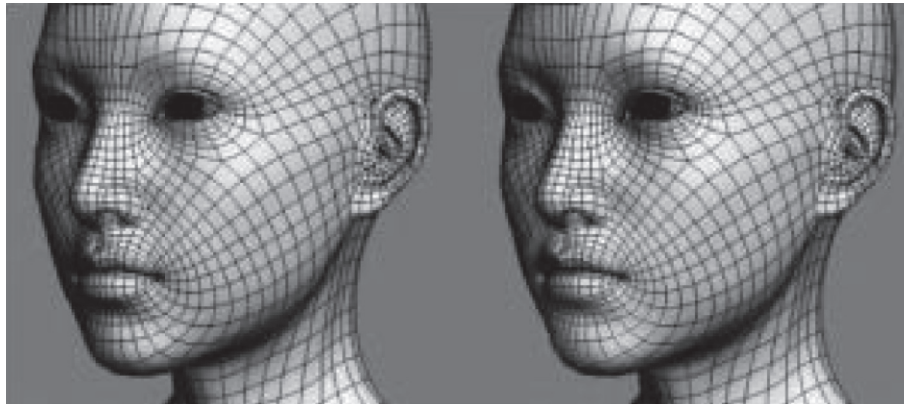


FIGURE 3: Role model.

4.3. Capture of the Face. Due to the principle of structured light adopted by the system, it is necessary to project light in the direction of the face, and then use the data that is on the surface of the object to be read to determine the shape of the face [30]. When choosing a face acquisition device, in addition to configure distance sensors, microphones, and front-facing cameras, it is also necessary to have infrared lenses, floodlights, floodlight sensing elements and dot matrix projectors arranged in sequence. Usually, the dot projector can project a dot matrix composed of more than 30,000 invisible light points to the face, then the face captured by the front-facing camera is simultaneously calculated to obtain the depth information of the facial expression, that is, the 3D model of real face. The four data interfaces that need to be built for simultaneous calculation are as follows:

- (i) 11FTFace Tracker: the main interface for face tracking.
- (ii) IFTResult: the result of face tracking.
- (iii) IFTImage: image buffer.
- (iv) IFTModel: model of 3D face.
- (v) The data that needs to be acquired in simultaneous calculation are as follows:
- (vi) FT_CAMERA_CONFIG: color or depth sensor data.
- (vii) FT_VECTOR2D: two-dimensional vector data.
- (viii) FT_VECTOR3D: 3D vector data.
- (ix) FT_Translate(XYZ): all input data required for face tracking.
- (x) FT_Rotate(X Y Z): 3D model of face angle data.

- (xi) FT_Scale(X Y Z): weight matrix data, as shown in Figure 4.

Compared with methods of face tracker, The accuracy of face recognition of T_j is 0.1 mm, which can exceed image 2, video 1mmcl and plane 0. When light conditions of R_i is not ideal, the method of obtaining facial information, such as the light $-\sigma$ and the received light s emitted by the dot projector will not affect the recognition efficiency of T_j , whose system of face tracker can be changed as follows:

$$T_j(x, y) = \begin{cases} 2 & \text{if } R_i(x, y) < -\sigma \\ 1 & \text{if } R_i(x, y) > \sigma \\ 0 & \text{if } -\sigma \leq R_i(x, y) \leq \sigma \end{cases} . \quad (1)$$

4.4. Capture of Gestures. Capture of gestures is a technical difficulty in virtual reality, which needs to be connected to the computer through the singer wearing a virtual head-mounted display device "Oculus quest2" and a hand tracker. After that, a depth-sensing camera is installed at the front of the head-mounted device and tilted downward by 13.4°, so that the singer can observe his hands in real time during the experience of virtual reality and track the changes of their fingertips in time. The gestures from left to right are: backward, stop, forward. If the position of the fingertip is within the zero-coordinate static zone (zc), no movement can be produced; However, when the fingertip extends forward beyond the static zone, the red progress bar of the subject's movement speed will increase linearly with the distance of the fingertip; In addition, when the finger joints move in the other direction and faces the palm downwards, the red progress bar will produce a subtle movement backwards. Specifically, it is a process of natural expansion and contraction of the palm. The process is based on the distance from the far end of the index finger of the singer's right hand to the center of the palm, and is proportionally enlarged by 2.74 times according to the size of each person's palm, so as to reduce the bending of the fingers. The noise caused by its setting parameters are as follows:

