

Innovative Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Lead Guest Editor: Tien-Wen Sung

Guest Editors: Tarek Gaber, Hamed Nassar, and Chao-Yang Lee





Innovative Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Wireless Communications and Mobile Computing

Innovative Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Lead Guest Editor: Tien-Wen Sung

Guest Editors: Tarek Gaber, Hamed Nassar, and
Chao-Yang Lee




Copyright © 2023 Hindawi Limited. All rights reserved.

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

Chief Editor

Zhipeng Cai , USA

Associate Editors

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

Academic Editors

Muhammad Inam Abbasi, Malaysia
Iftikhar Ahmad , Pakistan
Ghufran Ahmed , Pakistan
Hamza Mohammed Ridha Al-Khafaji ,
Iraq
Abdullah Alamoodi , Malaysia
Ihsan Ali , Malaysia
Jehad Ali , Republic of Korea
Marica Amadeo, Italy
Sandhya Aneja, USA
Mohd Dilshad Ansari, India
Eva Antonino-Daviu , Spain
Mehmet Emin Aydin, United Kingdom
Parameshchhari B. D. , India
Kalapraveen Bagadi , India
Ashish Bagwari , India
Dr. Abdul Basit , Pakistan
Alessandro Bazzi , Italy
Zdenek Becvar , Czech Republic
Nabil Benamar , Morocco
Olivier Berder, France
Petros S. Bithas, Greece
Dario Bruneo , Italy
Jun Cai, Canada
Weiwei Cai, USA
Xuesong Cai, Denmark
Gerardo Canfora , Italy
Rolando Carrasco, United Kingdom
Vicente Casares-Giner , Spain
Luis Castedo , Spain
Brijesh Chaurasia, India
Chin-Ling Chen, Taiwan
Lin Chen , France
Xianfu Chen , Finland
Zhen Chen, China
Hui Cheng , United Kingdom
Hsin-Hung Cho, Taiwan
Ernestina Cianca , Italy

Marta Cimitile , Italy
Riccardo Colella, Italy
Mario Collotta , Italy
Massimo Condoluci , Sweden
Antonino Crivello , Italy
Laurie Cuthbert , Macau
Antonio De Domenico , France
Floriano De Rango , Italy
Antonio De la Oliva, Spain
Margot Deruyck, Belgium
Liang Dong , USA
Praveen Kumar Donta, Austria
Zhuojun Duan, USA
Mohammed El-Hajjar , United Kingdom
Oscar Esparza, Spain
Maria Fazio , Italy
Mauro Femminella , Italy
Manuel Fernandez-Veiga , Spain
Gianluigi Ferrari , Italy
Jesus Fontecha , Spain
Luca Foschini , Italy
Alexandros G. Fragkiadakis , Greece
Sabrina Gaito , Italy
Ivan Ganchev , Bulgaria
Óscar García, Spain
Manuel García Sánchez , Spain
L. J. García Villalba , Spain
Miguel Garcia-Pineda , Spain
Piedad Garrido , Spain
Vincent Gauthier , France
Carlo Giannelli , Italy
Michele Girolami, Italy
Edoardo Giusto , Italy
Mariusz Glabowski , Poland
Carles Gomez , Spain
Antonio Guerrieri , Italy
Barbara Guidi, Italy
Tao Han, USA
Sherief Hashima, Egypt
Mahmoud Hassaballah , Egypt
Yejun He , China
Danfeng Hong, Germany
Andrej Hrovat, Slovenia
Chunqiang Hu , China
Xuexian Hu , China

Yan Huang , USA
Zhenghua Huang, China
Xiaohong Jiang , Japan
Yanxiang Jiang , China
Vicente Julian , Spain
Omprakash Kaiwartya, United Kingdom
Rajesh Kaluri, India
Dimitrios Katsaros, Greece
Muhammad Asghar Khan, Pakistan
Rahim Khan, Pakistan
Hasan Ali Khattak, Pakistan
Minseok Kim , Japan
Mario Kolberg , United Kingdom
Xiangjie Kong , China
Wen-Cheng Lai, Taiwan
Jose M. Lanza-Gutierrez, Spain
Pavlos I. Lazaridis , United Kingdom
Kim-Hung Le , Vietnam
Tuan Anh Le , United Kingdom
Xianfu Lei, China
Jianfeng Li , China
Peng Li , China
Wenjuan Li , Hong Kong
Xiangxue Li , China
Xingwang Li , China
Yaguang Lin , China
Zhi Lin, China
Liu Liu , China
Mingqian Liu , China
Xin Liu , China
Zhi Liu, Japan
Miguel López-Benítez , United Kingdom
Changqing Luo, USA
Chuanwen Luo , China
Lu Lv, China
Basem M. ElHalawany , Egypt
Imadeldin Mahgoub , USA
Rajesh Manoharan , India
Andrea Marin, Italy
Davide Mattera , Italy
Michael McGuire , Canada
Weizhi Meng , Denmark
Klaus Moessner , United Kingdom
Simone Morosi , Italy
Amrit Mukherjee, Czech Republic

Shahid Mumtaz , Portugal
Giovanni Nardini , Italy
Tuan M. Nguyen , Vietnam
Petros Nicopolitidis , Greece
Xin Ning , China
Rajendran Parthiban, Malaysia
Giovanni Pau , Italy
Matteo Petracca , Italy
Nada Y. Philip, United Kingdom
Marco Picone , Italy
Daniele Pinchera , Italy
Giuseppe Piro , Italy
Sara Pizzi , Italy
Javier Prieto , Spain
Rüdiger C. Pryss , Germany
Umair Rafique, Finland
Maheswar Rajagopal, India
Sujan Rajbhandari , United Kingdom
Dharmendra Singh Rajput , India
Rajib Rana, Australia
Luca Reggiani , Italy
Daniel G. Reina , Spain
Bo Rong , Canada
Mangal Sain , Republic of Korea
Jose Santa , Spain
Praneet Saurabh, United Kingdom
Hans Schotten, Germany
Patrick Seeling , USA
Muhammad Shafiq , China
Alireza Shahrabi , United Kingdom
Zaffar Ahmed Shaikh , Pakistan
Vishal Sharma , United Kingdom
Kaize Shi, Australia
Chakchai So-In, Thailand
Enrique Stevens-Navarro , Mexico
Zhou Su, Japan
Sangeetha Subbaraj, India
Yi Sun , China
Tien-Wen Sung, Taiwan
Ville Syrjälä , Finland
Suhua Tang , Japan
Pan Tang , China
Pierre-Martin Tardif , Canada
Sreenath Reddy Thummaluru, India
Mauro Tortonesi, Italy

Federico Tramarin , Italy
Tran Trung Duy , Vietnam
Fan-Hsun Tseng, Taiwan
Reza Monir Vaghefi, USA
Juan F. Valenzuela-Valdés , Spain
Lorenzo Vangelista, Italy
S Velliangiri, India
Quoc-Tuan Vien , United Kingdom
Enrico M. Vitucci , Italy
Shaohua Wan , China
Ding Wang , China
Huaqun Wang , China
Pengfei Wang , China
Yingjie Wang, China
Lifei Wei , China
Miaowen Wen , China
Dapeng Wu , China
Huaming Wu , China
Liang Wu, China
Ding Xu , China
YAN YAO , China
Jie Yang, USA
Long Yang , China
Qiang Ye , Canada
Changyan Yi , China
Ya-Ju Yu , Taiwan
Marat V. Yuldashev, Finland
Sherali Zeadally, USA
Hong-Hai Zhang, USA
Jie Zhang, United Kingdom
Jiliang Zhang, China
Lei Zhang, Spain
Wence Zhang , China
Yushu Zhang, China
Kechen Zheng, China
Xu Zheng, USA
Fuhui Zhou , USA
Meiling Zhu, United Kingdom
Zhengyu Zhu , China






Contents

Innovative Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Tien-Wen Sung , Chao-Yang Lee , Tarek Gaber , and Hamed Nassar 

Editorial (3 pages), Article ID 9870345, Volume 2023 (2023)

Industrial Internet of Things Intrusion Detection Method Using Machine Learning and Optimization Techniques

Tarek Gaber , Joseph B. Awotunde , Sakinat O. Folorunso , Sunday A. Ajagbe , and Esraa Eldesouky 






Research Article (15 pages), Article ID 3939895, Volume 2023 (2023)

Utilizing Artificial Intelligence and Lotus Effect in an Emerging Intelligent Drone for Persevering Solar Panel Efficiency

Faris A. Almalki , Amani A. Albraikan , Ben Othman Soufiene , and Obaid Ali 




Research Article (12 pages), Article ID 7741535, Volume 2022 (2022)

Performance Optimization of 3-DOF Application Scene Based on 360-Degree Panoramic Technology with Depth Information

Qinghua Wang , Linjuan Ma , Xiaotong Huang , Lan Lyu , and Fuquan Zhang 

Research Article (8 pages), Article ID 3655934, Volume 2022 (2022)

Artificial Intelligence of Things-Based Optimal Finite-Time Terminal Attractor and Its Application to Maximum Power Point Tracking of Photovoltaic Arrays in Smart Cities

En-Chih Chang , Chun-An Cheng , and Rong-Ching Wu 


Research Article (9 pages), Article ID 4213217, Volume 2022 (2022)

A Study on the Optimization Simulation of Big Data Video Image Keyframes in Motion Models

Jianbang Guo, Peng Sun , and Sang-Bing Tsai 




Research Article (12 pages), Article ID 2508174, Volume 2022 (2022)

Image Real-Time Detection Using LSE-Yolo Neural Network in Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Zheng Zhi-Xian and Fuquan Zhang 





Research Article (8 pages), Article ID 2608798, Volume 2022 (2022)

MidSiot: A Multistage Intrusion Detection System for Internet of Things

Nguyen Dat-Thinh , Ho Xuan-Ninh , and Le Kim-Hung 





Research Article (15 pages), Article ID 9173291, Volume 2022 (2022)

Real-Time 3D Pedestrian Tracking with Monocular Camera

Peng Xiao , Fei Yan , Jiannan Chi , and Zhiliang Wang 

Research Article (18 pages), Article ID 7437289, Volume 2022 (2022)

Gene Selection and Classification of scRNA-seq Data Combining Information Gain Ratio and Genetic Algorithm with Dynamic Crossover

Junhong Feng , Xishuan Niu , Jie Zhang , and Jian-Hong Wang 



Research Article (16 pages), Article ID 9639304, Volume 2022 (2022)

Enhanced Intelligent Smart Home Control and Security System Based on Deep Learning Model

Olutosin Taiwo , Absalom E. Ezugwu , Olaide N. Oyelade, and Mubarak S. Almutairi 

Research Article (22 pages), Article ID 9307961, Volume 2022 (2022)

Analysis and Research on Digital Reading Platform of Multimedia Library by Big Data Computing in Internet Era

Wanxia Zhang , Bo Liu, and Sang-Bing Tsai 

Research Article (10 pages), Article ID 5939138, Volume 2022 (2022)

Frequent-Pattern-Based Broadcast Scheduling for Conflict Avoidance in Multichannel Data Dissemination Systems

Chuan-Chi Lai , Yu-De Lin, and Chuan-Ming Liu 




Research Article (19 pages), Article ID 1863590, Volume 2021 (2021)

Stock Trading System Based on Machine Learning and Kelly Criterion in Internet of Things

Lili Chen , Lingyun Sun , Chien-Ming Chen , Mu-En Wu , and Jimmy Ming-Tai Wu 



Research Article (9 pages), Article ID 7632052, Volume 2021 (2021)

Research on a Power Grid Cascading Failure Prevention and Control Method considering WSN

Huiqiong Deng , Junyuan Wu , Jie Luo , Renwu Yan, Cheng Zhang, Peiqiang Li, Kuo-Chi Chang, and Rongjin Zheng

Research Article (12 pages), Article ID 9439977, Volume 2021 (2021)

The Hybrid Traffic Offloading Mode for Disaster-Resilient Communication Networks Based on User Mobility

Ang-Hsun Tsai  and Chung-Hsien Tsai 

Research Article (14 pages), Article ID 9403982, Volume 2021 (2021)





Rotated Black Hole: A New Heuristic Optimization for Reducing Localization Error of WSN in 3D Terrain

Qing-Wei Chai  and Jerry Wangtao Zheng 

Research Article (13 pages), Article ID 9255810, Volume 2021 (2021)

Editorial

Innovative Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Tien-Wen Sung ¹, **Chao-Yang Lee** ², **Tarek Gaber** ³, and **Hamed Nassar** ⁴

¹Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou, China

²Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

³School of Science, Engineering & Environment, University of Salford, Manchester, UK

⁴Department of Computer Science, Suez Canal University, Ismailia, Egypt

Correspondence should be addressed to Tien-Wen Sung; tienwen.sung@gmail.com

Received 1 September 2022; Accepted 1 September 2022; Published 8 June 2023

Copyright © 2023 Tien-Wen Sung 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 Internet of Things (IoT) consists of interconnected things with built-in and function-oriented sensors, essentially constituting a network of physical devices. These devices have the ability to gather measurement or observation data and then communicate or exchange data with each other by communication networks. IoT systems can be applied in various fields to improve human life, especially applications for smarter cities or smarter homes.

However, basic IoT systems have been unable to meet the requirements of a modern smart city or smart home that features various and complex functionality with hybrid communication networks. Modern applications of IoT systems must be assisted by powerful artificial intelligence (AI) technology to process and analyze big data and deal with the problems of finding an optimal solution, making the best decision, detecting events, and identifying objects. Artificial intelligence simulates natural intelligence as exhibited by humans or animals and can make the system capable of performing tasks without the assistance of humans and even perform tasks better than humans can. Modern AI technology usually utilizes evolutionary computation, nature-inspired algorithms, machine learning, or deep learning to solve the problems of optimization, decision making, event detection, and object identification. The integration of IoT systems and AI technology is very suitable for interconnected things to enhance intelligence, thus the artificial intelligence of things (AIoT),

which enables the establishment of innovative IoT systems and applications for the modern smart city and smart home.

This special issue aims to publish original and innovative research works that focus on challenging issues in the field of innovative AI-based IoT in smart cities and smart homes. After the review process for evaluating all submitted manuscripts, there are sixteen research papers accepted for publication in this special issue.

The paper titled “Industrial Internet of Things Intrusion Detection Method Using Machine Learning and Optimization Techniques” by T. Gaber et al. proposes a novel approach of intrusion detection for Industrial Internet of Things (IIoT). Two artificial intelligent algorithms, PSO and BA, are used for feature selection of network traffic. After applying the feature selection schemes, classification of malicious behaviors is performed with machine learning-based models for the IIoT-based network traffic. The experimental results show that the proposed approach obtained good performance.

The paper titled “Utilizing Artificial Intelligence and Lotus Effect in an Emerging Intelligent Drone for Persevering Solar Panel Efficiency” by F. A. Almalki et al. proposes a drone system with AI framework to help clean solar panels mounted on the top of buildings and enhance their efficiency. The evaluation shows the improvement of solar power efficiency by the proposed system.