- (1) β : beta coefficient, β represents slope coefficient = velocity/ γ (the distance from the index fingertip to the boundary of the static zone), forward movement $\gamma = (\text{position} \times 2.74) - (zc + dzw)$, backward movement $\gamma = (zc - dzw) - (\text{position} \times 2.74)$.
- (2) Dead zone: at the beginning of the test, the testers put their hands in a relaxed and gently bent position. Then, the zero rest position of the gesture can be determined when their fingers are in a comfortable position, as shown in Figure 5.
- (3) α : exponent-velocity = $(\beta \times \gamma)^\alpha$, when one parameter changes, the other parameters are fixed at their intermediate values. For example, $\beta = 21$ m/s, $dzw = 25$ mm, $\alpha = 1.0$. The order of the three parameters is coefficient, static zone width, and exponent α , which was randomized. For each parameter, participants completed a large (2m) experiment (30 goals) and a small (1m) experiment (30

goals), while the order between these three parameters is not random.

In this experiment, it is required to complete the last 24 of the 30 indicators at least. Repeated measurement was used to analyze the differences of time at diverse levels, so as to separate the details of small targets and large targets. The settings are shown in Tables 1 and 2.

4.5. Establishment of Multiview Camera. In order to better improve the stability of the camera in virtual reality, and to enable the singers to examine the comprehensive performance of their actions and facial expressions in the virtual space from angles of multiple camera, it is necessary to optimize the bottom layer of the program., the code is modified as follows:

```

Camera _camera;
(i) II Use this for initialization
(ii) void Start ( )
{
    _camera = Camera.main;
}
IIThe first 3 locks of unity, from low to high are
nothing/everything/default/transparent FX/ignore
raycast/water UI
(iii) From the 1st to the 3rd can be set optional
II The first is the cube layer; the second is the sphere
layer; and the third is the capsule layer
void Update ( )
{
    if (Input.GetKey Down (Key code.A) )
    {
        _camera.culling Mask = 1 << 1; ||cube, only render
the first one
    }
    if (Input. Get Key Down (Key Code. B) )
    {
        _camera.culling Mask = 1 << 2; ||sphere, only render
the second one
    }
    if (Input.Get Key Down (Key Code.S))
    {
        _camera.culling Mask = 1 << 3; ||capsule, only render
the 3rd one}

```

Through the method above, it can be observed that the stability of the multi-view camera is ideal, which is convenient for subsequent quantitative analysis of the singer's performance, as shown in Figure 6.

4.6. Evaluation of the Test. Subjects: 20 students majoring in vocal performance, 10 males and 10 females each, 5 people per time, who are divided into 4 groups according

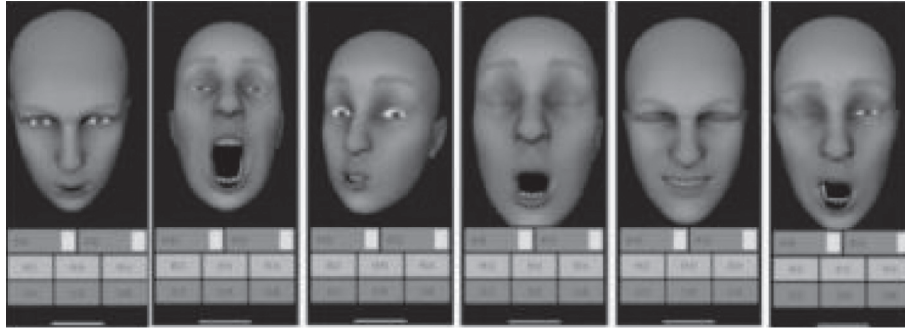


FIGURE 4: The effect after facial capture.

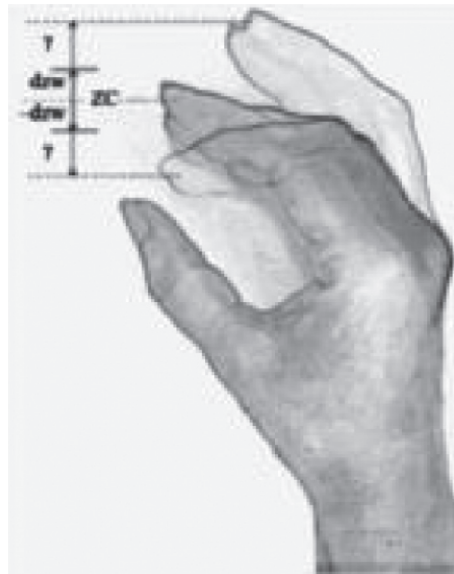


FIGURE 5: Zero stillness of gesture.