The paper titled “Performance Optimization of 3-DOF Application Scene Based on 360-Degree Panoramic Technology with Depth Information” by Q. Wang et al. proposes an optimization scheme to improve the spatial realism of an indoor 3-DOF application scenario. The experimental results show that the scheme is great in terms of visual effect and performance.

The paper titled “Artificial Intelligence of Things-Based Optimal Finite-Time Terminal Attractor and Its Application to Maximum Power Point Tracking of Photovoltaic Arrays in Smart Cities” by E.-C. Chang et al. utilizes AIoT technique to obtain the maximum power output from photovoltaic (PV) arrays. The method can save energy and reduce carbon emissions for the development of smart cities. The simulation results show that the proposed PV array system actually yields promising performance.

In the paper titled “A Study on the Optimization Simulation of Big Data Video Image Keyframes in Motion Models” by J. Guo et al., the signal of athletic sports video image frames is processed and studied according to the technology of big data. It proposes a fuzzy kernel extraction scheme based on the low-rank theory and is robust in fuzzy video forgery detection. The experimental results show that the efficiency of fuzzy video detection is improved compared to traditional video forgery detection methods.

The paper titled “Image Real-Time Detection Using LSE-Yolo Neural Network in Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes” by Z.-X. Zheng and F. Zhang proposes a visual image real-time detection LSE-Yolo neural network which helps bring a healthy and comfortable life in smart cities and homes.

The paper titled “MidSiot: A Multistage Intrusion Detection System for Internet of Things” by D.-T. Nguyen et al. focuses on the security issue for IoTs. It presents a collaborative intrusion detection system (IDS) deployed at both Internet gateways and IoT local gateways. The evaluation results indicate the proposed scheme could detect seven common cyberattacks targeting IoT devices with an average accuracy of 99.68% and outperforms state-of-the-art IDSs.

The paper titled “Real-Time 3D Pedestrian Tracking with Monocular Camera” by P. Xiao et al. deals with the pedestrian tracking issue which is a popular research area in computer vision. It proposes a target tracking method with a short-time prediction function to solve continuous tracking and occlusion judgment. The evaluation results show the method can achieve high accuracy and high tracking speed.

The paper titled “Gene Selection and Classification of scRNA-seq Data Combining Information Gain Ratio and Genetic Algorithm with Dynamic Crossover” by J. Feng et al. focuses on the single-cell RNA sequencing technology and presents a novel algorithm to address the gene selection and classification for scRNA-seq data. It utilizes information gain ratio to eliminate irrelevant genes roughly and utilizes genetic algorithm with dynamic crossover to choose high quality genes. The experimental results show it is superior to the other several competing algorithms in terms of classification accuracy.

The paper titled “Enhanced Intelligent Smart Home Control and Security System Based on Deep Learning Model” by O. Taiwo et al. focuses on the control and security system of smart homes. It presents a deep learning-based intelligent

home automation system for controlling home appliances, monitoring environmental factors, and detecting movement in the home and its surroundings. An experimental prototype for surveillance was implemented using an ESP32 camera in this study.

The paper titled “Analysis and Research on Digital Reading Platform of Multimedia Library by Big Data Computing in Internet Era” by W. Zhang et al. focuses on the analysis of digital reading platforms of multimedia library in China. Fourteen prefecture-level public libraries and fifty-eight libraries of higher education institutions in each region were accessed. It summarizes the problems found in the research and proposes solutions for the regional digital reading platform.

The paper titled “Frequent-Pattern-Based Broadcast Scheduling for Conflict Avoidance in Multichannel Data Dissemination Systems” by C.-C. Lai et al. considers the channel switching time and identifies the data conflict issue in an on-demand multichannel dissemination system. It models the considered problem as a data broadcast with conflict avoidance problem and proposes a frequent-pattern-based broadcast scheduling scheme to avoid data conflicts when assigning data items to time slots in the channels. The simulation results show that the proposed scheme can shorten the average access time compared with the existing heuristic methods.

The paper titled “Stock Trading System Based on Machine Learning and Kelly Criterion in Internet of Things” by L. Chen et al. proposes an AI-based stock trading system. A long short-term memory neural network is used to study stock price fluctuations, as well as genetic algorithms are used to obtain appropriate trading signals in the system. The Kelly criterion is used to determine the best investment score to control the risk of the transaction. The experiments show that the use of Kelly criterion for fund management reduces the risk of trading, and the return is higher.

The paper titled “Research on a Power Grid Cascading Failure Prevention and Control Method considering WSN” by H. Deng et al. proposes a WSN-based preventive control model to prevent cascading failures in the power grid. The nondominated sorting genetic algorithm II and particle swarm optimization are utilized to solve the issue. The method is verified in the IEEE39 node system.

The paper titled “The Hybrid Traffic Offloading Mode for Disaster-Resilient Communication Networks Based on User Mobility” by A.-H. Tsai and C.-H. Tsai provides a hybrid traffic offloading mechanism combining device-to-device (D2D) and local IP access (LIPA) modes for the disaster-resilient network. The method can prevent the local communication traffic from flowing into the core network and improve the system spectrum efficiency when the core network is under congestion.

The paper titled “Rotated Black Hole: A New Heuristic Optimization for Reducing Localization Error of WSN in 3D Terrain” by Q.-W. Chai and J. W. Zheng proposes the rotated black hole (RBH) algorithm which can improve the global search ability of the original black hole (BH) algorithm. The proposed RBH is also applied in reducing the error of position estimation of WSN unknown nodes in 3D terrain. The simulation results show that the proposed algorithm has better

search performance than other famous algorithms and has a good effect the localization problem of WSN in 3D terrain.

*Tien-Wen Sung
Chao-Yang Lee
Tarek Gaber
Hamed Nassar*

Conflicts of Interest

The guest editors declare that they have no conflicts of interest regarding the publication of this special issue.

Research Article

Industrial Internet of Things Intrusion Detection Method Using Machine Learning and Optimization Techniques

Tarek Gaber ^{1,2} Joseph B. Awotunde ³ Sakinat O. Folorunso ⁴ Sunday A. Ajagbe ⁵
and Esraa Eldesouky ^{6,7}

¹Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt

²School of Science, Engineering, and Environment, University of Salford, Manchester M5 4WT, UK

³Department of Computer Science, Faculty of Information and Communication Sciences, University of Ilorin, Ilorin 240003, Nigeria

⁴Department of Mathematical Science Olabisi Onabanjo University, Ago-Iwoye, 120107, Nigeria

⁵Department of Computer & Industrial Production Engineering, First Technical University, Ibadan, 200255, Nigeria

⁶Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

⁷Department of Computer Science, Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt

Correspondence should be addressed to Tarek Gaber; t.m.a.gaber@salford.ac.uk

Received 9 April 2022; Revised 29 August 2022; Accepted 30 September 2022; Published 30 April 2023

Academic Editor: Basem M. Elhalawany

Copyright © 2023 Tarek Gaber 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 emergence of the Internet of Things (IoT) has witnessed immense growth globally with the use of various devices found in home, transportation, healthcare, and industry. The deployment and implementation of the IoT paradigm in industrial settings lead to the architectural changes of Industrial Automation and Control Systems (IACS) plus the countless connectivity of industrial systems. This resulted in what is referred to as the Industrial Internet of Things (IIoT), which removes the barrier of connecting IACS to isolated conventional ICT platforms. In recent times, the IoT has started hacking our personal lives and not only our world, thus creating a platform for impending IoT cyberattacks. The widespread use of the IoT has created a rich platform for possible IoT cyberattacks. Machine learning (ML) algorithms have been driven solutions to secure wireless communication in IIoT-based systems, and their use in solving various cybersecurity challenges. Therefore, this paper proposes a novel intrusion detection model based on the Particle Swarm Optimization (PSO) and Bat algorithm (BA) for feature selection, and the Random Forest (RF) classifier for the classification of malicious behaviors in IIoT-based network traffic. An IIoT-based cybersecurity dataset, WUSTL-IIOT-2021 Dataset, was used to evaluate the performance of the proposed model using accuracy, recall, precision, and F1-score. The results of the two feature selection were compared to identify the most promising one. The results were compared with other recent state-of-the-art ML and multiobjective algorithms, and the results showed better performance. The RF along with BA classifier had proved to be the best classifier.

1. Introduction

The emergence of the Internet of Things (IoT) paradigm in Industrial Automation and Control Systems (IACS) is termed the Industrial Internet of Things (IIoT), and in recent years it has become very popular. The IACS have been utilized in recent time to keep an eye on industrial machines and processes, and thus the IIoT-based systems have become an essen-

tial part of every critical infrastructure in smart industries. The largest parts of these systems are the data acquisition and supervisory systems that repeatedly manage the IACSs. Real-time monitoring, interaction with the devices, analysing of data, and logging all the events that happen in the systems are the main roles of these systems. Hence, the arrival of the IoT paradigm in these systems enriches the security and network intelligence in the computerization and optimization of

industrial processes. Since the operations of IIoTs lead to a huge amount of data, and the majority of the applications are mission-critical and demand high availability, there is need for cyber-security to properly secure these systems.

Isolating IACSs from the outside world in the past has really helped to secure IACSs from intrusion and malicious external attack [1]. The recent improvements and usage of Internet communication with increased connectivity to transmit information have created more avenues for cyber-attacks like Denial-of-Service attack, Man-in-the-Middle (MITM) attack, Phishing Attack, Password Attack, SQL Injection Attack, and Cryptojacking against these systems [2, 3]. Cyber-attacks have a number of detrimental repercussions. When an attack is attempted, it may result in data breaches, which may cause data loss or manipulation. Companies suffer financial losses, a decrease in customer trust, and damaged reputations. In order to prevent or stop cyber-attacks, a cyber-security can use measures like IDS and antivirus to preventing unwanted digital access to networks, computer systems, and their parts. Hence, security is the most concerning issue in IIoT-based systems due to the sensitive nature of the industrial application.

To provide a secure environment, an intrusion detection system (IDS) has been an integral part of IIoT-based applications since the intrusion of crucial security concerns in 2010, the Stuxnet worm was exposed [4], and in December 2017, the attack reappeared with another powerful malware called Triton against the IACSs [5]. These attacks give rise to the awareness of the necessity to pay attention to the protection of these vital infrastructures' security [3]. The fundamental difference between regular information technology systems, and the IACSs necessities their priorities to secure common vulnerabilities, and in most cases their attacks are different [6]. Additionally, IACS traffic and data type are specifically different using certain IIoT communication protocols like Distributed Network Protocol 3 (DNP3), Building Automation Controls Network (BACnet), and Modbus [7]. Hence, with these special reasons, the security of IIoT-based applications must be properly considered when it comes to the designing of an IDS for IACSs.

The continuous growth of IoT-based systems and their related applications demands the improvement of network security and to maintain the security of any interconnected system that requires protection of its integrity, availability, and confidentiality [8]. The most common IIoT-based system threats that interrupt and attempt to terminate the integrity, availability, and/or confidentiality are cybersecurity and intrusions. IDS applications include the hardware devices or software services that monitor the network for malicious activities. The network intrusion detection system (NIDS) plays a prominent role in addressing various Internet attacks, and the IIoT has been identified as an integral part of the present machinery for industrial data transfer, necessitating the need for network security. The NIDS are used to safeguard the workstation structures from network intrusion and multiple grid invasions. Recent work has created new IDSs in response to the attacks and threats posed by various aggressive frameworks. However, the performance of current machine learning-based methods in terms of accuracy and high false alarm rate are still issues that need

urgent attention in order to reduce the irregularity of discovery methods of intrusion and malicious attacks.

Recently, feature selection has been identified as a modern method of getting an accurate and low false alarm rate in NIDSs [9, 10]. This method is used to select the most useful and fit features for a better classification result in NIDS models. This led to aggregated accuracy performance and a reduced error rate in their applications for detection of attackers [9]. Additionally, the datasets of features are very huge, and not all are always useful for the classification of the dataset as either normal or abnormal. Hence, the use of feature selection techniques is very necessary. The use of feature selection is, therefore, very important in the use of NIDSs in IIoT-based network traffic, helping in getting optimal results from the classification models used.

ML-based models have been previously used in securing IT-based systems [11–13] and IoT-based networks, but the suitability of these models has not been widely employed and still remains debatable according to authors in [12]. The ability to detect any penetration into the system is the main security concern of the IIoT-based devices. Sometimes the ML-based models of IDSs for IACS may not be able to properly detect the attack due to its design, which may not address the imbalance of the data, which is the main property of intrusion detection problems [14]. Hence, for better IDSs performance, the issue of imbalanced datasets for IIoT-based systems should be considered, addressing the questions of what are the true boundaries and how do various performance metrics react to them?

The application of the ML-based model for the IIoT-based network still faces various challenges, as given in the following:

- (i) *The Issue of Low Processing Ability.* The IoT-based devices have energy constraints with limited processing capacity due to their small size. This creates huge challenges since ML-based models require real-time processing of data, thus their implementation in such resource-constrained environments creates issues
- (ii) *Data Analytics.* Data is generated heterogeneously in the IoT environment and demands preprocessing before being applied to an ML-based model. This necessitates the processing memory space and power of IoT-based devices, making the provision of an efficient solution a challenge for diverse data

Inspired by the aforementioned challenges, the assumption of working on the imbalanced datasets by turning them into class balanced datasets, the aim of this paper is to design an efficient and yet accurate intrusion detection method for IIoT applications. Bioinspired optimizations (i.e., PSO [15] and BAT [16]) were used to get a subset of features helping to achieve this aim. Also, Random Forest (RF), k-Nearest Neighbor (k-NN), and MultiLayer Perceptron (MLP) classifiers were employed to measure the performance in terms of accuracy, precision, recall, F1-score, and ROC. Also, the experiments were conducted on a dataset (i.e., the WUSTL-IIOT-2021 Dataset) which was collected specifically for IIoT cybersecurity threats and attacks.

1.1. The Study Has the following Significant Contributions

- (i) Proposing an intrusion detection method for IIoT applications using bioinspired-based feature selection to enhance the performance of the intrusion detection system through reducing the number of the selected features while getting a high accuracy
- (ii) Investigating the effectiveness of the proposed feature selection method above by different types of machine learning algorithms (i.e., RF, k-NN and MLP). This was done with a relevant dataset, WUSTL-IIOT-2021 dataset, which was collected in IIoT environment
- (iii) Providing a thorough evaluation through two phases: (1) using the benchmark evaluation metrics of accuracy, precision, recall, F1-score, and ROC and (2) comparing the obtained results with the most related published work which showed better results for the proposed method

The rest of the paper is organized as follows: Section 2 presents the literature review on ML-based models for intrusion detection on IIoT networks. Section 3 explains the methodology employed in this study. Section 4 presents the experimental results of the study, while Section 5 concludes the paper with future direction.