TABLE 1: Beta coefficient (small target) (m/s).

	$\beta = 12$	$\beta = 21$	$\beta = 30$	P value
Total time (s)	74.5 (48.0)	62.6 (18.3)	82.0 (31.4)	0.18
Excellent control	4.5 (0.5)	3.8 (0.9)	3.6 (0.7)	0.0001
No shoulder fatigue	4.2 (0.9)	4.1 (1.0)	4.2 (1.1)	0.65

TABLE 2: Dead zone width (small target) (mm).

	dzw = 10	dzw = 25	dzw = 40	P value
Total time (s)	71.9 (22.8)	71.0 (32.5)	73.0 (38.2)	0.94
Excellent control	3.5 (0.7)	3.8 (0.9)	4.2 (0.9)	0.15
No shoulder fatigue	4.4 (1.0)	4.2 (0.9)	4.3 (1.0)	0.59

to the groups and gender. Virtual content: the content of the custom-made 360° virtual vocal video is divided into 6 songs according to emotional classification: positive (excited), neutral (comfortable), and negative (sad), which are respectively: positive emotional songs “My Motherland and Me,” “On the Field of Hope,” negative emotional songs “Where Has Time Gone,” “Mother in Candlelight,” neutral emotional songs “Baykal Lake” and “Pastoral.”

Experimental results: The data collection and analysis were carried out in the form of questionnaires and SAM, and the results were good, as shown in Table 3.

The high-fidelity vocal interaction was obtained through the SAM, and the corresponding analysis was made before and after intervention of virtual reality, as shown in Figure 7.

From the analysis of the scale data, it can be seen that the emotional changes produced by the use of the high-fidelity vocal interactive virtual system are much higher than the traditional music teaching, which is mainly due to the immersion and high simulation brought by virtual reality technology.

5. The Application Prospect of Virtual Reality Technology in Music Teaching

When VR is applied in education, students can be more focused and more active. This advantage comes from the immersion of VR itself, which cannot be provided by traditional teaching methods. In the teaching process of VR, students can have a higher degree of participation and better integrate into the whole process [31, 32]. For subjects that require a certain amount of imagination, virtual reality



FIGURE 6: Effect of multi-view camera.

TABLE 3: Data analysis of the experience of using the vocal interactive virtual system.

Index	Excellent (%)	Good (%)	Generally (%)	Poor (%)
Immersive experience	90	10	0	0
Intelligent interaction	95	5	0	0
Fluency	99	1	0	0
Module completeness	80	15	5	0

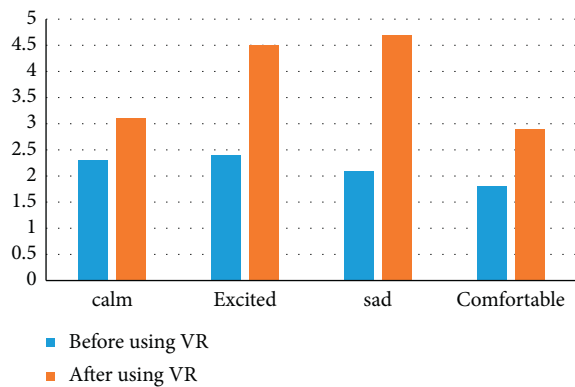


FIGURE 7: Comparison before and after the intervention of virtual reality.

technology can simulate some scenes that cannot be realized in ordinary class from diverse directions.

The virtual realization of music works and courses is mainly constructed through 3D animation and 3D roaming. In actual production, software such as 3DMAX and Unity3d is used to create multiple virtual scenes such as oceans, deserts, grasslands, forests, European castles and then they are placed in the module with different styles of music. After students enter the module and select the corresponding virtual scene, the music will be played together, which allows students to experience the music in combination with the environment. Moreover, it creates an immersive feeling and realizes situational teaching. For example, in the teaching of folk songs, students can enter the virtual scene of forests, mountains and rivers, and feel the charm of songs immersively.

Through the application of virtual reality technology, users can feel like entering a real concert hall or music classroom. That is to say, when users employ the virtual experience module, they can use the mouse to click and select different positions on the interface. Through the user's click and selection, the interface is displayed in a three-dimensional manner, which let the experienter have an immersive feeling. The specific content of the virtual

experience can be set according to the actual teaching of different majors. During the development of the system, many data interfaces are reserved, and the text, background music, pictures and other contents of the experience hall can be set according to the teaching tasks of different majors.