2. Related Work

The IIoT idea was created specifically for application in modern industry. Modern IIoT refers to the application of the standard IoT in various industrial projects and businesses. Numerous actuators, sensors, control systems, interfaces for communication and integration, cutting-edge security systems, networks for automobiles, household appliances, etc., are all included in the IIoT. The IIoT's nodes can all connect to the Internet. The capacities of many sectors, manufacturing facilities, asset management systems, sophisticated logistics systems, etc., have been substantially improved by the use of IIoT in contemporary businesses. Several applications, gadgets, and services can connect the real area to a virtual one thanks to the IIoT [17].

There are various ways for IIoT nodes to connect to the Internet, including through the use of Message Queue Telemetry Transport (MQTT), Modbus TCP, cellular networks, Long-Range Radio Wide Area Network (LoRaWAN), and other TCP/IP-based communication protocols [18]. The majority of IIoT nodes can also gather, process, and transfer data. Due to their capabilities, they are vulnerable to several privacy and security risks that could endanger IIoT systems and the applications they are a part of [19]. The fact that IIoT nodes are constantly active while carrying out data collecting, processing, and transmission is one of their major characteristics.

The perception layer, the network layer, the application layer, and the Cloud are the three main layers of the IIoT. These levels are founded on data flow. Additionally, each layer is vulnerable to different kinds of assaults and breaches that could jeopardize the IIoT systems. Access control breaches,

data corruption incidents, spoofing assaults, Distributed DoS, Operating System (OS) attacks, and jammer attacks are some frequent attacks and intrusions on the IIoT ecosystem. Many firms are employing intrusion detection systems to prevent these malicious assaults, ensuring that IIoT networks' security and active IIoT nodes' security are maintained (IDSs). Additionally, these IDSs can be set up at any layer.

There have been various approaches to solving the problem of identifying intrusions like ML-models, ensemble methods, deep learning methods, and the hybrid approaches enabled by feature selection [20, 21]. Through the analysis of collected information, the NIDS can detect attacks from various network traffic and systems [22]. Hence, the approach is widely used as a technique for network security. Various research have used both ML and DL methods for the purpose of intrusion detection in various environments like the World Wide Web, IoT-based systems, and Internet network traffic for the purpose of detecting and categorizing attacks such as [23, 24], among others. In recent time, ML and DL techniques, like SVM, RBM [25], Conventional Neural Network (CNN) [26], Artificial Neural Network (ANN) [27], Decision Tree (DT) and Random Tree (DT) [28], and clustering and K-NN algorithms [29], have been used for improving intrusion detection systems. The advantages of the ML-based IDS model are as follows:

- (i) The ML-based models can efficiently detect attacks with small variations since they are trained based on the behavior/pattern of the network for most scenarios
- (ii) The use of unsupervised learning models can easily detect zero-day attacks, especially if the model is trained based on this method
- (iii) Even in complex network environments, ML-based IDS gives higher detection accuracy and is faster

Machine learning approaches have been shown to provide effective intrusion detection systems during the recent years. They produce better outcomes than other alternative methods since they are applicable to different types of datasets and can analyze real-time data. Researchers usually use various approaches, including deep learning, heuristics, adaptive learning, decision trees, and semisupervised learning.

Priya et al. [30] proposed a two-phase intrusion detection model that was developed that includes SVM, NB, and DT in the first phase and an RF classifier for prediction using ensemble learning. In addition, to deliver better predictions, the results of the ANN classifier were integrated with those of the RF. The combined model is validated against the WUSTL-IIOT-2018, N_BaIoT, and Bot_IoT datasets. According to the conducted results of applying only the first phase, the Naïve Bayes classifier had the lowest accuracy, followed by the SVM and DT classifiers, while the DT classifier achieved the highest accuracy of 96%. The proposed method, on the other hand, incorporated ANN and RF predictions and attained a 99 percent accuracy rate for all the three datasets. A deep learning strategy was used to address another IIOT intrusion detection model by Raja [31]. The

proposed DL-TL-NIDS model had two levels of detection. The DNN is trained and evaluated at the first level to detect current assaults. Attacks that had a poor detection or low accuracy rate were classified as challenging attacks. These challenging attacks are input to second-level detection, which trains the Negative Selection Algorithm (NSA) and DNN models using the Dragonfly algorithm. Finally, the outputs of both models are combined using Dempster Shafer's combination rule.

Nevertheless, using bioinspired algorithms to extract key IIoT network features can assist in reducing processing costs and memory use and make it easier to apply various classification approaches to the selected features. In this section, we present related works on managing intrusion detection in the IIoT that use bioinspired algorithms for feature selection.

Keserwani et al. [32] suggested a hybrid metaheuristic approach for feature selection and deep learning for classification to identify intrusions in a virtualized cloud network. A deep sparse auto-encoder is utilized to classify the important features from the cloud network connections, which are identified using hybrid Gray Wolf Optimization (GWO) and PSO. The authors expanded on their previous work in [20] to include fetch attacks in the IoT world. The hybrid GWO-PSO is also utilized to extract key IoT network properties, which are then fed into a random forest classifier for improved attack detection accuracy. The proposed model was tested on the KDDCup99, NSL-KDD, and CICIDS-2017 datasets, and it achieved an accuracy of 99.66%.

Kasongo [19] proposed an IDS for IIoT by employing the genetic algorithm along with a random forest model, which was utilized in the fitness function of the genetic algorithm. The usage of Genetic Algorithms (GA) is motivated by the presence of a large number of features in current datasets, as well as a large number of network traces. As a result, the ML algorithms' training process is badly impacted and misled, as ML performance decreases as the number of features grows. Hence, the learning process becomes more difficult as the dataset's number of characteristics rises. Therefore, the genetic algorithm is utilized to improve the feature selection, and the author used tree-based methods such as RF, DT, and ET algorithms for each attribute vector, all of which were tested on the UNSW-NB15 general-purpose dataset.

Awotunde et al. [3] utilized the same dataset, together with the NSL-KDD dataset, to build a hybrid rule-based feature selection technique. The proposed research combines a deep feedforward neural network model and rule-based feature selection with IIoT applications to obtain relevant data that may be utilized to construct an intelligent NIDS (i.e., data gathered from TCP/IP packets). This research presents a three-tier methodology for intrusion detection in IIoT systems, in which a rule-based model is utilized for feature selection and a genetic tool is employed to create the characteristics with the highest values. Finally, the selected features are loaded into the ANN for use in the learning process.

The authors in [33] employed the Aquila optimizer (AQU) for feature selection in the CIC2017, NSL-KDD, BoT-IoT, and KDD99 datasets to assess the quality of the proposed IDS approach. A light feature extraction strategy based on CNN was adopted to extract relevant features from

the datasets utilized in this work. Following that, the AQU algorithm is used to pick a group of the best features that shows the datasets properties.

The ML-based IDS has a lot of advantages, like being faster and more accurate in both simple and complex environments. Furthermore, owing to the training nature of ML models, particularly through unsupervised learning techniques, several types of assaults may be easily spotted. Yet, several challenges still remain when applying machine learning models to IIoT networks. The bulk of recent datasets are large in size, both in terms of feature space dimension and the number of network traces. The presence of a large number of features in a dataset might have a detrimental influence on the training process of machine learning algorithms. The performance of the ML-based IDS has therefore deteriorated since performing an effective learning process becomes more difficult as the number of characteristics in a dataset grows. In order to obtain the essential features, an accurate method of feature selection is required. Another issue is the lack of real-world data collected by an IIoT system in order to assess the efficacy of present solutions.

Another point, which is not well-addressed in the literature, is the imbalance of the dataset used in building ML-based intrusion detection systems. Because of the imbalanced datasets, minority attacks may be missed. Also, the IDS model can identify the majority of attacks, but due to the imbalance, certain attacks may not be detected. As a result, these attacks need a high level of detection. Table 1 shows the summary of the main findings in the reviewed literature.

The application of feature selection has been helped in the area of feature reduction to transform features from high dimensional to a lower dimensional space without reducing the efficiency of the prediction algorithms. This technique is used to eliminate irrelevant features and variables from any dataset without reducing the data's usefulness to the classification model.

From the literature review, there has not been any work that applies feature selection on WUSTL-IIoT-2021 datasets for IDS to the best of our knowledge, this study will be the first to apply feature selection for IIoT-IDS system while testing it using a specialized IIoT-based dataset which would simulate the real case scenario. Though, the baseline model has applied various ML techniques on the dataset. Hence, this study applied the feature selection to further enhance the accuracy performance of the ML-based models while minimizing the computational cost.

3. Materials and Methods

3.1. Proposed IIoT Intrusion Detection Method. The proposed system aims at enhancing the performance of NIDSs for IIoT-based networks using feature selection techniques on the dataset. In recent years' various techniques like data mining and ML techniques have been used to resolve various problems involving optimization system performance. To improve the performance of NIDS for IIoT-based networks, the proposed model reduces the number of features used for the classification problem. Figure 1 presents the architecture of the model that has been proposed. The stages of the

TABLE 1: Summary of the main findings of the related work.

Method name	Accuracy (%)	Precision	Recall	F-score	Feature selection	Dataset
GWO-PSO-RF NIDS model [20]	99.88%	0.67	0.72	0.69	Hybrid GWO-PSO	KDDCup99
	99.24%	0.93	0.97	0.95		NSL-KDD
	99.87%	0.94	0.87	0.89		CICIDS2017
	99.66% (avg)					
DL-rule based feature selection model [3]	99.0%	0.97	0.99	0.98	Rule based selection	NSL-KDD
	98.9%	0.99	0.99	0.97		UNSW-NB15
GA-RF model [19]	87.61%	0.98	0.81	0.89	Genetic algorithm	UNSW-NB15
AQU-CNN [33]	99.99%	0.99	0.99	0.99	Aquila optimizer (AQU).	CIC2017
	77.38%	0.84	0.77	0.77		NSL-KDD
	99.99%	0.99	0.99	0.99		BoT-IoT
	99.92%	0.94	0.92	0.93		KDD99
Attack detection using ensemble classifier [30]	83%	0.86	0.84	0.83	—	WUSTL_IIoT-2018
	87%	0.88	0.88	0.87		N_BaIoT
	87%	0.88	0.88	0.87		Bot_IoT
						(Naïve Bayes)
DL-TL-NIDS [31]	99.97%	0.995	0.95	0.996	—	TON IoT
	99.86%	0.997	0.998	0.998		CICIDS-2017
	99.97%	0.997	0.997	0.997		CICIDS-2018

proposed model were discussed in detail in the following subsection. The method consists of preprocessing, feature selection, and classification.

3.2. The Preprocessing Stage. To provide appropriate data for the proposed model framework for the model optimization, various preprocessing steps were performing on the WUSTL-IIOT-2021 dataset. The following are the steps followed to reform the dataset used for the purpose of this study:

- (i) *Removing Features.* Features that are unique to the attacks are removed after downloading the dataset ('StartTime', 'LastTime', 'SrcAddr', 'DstAddr', 'sIpId', 'dIpId'), therefore, if not removed, the model would not be universal for unseen data since they expose the type of the attack to the model. Also, the attack cannot be included as a feature, hence, it is very necessary to remove them, and the main objective is to reduce the features of the dataset before classification
- (ii) *Label Encoding.* The traffic label is given string value to specify the type of attack in which it belongs, hence, it is very necessary to change the value encoded into numerical values
- (iii) *Data Binarization.* The data collected in the collection spans a wide range of values. This data presents the classifier with a variety of obstacles during the training process in order to correct such differences. As a result, each feature's values must be standardized. As a result, the lowest value for each characteristic should be 0. The maximum value, however, should be 1. It improves the homogeneity of the classifier. It keeps the discrepancy amongst each feature's values

- (iv) *Addressing Imbalance Data.* This was handled using resampling without replacement with a 20% sample size model for the dataset before classification

3.3. Feature Selection. The importance of feature selection in improving the performance of NIDSs cannot be overstated because it also improves the performance of IDSs. This is due to the fact that intrusion detection involves a huge number of features that take a long time to process. Hence, feature selection is very important to increase the detection rate (DR) and decrease the detection time and false alarm rate. This problem can be solved using bioinspired optimization methods. As a result, the feature selection method influences the amount of time required to examine traffic behavior and enhance the overall performance of the model. It is very challenging to select the subset of features in any given dataset, and when the dimensionality of the feature is high, it cannot be managed efficiently. They can provide high-quality solutions in a fair amount of time and with considerable diligence [34]. Two bioinspired metaheuristic algorithms were used for the purpose of feature selection, namely, PSO [15] and BA methods [16].

3.3.1. Particle Swarm Optimization. One of the most stunning tourist attractions is a flock of birds in flight. Herds and other forms of organizations, such as plants and terrestrial animals, are fascinating to observe and consider organizational behavior. It includes a variety of birds, but the overall exercise is fluid. It is straightforward, but visually complex. It appears to be arranged at random. It is breathtaking. The feeling of deliberate and concentrated dominance is the most humiliating. Furthermore, all the data suggest that the flock's movement is solely the result of each bird's recognition of the area. Bird-like objects called boids are employed in the flocking model [35]. Each boid is known for what happens in its

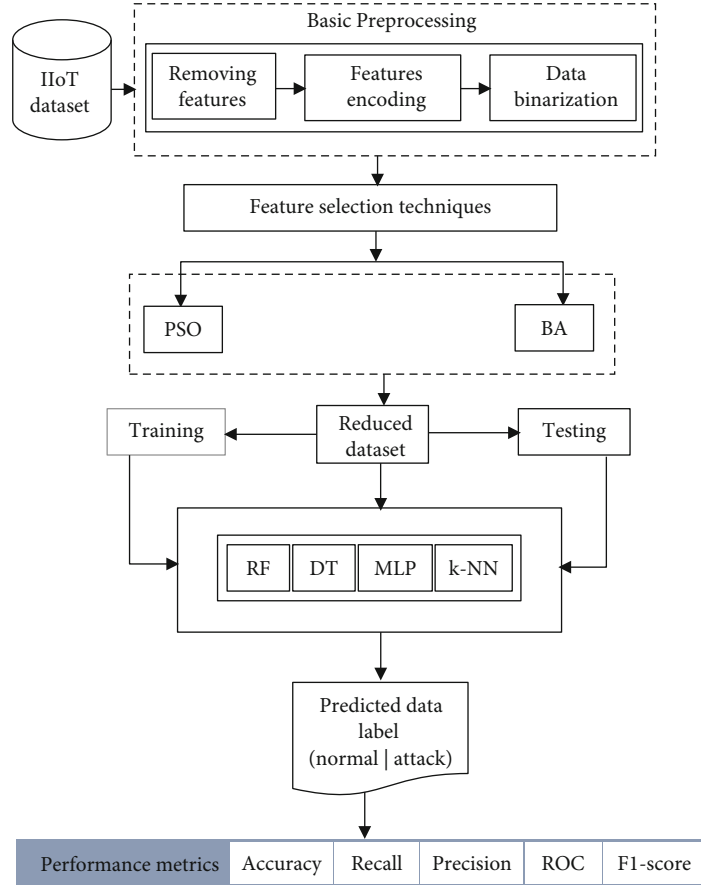


FIGURE 1: The proposed NIDS Architecture for IIoT.

immediate environs because of its position and speed. The three basic steering behaviors shown by boids are separation, alignment, and cohesion [36].

A PSO does not necessitate a thorough understanding of the situation, such as gradual changes [37]. It can be utilized when a problem requires access to data that is either unavailable or prohibitively expensive. Each particle's fitness score is determined in a swarm. The particle's best position is determined using a fitness score. Each particle's position indicates a potential solution to the optimization issue [38]. After that, the best global location among the particles is determined. It uses the best global and local locations to locate intriguing places for further research, as well as spots where all this information is shared with other particles, allowing particles to explore the solution space more effectively. It is a method of iterative optimization [39].

Each particle is defined by the original PSO formulas as a potential solution to a problem in N -dimensional space. Particle i 's position is denoted as $X_i = (x_{i1}, x_{i2}, \dots, x_{iN})$. Each component also remembers its prior optimal position, which is expressed as $P_i = (p_{i1}, p_{i2}, \dots, p_{iN})$. Because each particle in a swarm is rotating, it has a momentum, which may be expressed as $V_i = (v_{i1}, v_{i2}, \dots, v_{iN})$.

Among $pbest$, each particle knows its best value so far ($pbest$) and the best value in the group ($gbest$). This information is useful in determining how the particles in their

immediate vicinity have done. Using the following information, each particle tries to change its position:

- (i) the gap between $pbest$ where you are now and where you want to be
- (ii) the distance between where $gbest$ is now and where $gbest$ want to be

The notion of velocity can be used to illustrate this change. Each agent's velocity can be altered (3). Eberhart and Shi were the first to mention the incorporation of an inertia weight in the PSO algorithm in the literature [40]. Consider

$$V_{id} = w \times V_{id} + c_1 \times \text{rand}() \times (P_{id} - X_{id}) + c_2 \times \text{rand}() \times (p_{gd} - X_{id}), \quad (1)$$

where the index of the particle i is, $i \in \{1, \dots, n\}$, N population size, d dimension, $d \in \{1, \dots, N\}$, $\text{rand}()$ is uniformly distributed random variable between 0 and 1, V_{id} : velocity of particle i on dimension d , X_{id} current position of particle on dimension d , c_1 establishes the relative importance of the cognitive process, the factor of self-confidence, and the factor of motivation, c_2 defines the social component's proportionate influence, swarm confidence factor, P_{id} personal best or $pbest$

of particle i , p_{gd} global best or $gbest$ of the group, and w inertia weight.

The following equation can be used to change the existing position in the solution space, which is the searching point:

$$V_{id} = X_{id} + V_{id}. \quad (2)$$

Because all swarm particles tend to move towards better positions, the best position (i.e., optimum solution) can finally be attained by combining the efforts of the entire population. PSO is a basic, easy-to-implement, and computationally efficient method.

3.3.2. BAT Algorithm. This was created using the key concept of frequency tuning based on microbat echolocation. The echolocation features of microbats can be idealized as the following three rules in the typical bat algorithm:

All bats utilize echolocation to gauge distance, and in some mysterious way, they also 'know' the distance between food/prey and backdrop barriers.

Bats look for prey by flying at a random velocity v_i at position x_i with a fixed frequency f_{\min} , changing wavelength, and loudness A_o . Depending on the closeness of their target, they may automatically modify the wavelength (or frequency) of their radiated pulses as well as the rate of pulse emission $r \in [0, 1]$.

Even though loudness can change in a variety of ways, we assume that it ranges from a high (positive) A_o to a low (constant) A_{\min} .

The virtual bats require the following initialization parameters: the d -dimensional search space, position x_i , velocity v_i , and frequency f_1 . The following are the update rules for the new solution x_i^j and velocity v_i^j in each step t :

$$f_i = f_{\min} + (f_{\min} - f_{\max})\beta, \quad (3)$$

$$v_i^j = v_i^j(t-1) + [\hat{x}^j - x_i^j(t-1)]f_i, \quad (4)$$

$$x_i^j(t) = x_i^j(t-1) + v_i^j, \quad (5)$$

where $\beta \in [0, 1]$ denotes a uniformly distributed random vector. We know that the variable f_i is utilized to change the velocity and that the variable $x_i^j(t)$ represents the value of the position j for the bat i at the step t based on Equations (3), (4), and (5). The variable x^j denotes the current global best position, which is determined by comparing all of the m bats' answers.

Song and Gorla used a random walk technique for each bat to prevent them from falling into the local extremum and to boost their random searching ability [41]. Following the selection of a solution from the current best position, the random walk is used to generate a new solution for each bat, as described in

$$x_{\text{new}} = x_{\text{old}} + \epsilon \bar{A}(t), \quad (6)$$

where $\epsilon \in [-1, 1]$ is a random number that controls the walk's

direction and stride, and $\bar{A}(t)$ is the average volume of all bats in the step t .

In addition, according to Equation (7), the loudness A_i and the pulse rate r_i are updated for each step in Equation (5). When the prey is discovered, the loudness A_i is normally reduced and the pulse rate r_i is raised. For added convenience, the volume can be modified to any value.

$$A_i(t+1) = \alpha A_i(t), \quad (7)$$

$$r_i(t+1) = r_i(0)[1 - \exp(-\gamma t)], \quad (8)$$

where α and γ are both constants. The loudness $A_i(0)$ and the pulse rate $r_i(0)$ are normally chosen at random in the first phase of the bat algorithm. In general, $A_i(0) \in [1, 2]$ and $r_i(0) \in [0, 1]$ are set.

3.4. The Classifiers Models

3.4.1. Random Forest. This Bagging classifier uses a technique known as bootstrap aggregation, which is a form of ensemble technique. A number of different basic models (M_1, M_2, \dots, M_n) are blended. Using row sampling with replacement, distinct samples of records are delivered to each model. Some records may be repeated in the samples delivered to the models when row sampling with replacement is used. The voting classifier is used to combine the model outputs in order to make a judgment. A random forest is a bagging classifier in which numerous decision trees are utilized as models. Row and column sampling are used to provide input to each decision tree. The difficulty with the decision tree is that it has a low bias and a big variance. This indicates that the tree performs better in the training phase but poorly in the testing phase. The voting strategy lowers variance from high to low since the decision is based on the voting of numerous trees rather than a single tree [42].

3.4.2. Multilayer Perceptron (MLP). This is a type of ANN that feeds back information. The name MLP is confusing, referring to networks built of multiple layers of perceptrons (with threshold activation) in some cases and any feedforward ANN in others [43]. Multilayer perceptrons, especially those with a single hidden layer, are commonly referred to as "vanilla" neural networks [44]. There are at least three levels of nodes in an MLP: an input layer, a hidden layer, and an output layer. Each node, with the exception of the input nodes, is a neuron with a nonlinear activation function. Backpropagation is a supervised learning technique used by MLP during training. MLP is distinguished from a linear perceptron by its numerous layers and nonlinear activation. It can distinguish between data that is not linearly separable and data that is nonlinearly separable. If all of the neurons in a multilayer perceptron have a linear activation function, that is, a linear function that maps the weighted inputs to each neuron's output, then linear algebra shows that any number of layers may be reduced to a two-layer input-output model. In some MLP neurons, a nonlinear activation function is used that was made to model how often biological neurons fire or send out action potentials or pulses [45].

3.4.3. K-NN Algorithms. In classification and regression issues, the K-NN algorithm is used. It is a supervised learning technique that classifies an unknown instance based on the distance between the instance and k selected neighbors, with the class determined by the majority of neighbors voting [46]. The K-NN algorithm is frequently used in classification, with the goal of classifying new objects based on attributes and training examples. The K-NN technique is a classification approach based on learning data that is closest to the object. The K-NN algorithm is frequently used in classification, with the goal of classifying new objects based on attributes and training examples. The K-NN technique is a classification approach based on learning data that is closest to the object. This area is divided into divisions based on the training data's class label. A point in this space is designated as c class; if class c is the most frequently occurring point at k, then c is the correct answer. The Euclidean distance is used to determine how close or remote neighbors are [47].

4. Results and Discussion

4.1. The Dataset. The dataset used for the purpose of this study is WUSTL-IIoT-2021. This dataset consists of network data of IIoT-based systems that can be used for cybersecurity research. The dataset was captured using the IIoT testbed and presented by the authors in [48]. The goal of this testbed is to mimic real-world industrial systems as closely as possible while also allowing for real-world cyber-attacks. A total of 2.7GB of data was collected, spending about 53 hours. There are levels of preprocessing to clean the dataset by removing the rows with missing values, extreme outliers, and invalid entries resulting from corrupted values. After the preprocess stages, the final version is a little over 400 MB and can be used for the purpose of an intrusion detection experiment. Table 2 shows the statistics of the dataset.

The average data rate was 419 kbit/s, and the average packet size was 76.75 bytes, as shown in Table 3. This was purposefully focused around 90% of the attacks to DoS attacks because they are typically high in traffic and number of samples. Other forms of attacks are less common, and when they do occur, they simply convey a small amount of traffic data.

4.2. Evaluation Metrics. To assess the performance of the proposed model, the metrics in Equations employ many features, namely true positive (tp), false positive (fp), true negative (tn), and false negative (fn) [1]. The confusion matrix is a table that calculates the metric features as illustrated by Table 4 that estimates the true positive rate (TPR), false negative rate (FNR), true negative rate (TNR), and false positive rate (FPR). The main model assessors in Equations were derived from the table.

TPR is the ratio of class a instances correctly classified as class a as shown by

$$\text{TPR} = \text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}. \quad (9)$$

TABLE 2: The dataset characteristics.

Dataset	WUSTL-IIoT
Number of observations	1,194,464
Number of features	41
Number of attack samples	87,016
Number of normal samples	1,107,448

TABLE 3: Statistical information of the traffic types in our developed dataset.

Traffic's type	Percentage (%)
Normal traffic	92.72
Total attack traffic	7.28
Command injection traffic	0.31
DoS traffic	89.98
Reconnaissance traffic	9.46
Backdoor traffic	0.25

TABLE 4: Confusion matrix.

		Predicted Class	
		Class a	Class b
Real Class	Class a	tp	fn
	Class b	fp	tn

TNR is the ratio of class b instances correctly classified as class b as shown by

$$\text{TNR} = \frac{\text{tn}}{\text{tn} + \text{fp}}. \quad (10)$$

FPR is the ratio of class a instances incorrectly classified as class b as shown by

$$\text{FPR} = \frac{\text{fp}}{\text{fp} + \text{tn}}. \quad (11)$$

FNR is the ratio of class b instances incorrectly classified as class a as shown by

$$\text{FNR} = \frac{\text{tp}}{\text{tp} + \text{fn}}. \quad (12)$$

Accuracy is the percentage of correctly classified instances as presented by

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}}. \quad (13)$$

Precision is the ratio of the number of correct decisions made as shown by

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}. \quad (14)$$

Sensitivity is ratio of the number of tp by the number of all of the positive evaluations as shown by Equations ((15a) and (15b))

$$\text{Sensitivity} = \frac{\text{tp}}{\text{tp} + \text{fn}}. \quad (15a)$$

$$\text{Specificity} = \frac{\text{tn}}{\text{tn} + \text{fp}}. \quad (15b)$$

The F1-Score is the harmonic mean between the recall and precision as illustrated by

$$F1_{\text{Score}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (16)$$

Geometric Mean is the square root of the product of sensitivity and specificity as shown by

$$\text{Geo Mean} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \quad (17)$$

ROC shows the tradeoff between TPR and FPR as shown by

$$\text{ROC} = \frac{1 + (\text{TPR} - \text{FPR})}{2}. \quad (18)$$

PRC shows the trade-off between precision and recall for different threshold. A high PRC value shows both high recall and precision. It is a useful assessor especially when the classes are imbalanced.

Logistic Loss (Log Loss) measure the classification model performance based on the predicted probabilities of the real class. This value increases as the probability diverges from the real label. So, the lower the value, the better the performance of the model. The formula for Log Loss for multiclass classification is shown by

$$\text{LogLoss} = - \sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (19)$$

Where M is the number of labels, \log is the natural log, y is the class label, p is the predicted probability observation of o is of class c .

4.3. Feature Selection Schemes Results. All experiments were performed on a i7-8750H CPU @ 2.20GHz, 32 GB RAM and Windows 11 Pro. system. The study dataset was split into 80-20 of train-test ration. The general parameters used by all feature selection scheme (FSS) are k (4) which is the k -value in K-NN, the number of particles ($N = 10$) and the maximum number of iterations ($T = 30$). After application of the feature selection scheme, DstPkts, SrcBytes, DstLoss, pLoss, TcpRtt, IdleTime, and TotAppByte features were common to PSO and BA. Figures 2(a) and 2(b) show rate of convergence of the fitness function of PSO and BA.

Table 5 presents the optimum features that were selected for efficient classification performance by feature selection schemes considered in this study for detecting the attacks.

Figures 2(a) and 2(b) shows the rate of convergence of the fitness function of PSO and BA for the study dataset. For PSO, the convergence happens at the 8th iteration and the best fitness value is 0.00458. PSO started with the highest fitness value of 0.0082 and at the 2nd iteration, the PSO scheme did level of exploration and gradually switched between exploration and exploitation which converges at the 8th iteration. Likewise, for BA scheme, the convergence occurs at the 13th iteration with the best fitness value is 0.00419. The highest fitness of 0.00537 is steeply decreased by switching between exploration and exploitation. So, at their individual best fitness function, the schemes make the search for the global optimal solution.

4.4. Evaluation Results of the Classifiers. The proposed model is assessed based on the RF, K-NN, and MLP machine learning classifiers. The results of the performed experiment that is based on the aforementioned classifiers are presented in Table 6. These outcomes are based on the two-feature selection scheme (PSO and BA) adopted for the study and RF, K-NN, and MLP. It is observed that the highest recall rate of 0.996 was obtained from RF based on the dataset created from BA scheme which was closely followed by RF on the original dataset with a value of 0.98. Likewise, for accuracy metric as observed from Table 5, it is observed that RF on dataset created from BA scheme scored the highest value of 99.99%. Similarly, for F1_Score and Precision metrics, RF on dataset created from BA scheme still scored the same highest values of 0.996 and 0.996, respectively. So, based on classification report, RF on dataset created from BA scheme gave a superior performance compared to other models. MLP classifier performed poorly for both schemes and the aforementioned metrics.