The analysis of functional module can be designed with students' major in music production as the main object, in which the effect of actual works of art is taken as an example. Detailed explanation and elaboration of the professional techniques and principles used in the works is received, and in this way, the effect of virtual reality is fully utilized to make the corresponding class more vivid, which provides students a better interactive experience. In musical instrument teaching, students can check the performance from different angles, which is almost the same as the effect of live teaching. Others, such as classical music and vocal music, also have the same advantages for teaching in the application.

The interaction model is the main functional module of learning and resource sharing. This function needs to provide the Unity 3D default plug-in, which is implemented by self-coding. The rotation angle range is calculated by moving the distance of the mouse, and the angle of view is controlled by Clamp Angle to complete the calculation of the angle. Then according to the relative distance between the camera and the object, the relative coordinates of the camera are calculated, and values are assigned, finally the setting of the entire camera position and angle parameters are completed to ensure the effectiveness of the perspective interaction.

6. Conclusion

In the process of education and teaching, virtual reality technology can realize interactive teaching through human-computer interaction, which brings convenience to teachers and students, and prompts the birth of a new teaching mode. Due to the limitations of musical equipment, conventional music teaching is carried out in a relatively enclosed environment, and its guidance is implemented one-on-one by teachers. In particular, courses such as vocal music, piano,

and instrumental music performance usually require students to “feel with their hearts”, which is highly subjective. The infinite extensibility and abundant teaching expressiveness of virtual reality interactive teaching can make the teaching more attractive. Moreover, using information technology to integrate virtual reality technology into music teaching, as well as building a specific place in an abstract way to provide students a “realistic” learning environment, can stimulate students’ autonomous learning it.