Table 7 presented the result of the dataset analysis metrics efficient for evaluation of imbalanced dataset. Since sampling without replacing to the tune of 20% was applied to the dataset to treat the imbalance, geometric mean, Precision-Recall-curve, and log loss are better suited metrics. Based on geometric mean, RF gave a superior performance based on BA scheme with a value of 0.996 while RF and K-NN scored the value of 1 on dataset created from BA scheme for PRC, Log Loss and ROC metric, respectively. MLP classifier performed poorly for both schemes and the aforementioned metrics. These metrics was used since they are best in measuring the imbalanced cases. The results revealed that BA with RF performed better across all the performance metrics when compare with the PSO feature selection algorithm. The best of all the classifiers is the RF with the BA classifier.

4.4.1. Confusion Matrix (CM). The table of CM is used to define the performance of any classification models. This is used here to visualize and summarize the results of the performance of the proposed classifiers. This CM table shows the detection rate of each of the classes. Based on the results presented in Tables 5 and 6, it could be deduced that RF and DTC produced results that were similar. So, further analysis to reveal which model and on which performed best. Figures 3(a)–3(c) are confusion

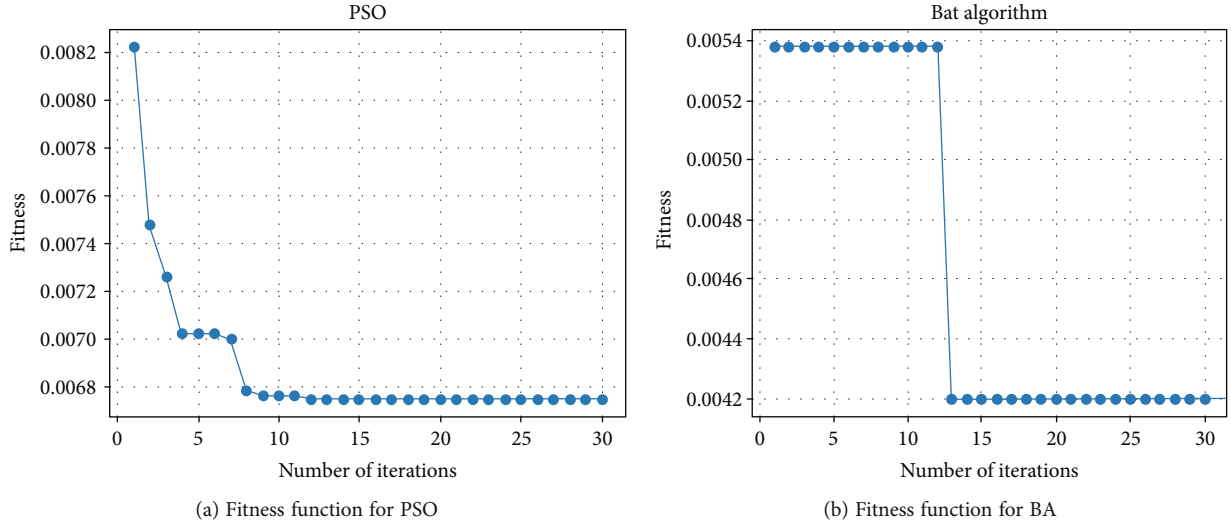


FIGURE 2: The convergence of the fitness function of PSO and BA.

TABLE 5: Selected Features.

Name of FSS	Parameters	Fitness value	No. of features	Index of selected features
Particle swarm optimization (PSO)	w=0.9 c1 = 0.5 c2 = 0.5 Fmax = 2 # maximum frequency Fmin = 0 # minimum frequency	0.00458	17	[mean, DstPkts, SrcBytes, TotBytes, SrcLoad, DstLoad, SrcRate, SrcLoss, DstLoss, pLoss, proto, TcpRtt, IdleTime, max, sTtl, TotAppByte, SynAck]
Bat algorithm (BA)	Alpha = 0.9 # constant Gamma = 0.9 # constant A = 2 # maximum loudness r = 1 # maximum pulse rate	0.00419	16	[sport, Dport, SrcPkts, Dst, Pkts, SrcBytes, DstLoss, loss, pLoss, DstJitter, Tcp, Rtt, IdleTime, sum, min, sDSb, TotAppByte, DstJitAct]

TABLE 6: Classification report.

Assessor	FSS	RF	K-NN	MLP
Recall	PSO	0.958	0.94	0.792
	BA	0.996	0.99	0.819
	Original	0.98	0.94	0.34
Precision	PSO	0.954	0.946	0.882
	BA	0.996	0.966	0.94
	Original	0.992	0.974	0.579
Accuracy	PSO	0.997	0.995	0.967
	BA	0.999	0.998	0.459
	Original	0.993	0.975	0.467
F1_Score	PSO	0.956	0.942	0.832
	BA	0.996	0.978	0.266
	Original	0.982	0.951	0.297

TABLE 7: Imbalance report.

Assessor	FSS	RF	K-NN	MLP
Geo mean	PSO	0.978	0.97	0.878
	BA	0.996	0.996	0.162
	Original	0.992	0.968	0.254
Log loss	PSO	0.036	0.037	1.024
	BA	0.0075	0.0075	18.66
	Original	0.0008	0.042	18.3309
PRC	PSO	0.0081	0.072	18.3021
	BA	1	1	0.319
	Original	1	0.998	0.326
ROC	PSO	0.98	0.98	0.88
	BA	1	1	0.55
	Original	1	0.99	0.57

matrixes based on the original and the dataset obtained from feature selected schemes (PSO, and BA) for and RF, K-NN, and MLP models. The class labels: Backdoor, CommInj, DoS, Recon, and normal represented by 0.0, 1.0, 2.0, 3.0, and 4.0.

From Figures 3(a) which presents BA scheme on RF, the classification performance for class Backdoor represented by 0.0 was 100% because all 54 instances were correctly classified. For class label CommInj represented by 1.0, 66 instances were correctly

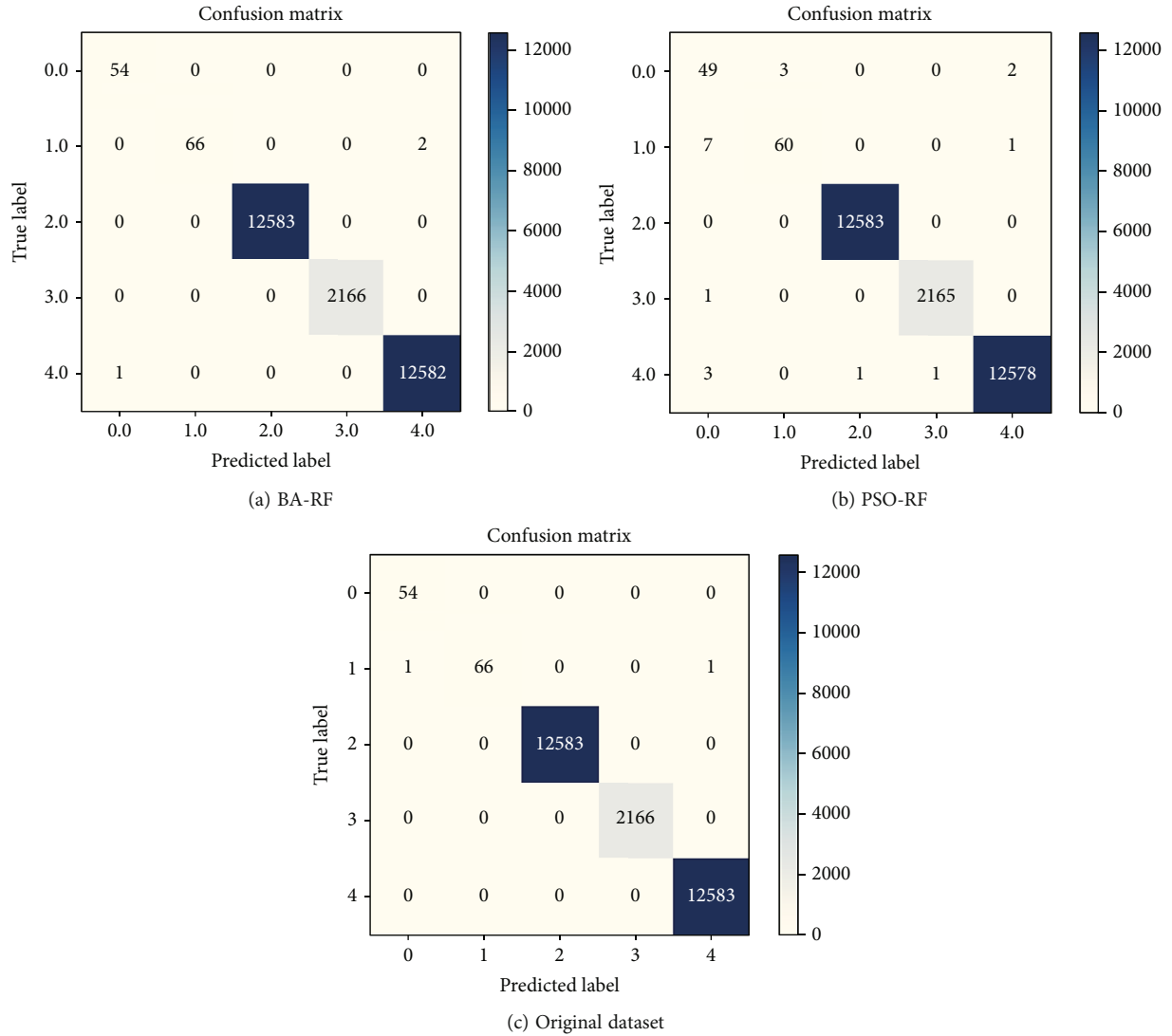


FIGURE 3: (a), (b), and (c) showing confusion matrix for BA, PSO, and original dataset on RF.

classified out of 68, 2 instances were misclassified as normal. For DoS:2.0 and Recon:3.0, their classification performance was 100%. For normal:4.0 class which is the majority class, 12582 instances were rightly classified out of 12583 instances while 1 instance was misclassified as Backdoor attack.

Similarly, for Figure 3(b) which presents PSO scheme on RF, out of 54 instances for class Backdoor represented by 0.0, only 49 instances were correctly classified, 3 instances were misclassified as CommInj and 2 instances as normal. For class label CommInj represented by 1.0, 60 instances were correctly classified out of 68, 7 instances were misclassified as Backdoor while 1 instance is misclassified as normal. For DoS:2.0, the classification was 100%. For Recon:3.0, 2165 out of 2166 were rightly classified while 1 was misclassified as Backdoor. For normal:4.0 class which is the majority class, 12578 instances were rightly classified out of 12583 instances while 3 instances were misclassified as Backdoor, 1 instance were wrongly classified as DoS and Recon, respectively.

From Figure 3(c) which presents original dataset on RF model, the classification performance for class Backdoor

represented by 0.0 was 100% because all 54 instances were correctly classified. For class label CommInj represented by 1.0, 66 instances were correctly classified out of 68, 1 instance were misclassified as Backdoor and 1 instance were misclassified as normal. For DoS:2.0, Recon:3.0, and normal:4.0, their classification performance was 100%.

4.4.2. Comparison of the Proposed Model with Existing Models. In recent years, researchers have attempted to resolve the issues of intrusion detection in the IoT network. As mentioned earlier, these researches are carried out using various techniques such as ML, semisupervised learning, adaptive, heuristic, decision tree, and DL. However, bio-inspired algorithms have been employed to extract the relevant features of the IoT networks in order to decrease processing cost, memory, and pave a smooth way to apply various classification techniques from the selected features. In this section, we present the related works on handling intrusion detection in the IIoT that employs bioinspired algorithms. At last, they are summarized in Table 8 comparatively.

TABLE 8: Comparison of the proposed model with state-of-the-art with existing models.

Method name	Accuracy (%)	Precision	Recall	F-score	Feature selection	Dataset
GWO-PSO-RF NIDS model [20]	99.88%	0.67	0.72	0.69	Hybrid GWO-PSO	KDDCup99
	99.24%	0.93	0.97	0.95		NSL-KDD
	99.87%	0.94	0.87	0.89		CICIDS2017
	99.66% (avg)					
DL-Rule Based Feature Selection Model [3]	99.0%	0.97	0.99	0.98	Rule based selection	NSL-KDD
	98.9%	0.99	0.99	0.97		UNSW-NB15
GA-RF model [19]	87.61%	0.98	0.81	0.89	Genetic algorithm	UNSW-NB15
AQU-CNN [33]	99.99%	0.99	0.99	0.99	Aquila optimizer (AQU).	CIC2017
	77.38%	0.84	0.77	0.77		NSL-KDD
	99.99%	0.99	0.99	0.99		BoT-IoT
	99.92%	0.94	0.92	0.93		KDD99
Proposed model	99.99%	0.996	0.996	0.996	BA	WUSTL-IIoT
	95.68%	0.997	0.958	0.956	PSO	

In [20], the authors used an hybrid GWO-PSO for feature selection before employing RF for classification of the dataset used to test the performance of the proposed model. The model performs reasonable better with average of 99.66%, but the proposed model still performs better with BA used for feature selection on the dataset used. The authors in [19] had used the genetic algorithm along with random forest model which was employed in the fitness function of the genetic algorithm proposed an IDS for IIOT. The reason behind the use of GA is the presence of the high number of features in the modern datasets as well as the number of network traces. As a result, the training process of the ML algorithms are negatively impacted and mislead as the ML performance reduces as the feature numbers increases. It is harder to perform the learning process as the number of attributes increases in the dataset. Therefore, the genetic algorithm is used for enhancing the feature selection and for each attribute vector, the author implemented Tree-based algorithms such as RF, DT, and ET algorithms which is conducted on the UNSW-NB15 general-purpose dataset.

The same dataset has been used along with the NSL-KDD dataset to implement a hybrid rule-based feature selection approach by authors in [3]. The proposed study integrates deep feedforward neural network model and rule-based feature selection with the applications of the IIOT in order to gather the relevant information that can be used to develop an intelligent NIDS (i.e., information is captured from TCP/IP packets). This study is a three-tier model for intrusion detection in IIoT systems in which a rule-based model is used for feature selection along with a genetic tool were used for feature selection and to generate attributes with the greatest values. At the end, the features that have been selected are loaded into the ANN for learning purposes.

The authors in [33] presented using feature selection for IDS in IoT-based system to remove irrelevant parameters before applying the DL model on the dataset. The proposed performed very well on the dataset used with an accuracy of 99.99%, and the model reduced the computational time of the proposed system. The proposed model performance was reasonably well with compared to the existing similar work

TABLE 9: Comparison of the proposed model with baseline model.

Technique	Accuracy (%)	Precision (%)	Feature selection
[49]	99.99	99.95	—
[2]	98.40	—	—
[48]	99.99	97.44	—
Proposed model	99.99	99.96	BA

in IoT-based systems. The accuracy of BA feature selection did well when compared with the existing models in this area.

To really show the importance of employing feature selection on the dataset before classification models, the proposed model used the baseline methods to compare the proposed model. Table 9 displays the comparison of the proposed model with the baseline model that used and created the dataset used.

In [49], RF performs better when compared with other ML-models used for the classification of the dataset with 99.99%, and Naïve Bayes has the least performance in term of accuracy with 97.48%, both RF and Naïve Bayes perform better in term of precision with 97.44%, but according to the authors, accuracy is not the best performance metric when it comes to the classification of huge amount of data, the sensitivity (precision) metric. Therefore, it can be said that the proposed model using feature selection with RF performs better than the baseline models. The computational time of the proposed models is very fast since the number of parameters used is reasonably reduced when compare with the baseline model. The same authors in [2] recorded an accuracy of 99.99%, and 99.95% of precision. In another study by the same authors in [48], the RF and Naïve Bayes performs better with precision of 97.44%, and the least of all the classifiers is the Logistic Regression with 47.44%.

Therefore, the proposed model performs reasonably better in terms of precision when compared with the baseline model. Hence, the model is optimal when in use in a real-world IIoT-based environment with huge amounts of unstructured and unlabeled datasets. The use of feature selection greatly reduces the computational time used in processing the dataset when

compared with the baseline, thus automatically reduces the data dimensionality and examines high-level functionality with effective accuracy and precision. Although our results seem similar to the other related work, as can be seen in Table 8, our proposed method has been tested on a more relevant dataset, WUSTL-IIoT, which is specifically collected for IIoT environment. So, our results would be more reliable than other related work.

5. Conclusions

The emergence of various cybersecurity techniques associated with IIoT-based network traffic has become critical to securing the IIoT environment from attackers and intruders from the outside world. Big data enabled with ML-based classifiers is a powerful tool for the analysis of huge data with the intention of securing the IIoT technology. The technologies have been proven helpful in the security of the IIoT-based system. However, the divergent implications and fundamental differences between IACS and traditional IT systems for counter-cyberattacks are distinct. Thus, special attention is required to provide security for the IIoT. Therefore, this study proposes a feature selection scheme with ML-based models for the classification of NIDS in IIoT-based traffic. The PSO and BA are used for feature selection to reduce the parameters used for the classification of the IIoT-based dataset used. For the classification, three different ML-based models are used to classify the dataset. The ML techniques were used to handle the new types of attacks like command injection, SQL injection, and backdoors after applying the feature section schemes to the dataset. The dataset used for the proposed model is the WUSTL-IIoT cybersecurity research. The experimental results show that the proposed model performs greatly better when compared with the baseline model, which created the testbed dataset with an accuracy of 99.99%, and 99.96% for precision. The feature extraction on the dataset reduces the computational time of the proposed model, which is very necessary when considering the use of an IIoT-based system. Future work will consider the use of a deep learning model for the classification of the dataset for ranking the attack traffic from the normal traffic. The security of the proposed system can be enhanced using the blockchain and various encryption techniques.

Data Availability

The data used in the study can be found in: <https://www.cse.wustl.edu/~jain/iiot2/index>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study is supported via funding from the Prince Sattam bin Abdulaziz University (project number PSAU/2023/R/1444).

References

- [1] M. Iaiani, A. Tugnoli, S. Bonvicini, and V. Cozzani, "Analysis of cybersecurity-related incidents in the process industry," *Reliability Engineering & System Safety*, vol. 209, article 107485, 2021.
- [2] M. Zolanvari, M. A. Teixeira, and R. Jain, "Effect of imbalanced datasets on security of industrial IoT using machine learning," in *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 112–117, Miami, FL, USA, November 2018.
- [3] J. B. Awotunde, C. Chakraborty, and A. E. Adeniyi, "Intrusion detection in industrial internet of things network-based on deep learning model with rule-based feature selection," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 7154587, 17 pages, 2021.
- [4] A. Sajid, H. Abbas, and K. Saleem, "Cloud-assisted IoT-based SCADA systems security: a review of the state of the art and future challenges," *IEEE Access*, vol. 4, pp. 1375–1384, 2016.
- [5] S. Gibbs, "Triton: hackers take out safety systems in watershed attack on energy plant," *The Guardian*, 2017, <https://www.theguardian.com/technology/2017/dec/15/triton-hackers-malware-attack-safety-systems-energyplant>.
- [6] K. Stouffer, J. Falco, and K. Scarfone, "Guide to industrial control systems (ICS) security," *NIST Special Publication*, vol. 800, no. 82, pp. 16–16, 2011.
- [7] S. Varghese, *Digital Twin-based Intrusion Detection for Industrial Control Systems*, 2021.
- [8] T. Gaber, A. El-Ghamry, and A. E. Hassanien, "Injection attack detection using machine learning for smart IoT applications," *Physical Communication*, vol. 52, article 101685, 2022.
- [9] R. O. Ogundokun, J. B. Awotunde, P. Sadiku, E. A. Adeniyi, M. Abiodun, and O. I. Dauda, "An enhanced intrusion detection system using particle swarm optimization feature extraction technique," *Procedia Computer Science*, vol. 193, pp. 504–512, 2021.
- [10] J. B. Awotunde and S. Misra, "Feature extraction and artificial intelligence-based intrusion detection model for a secure internet of things networks," in *Illumination of Artificial Intelligence in Cybersecurity and Forensics*, pp. 21–44, Springer, 2022.
- [11] M. Al-Omari, M. Rawashdeh, F. Qutaishat, M. Alshira'H, and N. Ababneh, "An intelligent tree-based intrusion detection model for cyber security," *Journal of Network and Systems Management*, vol. 29, no. 2, pp. 1–8, 2021.
- [12] J. B. Awotunde, R. G. Jimoh, S. O. Folorunso, E. A. Adeniyi, K. M. Abiodun, and O. O. Banjo, "Privacy and security concerns in IoT-based healthcare systems," in *The Fusion of Internet of Things, Artificial Intelligence, and Cloud Computing in Health Care*, pp. 105–134, Springer, Cham, 2021.
- [13] S. S. Mohammed, R. Hussain, O. Senko et al., "A new machine learning-based collaborative DDoS mitigation mechanism in software-defined network," in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–8, Limassol, Cyprus, October 2018.
- [14] J. B. Awotunde, K. M. Abiodun, E. A. Adeniyi, S. O. Folorunso, and R. G. Jimoh, "A deep learning-based intrusion detection technique for a secured IoMT system," in *International Conference on Informatics and Intelligent Applications*, pp. 50–62, Springer, 2021.
- [15] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

- [16] X. S. Yang and A. H. Gandomi, "Bat algorithm: a novel approach for global engineering optimization," *Engineering Computations*, vol. 29, no. 5, pp. 464–483, 2012.
- [17] A. S. Lalos, A. P. Kalogeras, C. Koulamas, C. Tselios, C. Alexakos, and D. Serpanos, "Secure and safe IIoT systems via machine and deep learning approaches," in *Security and Quality in Cyber-Physical Systems Engineering*, S. Biffl, M. Eckhart, A. Lüder, and E. Weippl, Eds., pp. 443–470, Springer, Cham, 2019.
- [18] B. Valeske, A. Osman, F. Römer, and R. Tschuncky, "Next generation NDE sensor systems as IIoT elements of industry 4.0," *Research in Nondestructive Evaluation*, vol. 31, no. 5-6, pp. 340–369, 2020.
- [19] S. M. Kasongo, "An advanced intrusion detection system for IIoT based on GA and tree based algorithms," *IEEE Access*, vol. 9, pp. 113199–113212, 2021.
- [20] P. K. Keserwani, M. C. Govil, E. S. Pilli, and P. Govil, "A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO-PSO-RF model," *Journal of Reliable Intelligent Environments*, vol. 7, no. 1, pp. 3–21, 2021.
- [21] F. E. Ayo, S. O. Folorunso, A. A. Abayomi-Alli, A. O. Adekunle, and J. B. Awotunde, "Network intrusion detection based on deep learning model optimized with rule-based hybrid feature selection," *Information Security Journal: A Global Perspective*, vol. 29, no. 6, pp. 267–283, 2020.
- [22] D. Moon, H. Im, I. Kim, and J. H. Park, "DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks," *The Journal of Supercomputing*, vol. 73, no. 7, pp. 2881–2895, 2017.
- [23] G. Prethija and J. Katiravan, "Machine learning and deep learning approaches for intrusion detection: a comparative study," in *Inventive Communication and Computational Technologies*, pp. 75–95, Springer, 2022.
- [24] O. A. Abisoye, O. S. Akanji, B. O. Abisoye, and J. Awotunde, "Slow hypertext transfer protocol mitigation model in software defined networks," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, pp. 1–5, Sakheer, Bahrain, October 2020.
- [25] M. Mayuranathan, M. Murugan, and V. Dhanakoti, "Retracted article: best features based intrusion detection system by RBM model for detecting DDoS in cloud environment," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3609–3619, 2021.
- [26] M. T. Nguyen and K. Kim, "Genetic convolutional neural network for intrusion detection systems," *Future Generation Computer Systems*, vol. 113, pp. 418–427, 2020.
- [27] A. Shenfield, D. Day, and A. Ayes, "Intelligent intrusion detection systems using artificial neural networks," *Ict Express*, vol. 4, no. 2, pp. 95–99, 2018.
- [28] W. Lian, G. Nie, B. Jia, D. Shi, Q. Fan, and Y. Liang, "An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning," *Mathematical Problems in Engineering*, vol. 2020, Article ID 2835023, 15 pages, 2020.
- [29] C. Long, Y. Zhang, J. Wei, W. Wan, J. Zhao, and G. Du, "A hybrid intrusion detection algorithm based on Gaussian mixture model and nearest neighbors," in *2019 IEEE 44th Conference on Local Computer Networks (LCN)*, pp. 117–120, Osnabrueck, Germany, October 2019.
- [30] V. Priya, I. S. Thaseen, T. R. Gadekallu, M. K. Aboudaif, and E. A. Nasr, "Robust attack detection approach for IIoT using ensemble classifier," <https://arxiv.org/abs/2102.01515>.
- [31] K. Raja, K. Karthikeyan, B. Abilash, K. Dev, and G. Raja, "Deep learning based attack detection in IIoT using two-level intrusion detection system," 2021.
- [32] P. K. Keserwani, M. C. Govil, and S. E. Pilli, "An optimal intrusion detection system using GWO-CSA-DSAE model," *Cyber-Physical Systems*, vol. 7, no. 4, pp. 197–220, 2021.
- [33] A. Fatani, A. Dahou, M. A. A. al-qaness, S. Lu, and M. A. Abd Elaziz, "Advanced feature extraction and selection approach using deep learning and Aquila optimizer for IoT intrusion detection system," *Sensors*, vol. 22, no. 1, p. 140, 2021.
- [34] O. Almomani, "A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms," *Symmetry*, vol. 12, no. 6, p. 1046, 2020.
- [35] A. P. Bhopale and A. Tiwari, "Swarm optimized cluster based framework for information retrieval," *Expert Systems with Applications*, vol. 154, article 113441, 2020.
- [36] M. M. Khan, K. Kasmarik, and M. Barlow, "Autonomous detection of collective behaviours in swarms," *Swarm and Evolutionary Computation*, vol. 57, article 100715, 2020.
- [37] A. Tharwat, T. Gaber, A. E. Hassanien, and B. E. Elnaghi, "Particle swarm optimization: a tutorial," in *Handbook of Research on Machine Learning Innovations and Trends*, A. Hassanien and T. Gaber, Eds., pp. 614–635, IGI Global, 2017.
- [38] W. Niu, Z. Feng, C. Cheng, and X. Wu, "A parallel multi-objective particle swarm optimization for cascade hydropower reservoir operation in Southwest China," *Applied Soft Computing*, vol. 70, pp. 562–575, 2018.
- [39] E. H. Houssein, A. G. Gad, K. Hussain, and P. N. Suganthan, "Major advances in particle swarm optimization: theory, analysis, and application," *Swarm and Evolutionary Computation*, vol. 63, article 100868, 2021.
- [40] R. C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512)*, pp. 84–88, La Jolla, CA, USA, July 2000.
- [41] F. Liu, X. Yan, and Y. Lu, "Feature selection for image steganalysis using binary bat algorithm," *IEEE Access*, vol. 8, pp. 4244–4249, 2019.
- [42] S. Y. Kim and A. Upneja, "Majority voting ensemble with a decision trees for business failure prediction during economic downturns," *Journal of Innovation & Knowledge*, vol. 6, no. 2, pp. 112–123, 2021.
- [43] N. V. Sailaja, M. Yelamarthi, Y. H. Chandana, P. Karadi, and S. Yedla, "Early detection of sepsis on clinical data using multi-layer perceptron," in *Machine Learning Technologies and Applications*, pp. 223–233, Springer, 2021.
- [44] S. I. Altalbany and A. A. Abualhussein, "Performance comparison of neural networks (MLP, RBFNN, ERNN, JRNN) models for stock prices forecasting to Bank of Palestine," *Muthanna Journal of Administrative and Economic Sciences*, vol. 11, no. 1, pp. 8–28, 2021.
- [45] N. Wagh and S. D. Agashe, "Data-driven frameworks for system identification of a steam generator," in *Data Management, Analytics and Innovation*, pp. 441–452, Springer, 2022.
- [46] M. A. Fauzi, A. T. Hanuranto, and C. Setianingsih, "Intrusion detection system using genetic algorithm and K-NN algorithm on DoS attack," in *2020 2nd International Conference on*

Cybernetics and Intelligent System (ICORIS), pp. 1–6, Manado, Indonesia, October 2020.

- [47] J. K. Seth and S. Chandra, “MIDS: Metaheuristic based intrusion detection system for cloud using k-NN and MGWO,” in *International Conference on Advances in Computing and Data Sciences*, pp. 411–420, Springer, 2018.
- [48] M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, “Machine learning-based network vulnerability analysis of industrial internet of things,” *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6822–6834, 2019.
- [49] M. Zolanvari, A. Ghubaish, and R. Jain, “ADDAI: anomaly detection using distributed AI,” in *2021 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, pp. 1–6, Xiamen, China, December 2021.

Research Article

Utilizing Artificial Intelligence and Lotus Effect in an Emerging Intelligent Drone for Persevering Solar Panel Efficiency

Faris A. Almalki ¹, Amani A. Albraikan ², Ben Othman Soufiene ³, and Obaid Ali ⁴

¹Department of Computer Engineering, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

²Department of Computer Science, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

³PRINCE Laboratory Research, ISITCom, Hammam Sousse, University of Sousse, Sousse 4023, Tunisia

⁴Ibb University, Department of Computer Science & Information Technology, Ibb, Yemen

Correspondence should be addressed to Obaid Ali; obaid.ali2016@gmail.com

Received 5 January 2022; Revised 23 March 2022; Accepted 12 April 2022; Published 30 April 2022

Academic Editor: Tarek Gaber

Copyright © 2022 Faris A. Almalki 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.