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. Mohd and H. Abid, “Virtual reality applications toward medical field,” *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 600–605, 2020.
- [2] X. Wang, X. Geng, and S. Han, “Research of industrial robot training simulation teaching system based on unity3d,” *Journal of the Staff and Worker’s University*, vol. 18, 2019.
- [3] W. K. You, “Design of Virtual Teaching System for Robot Noumenon Based on unity3d,” *Mechanical Research & Application*, vol. 32, 2018.
- [4] H. A. Gandhi, S. Jakymiw, R. Barrett, H. Mahaseth, and A. D. White, “Real-time interactive simulation and visualization of organic molecules,” *Journal of Chemical Education*, vol. 97, no. 11, pp. 4189–4195, 2020.
- [5] Y. Wang and Q. Yang, “Application and exploration of computer-aided vr teaching system in civil and infrastructure engineering course,” *Education Modernization*, vol. 35, 2019.
- [6] M. Weng, X. Kong, L. Huang, and B. Li, “A Low-Cost Wireless System Implementation for Interactive and Immersive Teaching,” in *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 322–323.
- [7] K. Abhishek, V. R. Kumar, S. Datta, and S. S. Mahapatra, “An integrated multi-response optimisation route combining principal component analysis, fuzzy inference system, nonlinear regression and jaya algorithm: a case experimental study on machining of gfrp (epoxy) composites,” *International Journal of Industrial and Systems Engineering*, vol. 32, no. 3, p. 497, 2019.
- [8] Y. Nissim and E. Weissblueth, “Virtual reality (vr) as a source for self-efficacy in teacher training,” *International Education Studies*, vol. 10, no. 8, p. 52, 2017.
- [9] A. Vaninsky, *Teaching and Learning Mathematics in the AR/VR Environment*, 2017.
- [10] Y. Liu and M. E. Center, “Experimental research and implementation of multimedia teaching system based on vr technology,” *Journal of Xi’an University(Natural Science Edition)*, vol. 38, 2019.
- [11] P. I. Jian, L. I. Wendong, and D. C. School, “Construction and Application of Steam Curriculum System in Vocational Schools Based on Vr/ar Technology,” *Journal of Liaoning Normal University(Natural Science Edition)*, vol. 23, 2019.
- [12] G. W. Hou, Q. K. Chen, B. Gao, X. L. Zhu, S. Q. Jiao, and S. J. University, *Development of Virtual Simulation Teaching System for “plastic Forming and Mould Design, Die & Mould Industry, Liaoning, China*, 2018.
- [13] M. A. Shao-Bin, C. W. Zhang, and W. J. Wang, “A Visual Interactive Teaching Platform Based on Ar,” *Journal of Lanzhou University of Arts and Science(Natural Science Edition)*, vol. 28, 2019.
- [14] J. Yan, “Practice and thoughts of the informatization teaching for “rail passenger service” course in higher vocational colleges,” *Journal of Yueyang Vocational and Technical College*, vol. 28, 2019.
- [15] X. U. Wen-Shuo, J. J. Wang, J. W. Qiao, F. Wang, and Y. P. Lun, “Application and Development of Vr Technology in the experiment Teaching of the Measurement and Control Technologycourses,” *Computer Knowledge and Technology*, vol. 15, 2019.
- [16] S. Serafin, A. Adjorlu, N. Nilsson, L. Thomsen, and R. Nordahl, ““Considerations on the use of virtual and augmented reality technologies in music education,” in *Proceedings of the IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*, Los Angeles, CA, USA, March 2017.
- [17] Z. Y. Yang, G. Y. Zhao, L. I. Yu-Sheng, L. I. Zhi-Yong, and Q. Z. Zhao, “Design and Application of Vr Cognition experiment for Process Fixture,” *Education Teaching Forum*, vol. 46, 2019.
- [18] M. Jiang, Y. Bai, and B. Wang, “Application of virtual reality technology in the teaching of bridge superstructure construction,” *The Theory and Practice of Innovation and Entrepreneurship*, vol. 18, 2019.
- [19] J. Chen, P. Hu, and Z. Huang, “Design and Realization of 3D Virtual Campus Roaming System Based on Unity3D,” *Non-seismic and Non-conventional Exploration Methods for Oil and Gas in Cuba*, pp. 485–492, 2018.
- [20] A. J. Wang, L. I. Zhong-Yong, and S. B. Yan, “Design and Implementation of Museum Roaming System Platform Based on unity 3d,” *Journal of Hunan Institute of Engineering(Natural Science Edition)*, vol. 22, 2019.
- [21] Y. U. Hongwen, Q. Shan, and Y. Xiang, “Design of virtual assembly system of energy-saving vehicle based on unity 3d,” *Journal of Shanghai University of Engineering Science*, vol. 28, 2018.
- [22] L. Dan, B. J. Xiao, J. Y. Xia, and K. R. Wang, “Preliminary realization of immersive east system using virtual reality,” *Fusion Engineering and Design*, vol. 128, pp. 198–203, 2018.
- [23] F. Zhang, T. Wang, K. Huang, J. Mao, M. Li, and Z. Wang, “One improved real-time infrared simulation system based on unity3d,” *Journal of Computer-Aided Design & Computer Graphics*, vol. 30, no. 7, p. 1177, 2018.
- [24] C. Deng, S. Tan, and C. Wang, “Design and Implementation of Electric Safety experiment Training System Based on 3d Virtual Simulation Technology,” *Research and Exploration in Laboratory*, vol. 38, 2019.
- [25] A. J. Wang and L. I. Zhong-Yong, “Design and Implementation of Virtual Memorial Hall Based on unity 3d,” *Journal of Hebei North University(Natural Science Edition)*, vol. 22, 2019.
- [26] A. Sing, A. Ibrahim, N. G. Weng, M. Hamzah, and W. C. Yung, *Design and Development of Multimedia and Multi-Marker Detection Techniques in Interactive Augmented Reality Colouring Book*, Computational Science and Technology, Kota Kinabalu, Malaysia, 2020.
- [27] H. Cui, Z. Jiao, and X. Yang, *Artillery Exterior Trajectory Scene Simulation System Based on unity3d*, Ordnance Industry Automation, Thane, Maharashtra, India, 2017.

- [28] W. Pan, *Design of Virtual Training System of Railway Passenger Car Bogie Based on unity3d*, China Computer & Communication, Beijing Shanghai, China, 2019.
- [29] H. Gong, Q. Zhou, and S. Feng, *Design of Virtual Simulation System of Liupao tea Planting Teaching Based on unity3d*, Modern Information Technology, Turks & Caicos Islands., 2019.
- [30] P. Abichandani, W. McIntyre, W. Fligor, and D. Lobo, "Solar Energy Education through a Cloud-Based Desktop Virtual Reality System," *IEEE Access*, vol. 7, pp. 147081–147093, 2019.
- [31] M. Y. Shen, Y. C. Wang, and K. P. Cui, *Design and Making of Interactive Courseware for Earthquake Safety Based on unity3d*, Computer Knowledge and Technology, United States, 2019.
- [32] S. Alfalah, J. Falah, A. Tasneem, E. Mutasem, M. Nadia, and F. Orwa, "A comparative study between a virtual reality heart anatomy system and traditional medical teaching modalities," *Virtual Reality*, vol. 23, no. 3, pp. 229–234, 2018.