Solar energy has been receiving increasing attention due to the wide range of functionality it now supports and the promise of reduced environmental pollution. However, there are a few ongoing challenges, one of which relates to regular maintenance, especially for those high roof-mounted solar panels whose efficiency might have been degraded because of dust and dirt particle deposits. This paper is aimed at developing an autonomous drone which is equipped with a camera, a GPS sensor, a transceiver, and a pump conical tank to hold water with lotus effects, for the purpose of maintenance cleaning of solar panels which have been mounted on the roof top of buildings. The proposed system considers the combination of MobileNet and VGG-16 CNN techniques together for object detection and recognition that will enable assessment of solar panels, alongside the lotus effect physical technique that will enable maintenance of the solar panels. This would also put out of danger those human lives who have previously risked carrying out such high installation inspections and maintenance. A pump tank is designed and fabricated using a 3D printer to hold water with the lotus effect and is fitted on the drone. The proposed intelligent drone has been prototyped and deployed on-site at Taif University. The experimental results confirm the ability to detect dust and dirt particle deposits on solar panels while flying autonomously before proceeding to maintain them. The cleaning process with the use of the lotus effect has led to an increase in the power efficiency of the solar panels by approximately 31%, while the accuracy of the MobileNet framework for detecting particle deposits on the solar panels and thereafter making maintenance decisions rose to approximately 99%.

1. Introduction

The importance of relying on renewable energy means such as solar panels globally has been prompted. Thus, researchers, engineers, and IT specialists are striving to participate in employing advanced technology to achieve a better life for our planet. However, many researchers have identified the solar panels' efficiency reduction, due to deposited solid particles on their surfaces which decrease the absorbed solar radiation on which the device depends for generating electricity or heat. Further, efficiency reduc-

tion due to dirty surfaces can be as high as 60% if the installation is in dusty or polluted atmosphere and not cleaned frequently [1–3].

Thus, it is rational to have reliable tools that could save natural resources and adapt technology wisely, especially where it is widely accepted that organizations and research centers tend to reduce environmental pollution by reducing the use of depleted energy sources and replacing them with clean, renewable energy sources. Further, there are various pillars of the Fourth Industrial Revolution (4IR) that have gained attraction from both industry and academia sectors

to employ advanced technology such as artificial intelligence (AI), Unmanned Aerial Vehicles (UAVs), and Internet of Things (IoT) to propose innovative solutions towards green energy. Unquestionably, integrating UAV with IoT and AI can wider horizons of smart applications that could save time, resources, and even human lives [4, 5].

UAVs are sophisticated instruments that can be either preprogrammed or remotely controlled. UAVs are diverse in sizes, shapes, and functions. Further, the UAV system is considered a satellite, but without the distance penalty along with list of advantages such as proficiency, reliability, flexibility, portability, applicability, rapid deployment, Line of Sight (LoS) connectivity, low latency, and low cost of manufacturing, launching, and maintenance. Further, when coupling UAVs with the AI framework, more applications can be benefited such as remote sensing, high-resolution imaging, monitor disaster-relief missions, atmospheric studies, service delivery, surveillance, and empowering smart cities [6–8]. Using AI techniques and UAVs is extraordinarily priceless since it leads to pairing the real-time machine learning ability with the exploratory abilities of unmanned drones offering ground-level operators a human-like eye in the sky. This integration not only becomes help in computer vision and image processing from aerial imaging aspects but also becomes an enabler in wireless communications via drone to fulfil the increasing and diverse requirements across a large range of application scenarios [9–12].

The term “lotus effect” refers to water resistance and self-cleaning. Figure 1 shows the functional principle of lotus effect and how a droplet can slide off carrying dust and dirt, leaving the surface clean. The self-cleaning property of waterproof surfaces is called the lotus effect. The hydrophobic property of lotus flower has been simulated to obtain beneficial results in most areas, one of which is nanotechnology. The lotus effect can be applied on various surfaces, such as wood, metal, plastic, and glass, through chemical products for the sake of protection [13].

This paper is aimed at building and implementing an autonomous drone quipped with various payloads and paired with the MobileNet technique to enhance efficiency of solar panels installed in high places like roof of houses and buildings. This work’s contributions are noteworthy shift from the existing work, as presented in the paper’s aim and objectives by the end of the following section. The rest of this paper is organized as follows. Related studies are discussed in Section 2. The proposed model and implementation step are presented in Section 3. Section 4 presents the simulation results and validation discussions. Then, Section 5 concludes with a discussion focused on future perspectives.

2. Related Study

Solar panels are affected daily by various environmental factors such as dust and dirt, which reduce the efficiency of the solar panels. Therefore, trained workers are one of those approaches to maintain the cleaning process manually. Yet, higher effort, long time, high cost, and risk are issues associated with this approach. More advanced solutions are still

presented day after day, such as robotics or drones. This section is aimed at presenting a representative illustration of the related research works that have been considered in the literature, where a set of criteria have been used to source and review these related research works that meet this research scope. The criteria were as follows: (a) UAV platform type, (b) standalone topology, (c) aerial imaging, (d) AI framework if applicable, and (f) low-altitude missions. This section concludes by highlighting research gaps and our own research motivations.

Solar panels are exposed to several factors related to weather conditions throughout the year such as dirt, dust accumulation, atmospheric pollution, and bird droppings. Thus, many solar panel cleaning techniques have been discussed that could positively enhance the solar panel’s performance. Further, researchers believe that energy yield loss is caused by dust deposition on photovoltaic panels as claimed by [14], where it presents a comprehensive review of automated cleaning systems of solar panels. Several methods like pneumatic suction pump, air pump, water pump, drones, jet spray compressed air, self-cleaning brush, or using a central water pumping throughout the solar cell panels were used. Results of various solar panel cleaning techniques emphasize the importance of having reliable cleaning tools of solar panels. This study has opened various options to explore and experimentally carry out.

An intensive review is introduced in [15] where the review emphasizes the existence of different solutions for cleaning photovoltaic panels. Two of those solutions were utilizing a flying robot and an automatic arm (e.g., microfiber brush) to sweep for cleaning photovoltaic panels. Similar work is presented in [16] where self-inspection cleaning devices for solar panels based on machine vision were highlighted. The work has discussed an automatic arm with microfiber brush that can be fixed on the frame of the solar panels as a cleaning method. Additionally, an autonomous robot with wheels and fans is presented that allow the vehicle to fly and reach roof-mounted solar panels and then walk on the panels for an inspection mission.

Uddin et al. in [17] presented a design of a drone for cleaning the high-rise building windows, where the drone can spray water and use microfiber brush washes. Also, the designed drone was remotely controlled to allow workers have better security and maintenance of the surrounding area. Experimental results show the applicability of using a UAV system for cleaning purposes, which can be done in costly and efficient manners. Saurav and Jung in [18] have emphasized the importance of using an autonomous drone for cleaning solar panels in the field. The aim was not only for appropriate maintenance but also for the sake of efficiency, safety, and cost-effectiveness, where these aerial vehicles can lift its own weight and extra payload weights and fly autonomously not carrying out solar panel detection and cleaning.

Soman in [19] uses a drone with RGB and near-infrared data for rooftop detection, where this proposal would be helpful in solar panel deployment. An experimental investigation is presented in [20] on cleaning solar panels in desert climates using a UAV along with brush, microfiber wiper,

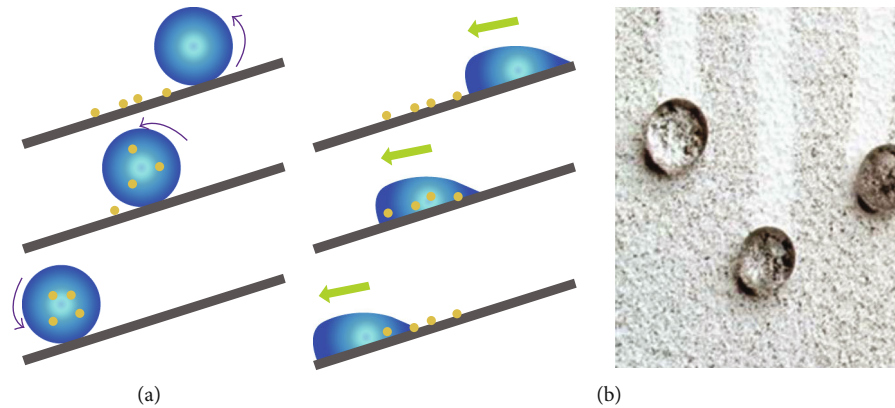


FIGURE 1: Working principle of the lotus effect: (a) a drop with a very high contact angle washes out contamination particles, (b) while a drop with a low contact angle does not clean the surface.

and vacuum cleaner. The cleaning mission has been carried out remotely using a remote control and camera. Results infer the effectiveness of aerial cleaning where the enhancement in the average power output of solar panels cleaned was around 7%. A fully autonomous UAV is utilized in [21] to fly over solar panels for complete inspections. The drone is associated with a thermal camera to collect thermal and regular video that can be postprocessed via a stitching algorithm. After three different cases, the proposed algorithm shows good level of detecting solar panels.

A drone-supported thermal imaging is used for estimating the efficiency of photovoltaic solar panels as presented in [22]. The proposed model design has focused on several diagnostic methods, all of which reduce the power output and accordingly the financial revenue. The drone along with its detection framework uses visual inspection, thermographic inspection, electrical measurement, and electroluminescence imaging. Results show the effectiveness of deploying these inspection techniques for predictive maintenance.

A drone that is payloaded with a thermographic camera and radiometric sensor to detect dust from solar panels is presented in [23]. Experimental results show the effectiveness of using such an aerial method to monitor the conditions of photovoltaic solar panels for the sake of maintainability, operating costs, availability, reliability, safety, and energy efficiency. Further improvement is introduced in [24] where a convolutional neural network (CNN) novel method is employed to enhance detection. Results show a good level of accuracy of the approach under different real scenarios. Recently, CNN has been explored in surface defect detection in strip steel, textiles, and building [25], where several techniques in intelligent inspection have been discussed. A deep learning (DL) method for detecting faults in solar panels is presented in [26], where the proposed model includes a drone equipped with a thermal camera and GPS sensor. The results of the proposed model show high performance in coupling an AI framework along with drone aerial imaging.

Weimer et al. in [27] employed a CNN to detect surface defects in textile and steel datasets. The impact of the CNN model's depth and breadth on test outcomes was explored.

Wang et al. in [28] developed a new CNN model structure. The model inputs all defect-free and defect samples and produces a 12-class classifier with six nondefective and six defective samples. However, since the dataset is limited, there is a risk of overfitting. The authors in [18] develop a robust and smart technique for recognizing and detecting solar panels experimentally. A TensorFlow-based real-time custom object detector along with its DL capabilities is used. The DL with its Raspberry-pi Interface and Jetson Nano show reasonable detection. However, although the proposal includes having a drone, no real drone deployment has been shown.

A flaw identification algorithm based on transfer learning was suggested in [29] to accomplish weight sharing and reduce overfitting. Further, Pierdicca et al. in [30] utilized a CNN to recognize solar cell remote sensing pictures and identify damaged cells in the solar cell module. Findings show that CNN has successfully identified solar cell defects. The drawback is that owing to the low-resolution remote sensing pictures of solar modules, the precision of CNN in this work is only approximately 70%. Deitsch et al. in [31] used a CNN to identify various electroluminescence faults in solar cells. Compared to traditional machine vision methods, the algorithm in this study achieves 88.36% accuracy on the dataset, a 6% improvement. At the same time, the algorithm's detection speed fulfils the demands of real-time manufacturing.

Kim et al. in [32] addressed an AI framework coupled with a drone for real-time remote sensing and energy efficiency. Furthermore, Pierdicca et al. in [33] presented solAIr; a mask R-CNN was used for detecting anomalous cells in photovoltaic pictures captured by drones with a thermal infrared sensor. Moreover, Kim et al. in [34] applied the Canny edge operator and image segmentation techniques to analyse IRT pictures obtained by drone. All the mentioned techniques show respected results in detection and inspection solar panel missions. Two self-developed techniques for detecting panels are compared based on classical techniques and the other on deep learning, both with a similar postprocessing phase [35]. The first technique relies on edge detection and classification, whereas the second relies on convolutional neural networks trained on regions to identify

a panel. The second technique uses deep learning to train pictures that have undergone three distinct preprocessing processes. Table 1 shows a comparison of the related research windup against the proposed work.

Based on the windup-related studies in Table 1, the research gaps have been identified, so we draw our own research motivations to inform our own proposed model. Clearly, the comparative analysis between approaches presented in Table 1 has highlighted the unresolved issues which are mainly two issues: (1) full autonomous aerial vehicle and (2) considering an effective cleaning approach while doing the intelligent aerial inspection. To our best knowledge and based on the related study, no work has been done to couple a fully autonomous drone with an AI framework to achieve two goals: (1) flying the drone autonomously using VGG-16 CNN and (2) detecting objects/dust on solar panels for cleaning purpose using MobileNet, let alone considering the lotus effect for cleaning and long-term protection. Therefore, our proposed novel contribution is motivated to cover this gap. To illustrate, the process can be seen from two angles: software and hardware. The software consists of two brains: the first brain is related to training the drone to fly autonomously using VGG-16 CNN, while the second brain is related to detection and recognition of solar panels using MobileNet. So, these two brains are working on the go; hence, while the drone flies autonomously, it can also detect solar panels and recognize the need of cleaning or not. The hardware part including the camera, GPS, and pump conical tank is synchronized with the software part, so once the second brain recognizes that solar panels need cleaning, the conical tank pumps water towards solar panels and then applies the lotus effect, whereas the GPS sensor reports the solar panels' coordinates to the control station.

This paper is mainly aimed at building and implementing an autonomous drone equipped with various payloads and paired with VGG-16 CNN and MobileNet techniques that work harmoniously together not separately to enhance efficiency of roof-mounted solar panels, as well as decrease personnel effort and risk. To achieve our aim, these objectives need to be pursued:

- (i) Determining drones and camera settings (e.g., flight path, altitude, camera focal length, and frame rate)
- (ii) Developing an AI framework with the two software brains
- (iii) Testing the efficiency of solar panels before and after cleaning and applying the lotus effect

From these objectives, we can drive the paper's contribution that is predominantly seen as a comprehensive framework that consists of a computer software technique (MobileNet and VGG-16 CNN), as well as a physical technique (lotus effect). This combination would not only enhance efficiency of solar panels but also solve issues on endangerment of human lives in such high installation places in a more efficient and timely fashion. MobileNet is used for object detection and recognition capability to iden-

tify roof-mounted solar panels and evaluate their status, where this guarantees real-time processing and thus reduces energy consumption on IoT devices, while VGG-16 CNN is used for its simplicity and accuracy to train the drone to fly autonomously. The lotus effect is applied for long-term protection, where this work has designed a pump tank using a 3D printer.

3. Proposed Work

This paper is aimed at evaluating solar panels mounted on the roof using an autonomous and intelligent drone. This has been done by coupling a fully autonomous drone paired with VGG-16 CNN and MobileNet techniques that work harmoniously together not separately to enhance efficiency of roof-mounted solar panels, as well as decrease personnel effort and risk. Further, the proposed system is aimed at not only detecting issues on solar panels but also cleaning solar panels with water and the lotus effect for long-term protection. Figure 2 illustrates a block diagram of the proposed work architecture, which consists of the sky segment and ground segment. The former contains a drone platform with its payloads including a camera paired with the AI framework, a GPS sensor, a 3D conical tank designed and fabricated using a 3D printer to hold water with the lotus effect, and a transceiver module, which is responsible for wireless communication between the drone and ground segment before transmitting the gathered data to the cloud for further storage and analysis. However, the latter segment includes a ground control center (GCC), which acts as a focal point between two segments and host gateways to external networks.

This section is mainly focusing on the architecture of the proposed intelligent AI framework in two phases as shown in Figure 3. Phase one deals with how to train a network to fly a drone autonomously using the VGG-16 CNN technique. Phase 2 deals with solar power detection and recognition to identify roof-mounted solar panels and evaluate their status for cleaning purposes using the MobileNet technique.

Phase 1. It starts with two preflight settings. The drone setting includes the drone's flight path, speed, and altitude, while the camera setting includes focal length, shutter speed, frame rate, aperture, and configuration (e.g., vertical and oblique). Thus, these preflight settings are vital to be tuned before drone flight execution, which in turn helps the drone fly autonomously. To achieve the autonomous flying, a VGG-16 CNN network is trained to obtain this goal, which also leads to precision aerial imagery. Researchers in [36] show an intensive study on the most frequent tool in support of autonomous flying, where it claims that the VGG-16 CNN is one of the most frequent approaches due to its simplicity and accuracy to train the drone to fly autonomously. The considered VGG-16 network is programmed, trained, and calibrated with the camera using the VGG-16 network support package in MATLAB, which has so many built-in functions and toolboxes to interact with hardware and provide additional processing capabilities [37].

Clearly, the proposed AI system for the autonomous drone allows flying and recording the current front view

TABLE 1: Comparison of the related research windup against the proposal.

Ref.	Vehicle type	Autonomous	Altitude	Cleaning approach	AI framework	Camera type
[15]	Flying robot	×	16 m	Microfiber brush	×	RGB
[16]	Automatic arm	×	×	Microfiber brush	√	×
[17]	Quad copter	×	32 m	Water spray	×	RGB
[19]	Drone	×	20 m	×	√	RGB
[21]	UAV	√	20 m	×	√	Thermal
[22]	Drone	×	14 m	×	√	Thermal
[23]	UAV	×	27 m	×	√	Thermal
[24]	UAV	×	50 m	×	√	Thermal
[26]	Drone	×	32 m	×	√	Thermal
[30]	Drone	×	50 m	×	√	Thermal
[32]	Drone	×	18 m	×	√	RGB
[33]	Skyrobotic	√	50 m	×	√	Thermal
[34]	UAV	√	16 m	×	√	Thermal
[35]	UAV	√	34-56 m	×	√	Thermal
Proposed model	Drone	√	10-30 m	Water spray and lotus effect	√	√

√ refers to an element that existed. × refers to an element that did not exist.

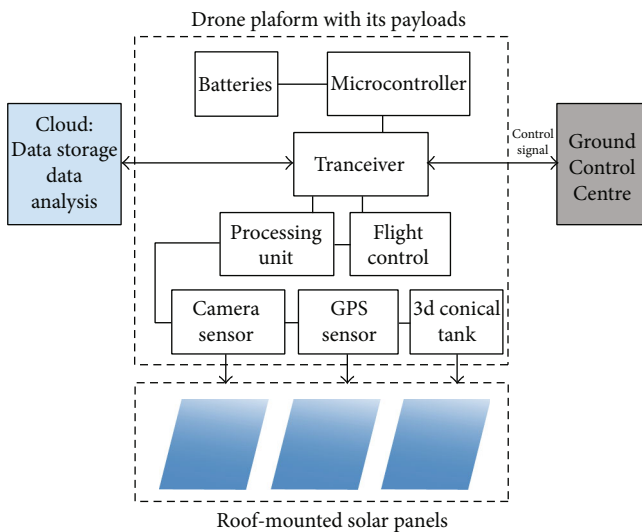


FIGURE 2: A block diagram of the proposed work architecture.

and yaw of the drone to generate more data for training and validation. Then, the recorded yaw is transformed to qualified yaw commands, before the VGG-16 CNN and a regressor get combined to train the neural network and control the drone path. Further, the regressor helps in generalizing real-world drone path settings in the provided environment. The yaw data take into consideration the drone's movement budget, which get updated regularly to have safe navigation and then landing. However, the yaw angle facilitates the drone's rotation while navigating at fixed velocity.

To train the VGG-16 neural network, the state was reset and a random UAV starting position and random movement budget were chosen. The training episode continues if the movement budget is greater than zero and the UAV has not landed. The episode finishes when either the drone

lands or the movement budget decreases to zero. Then, a new episode starts unless the maximum number of episodes is reached. The training function chosen here is backpropagation, which is considered due to its speed of convergence, moderate-sized feedforward NN with accurate update weights and biases near an error minimum. Therefore, the Mean Squared Error (MSE) between the predicted yaw and the actual yaw can be calculated as per

$$\text{MSE} = \frac{1}{n} \sum_i (y_i - p_i)^2, \quad (1)$$

where n refers to the number of training samples, i refers to a sample, y_i refers to the actual yaw, and p_i refers to the predicted yaw. The actual yaw can be computed as the angle between the current drone movement and the next drone movement to which it is heading. Thus, the actual yaw is considered successful if it is obstacle-free.

Phase 2. When a drone flies autonomously, a better precision aerial imagery can be obtained. This motivates us to apply a MobileNet for object detection and recognition capability to identify roof-mounted solar panels and evaluate their status. This mission can be done by coupling the camera on board the drone with lightweight computing the MobileNet. Further, this computer vision tool goes beyond processing of images which can reduce noise or recognize objects based on scale, color, or position. Instead, computer vision can understand the scene by identifying and classifying objects from images and videos [38, 39]. The MobileNet tool is seen as small with few parameters and addition operation. Hence, it is widely recognized in mobile-based computer vision applications.

The MobileNet considered here is MobileNetV2, which set between V1 with low computation and V3 with power-consuming computation that is suited to large data. The

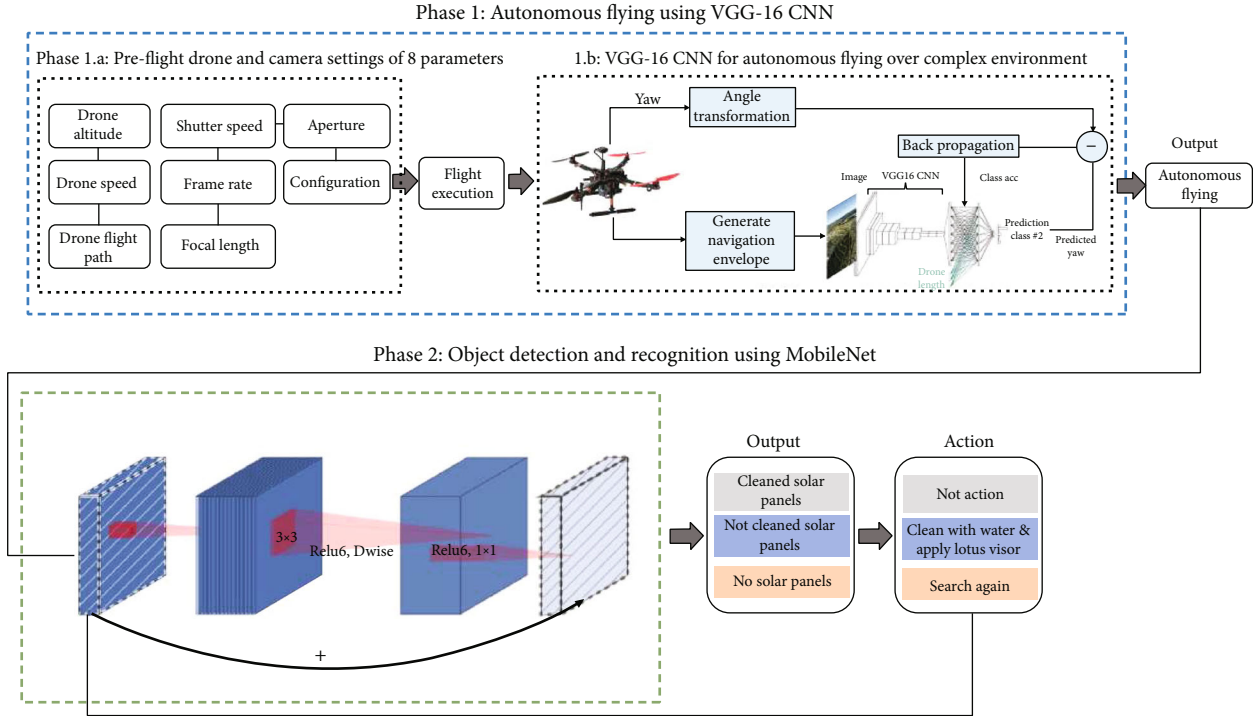


FIGURE 3: The proposed intelligent AI framework in two phases.

MobileNetV2 focuses on mobile and embedded vision applications to boost the efficiency of the NN by reducing the number of parameters by not compromising on performance. This MobileNetV2 structure, as shown in Figure 3, is based on depthwise separable convolution “pointwise,” which uses a single filter to every input channel, while the pointwise convolution is utilized to merge the outputs of depthwise convolution. This can help in reducing the number of parameters, which leads to improving network performance and low resource use.

Batch Normalization (BN) is applied to make the neural networks faster and more stable via normalizing the samples in a minibatch throughout the training process. This also would enable a higher learning rate, hence leading to enhancing the model generalization accuracy. The rectified linear unit (ReLU) is selected as the activation function for the convolutional layers attached to the flatten layer. The output feature map for regular convolution is represented as per

$$G_{k,l,n} = \sum_{i,j,m} K_{i,j,m,n} \cdot F_{k+i-1,j-1,m}, \quad (2)$$

where K refers to the kernel of the size of depthwise convolution, F refers to the filtered output feature map, and i, j, m, n refer to the number of feature channels, batch size, height, and width of each feature map, respectively. The cost of computation for a depthwise convolution is represented as per equation (3), while the depthwise separable convolution computational cost is represented as per equation (4).

$$\text{Computational cost} = D_k \cdot D_k \cdot M \cdot D_F \cdot D_F, \quad (3)$$

$$\text{Separable computational cost} = D_k \cdot D_k \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F, \quad (4)$$

where D_k denotes the kernel size, D_F denotes the feature map size, M denotes the number of input channel, N denotes the number of output channel, $D_k \cdot D_k \cdot M \cdot D_F \cdot D_F$ represents the depthwise convolution, and $M \cdot N \cdot D_F \cdot D_F$ represents the pointwise convolution. By using convolution as a two-step process of filtering and grouping, the cost of computation is given as per

$$\frac{D_k \cdot D_k \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_k \cdot D_k \cdot M \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_k^2}, \quad (5)$$

where both the width multiplier (α) and resolution multiplier (ρ) are introduced to minimize computational cost and control the image resolution, respectively, as per

$$\alpha = D_k \cdot D_k \cdot \alpha M \cdot D_F \cdot D_F + \alpha M \cdot \alpha N \cdot D_F \cdot D_F, \quad (6)$$

$$\rho = D_k \cdot D_k \cdot \alpha M \cdot \rho D_F \cdot \rho D_F + \alpha M \cdot \alpha N \cdot \rho D_F \cdot \rho D_F, \quad (7)$$

where α and ρ values range between 0 and 1. To note, when these two parameters are equal, the MobileNet baseline is achieved.

Since the MobileNetV2 has pretrained models in MATLAB called “MobileNet-v2 Support Package” that helped us to modify these models to be suitable for our proposed framework without requiring training a network from the scratch, the considered approach is adopted on the concatenated feature maps to compound multiple features together and decrease the number of channels and scale of

concatenated feature maps at the same time. The considered inception module has accomplished immense success on computer vision missions due to its efficient and delicate design. The refined feature output can be one of the following three: cleaned solar panels, not cleaned solar panels, and no solar panels. Subsequently, three actions can be taken based on the output as follows: no action, clean with water and apply lotus visor, and search again, respectively.

4. Simulation and Discussion

After defining the main components of the proposed intelligent AI framework in two phases, this section starts with outlining the hardware specifications of the experiment setup and the testbed structure. Then, practical measurements and results are presented. Figure 4 shows the final drone installation with all of its hardware elements. We can see that the UAV contains a hexacopter frame, kit 6-axis drone frame, Electronic Speed Controllers (ESC), 6 propellers, brushless DC motor, DC electric source, drone battery, and Pixhawk controller using a Mission Planner software tool. A camera for aerial imaging, a GPS sensor for coordinating location of the drone, a transceiver module for connecting the drone with the GCC, and a battery for powering components are the four main parts connected to the Raspberry Pi 3 B+ microcontroller, which is fitted on board the drone. The conical water and lotus effect tank system is fabricated using a 3D printer. For the sake of the experiment, the tank has been designed to meet the size and the weight capability of the drone, yet scalability is possible.

The experiment had taken place at the Taif University campus on the 19th and 20th of March 2021 at 1 p.m. at altitudes ranging from 10 to 30 meters. The university new campus's longitude and latitude are 40.4867323 and 21.3320348, respectively, which is in Saysad District in Taif City, Saudi Arabia. A Mission Planner software has been used when flying the drone at the university campus. Additionally, choosing weekend is to minimize the risk of any possible accident when flying a drone over a number of people.

4.1. Experimental Setup. This section contains the experimental setup along with visual results of the designed drone along with its AI framework, as shown in Figures 5–7. However, before presenting the visual results of the experiment, a brief discussion is highlighted about the process cycle of utilizing an autonomous drone equipped with various payloads and coupled with the CNN framework to clean roof-mounted solar panels. The process can be summarized as follows:

- (i) The designed drone was trained to fly autonomously by relying on the VGG-16 CNN network
- (ii) With the help of the GPS, the drone should be able to locate the solar panels and report that to the GCC
- (iii) The camera that fitted on board the drone should be able to take pictures of the solar panels and send them to the cloud for deeper analysis and storage purposes

(iv) By relying on the MobileNetV2 framework, the drone can determine whether solar panels are defective or not; thus:

- (1) If the solar panels are cleaned, no action is needed
- (2) If the solar panels are not cleaned, clean with water and apply lotus visor
- (3) If there are no solar panels, relaunch the drone flight and search again

The size of a solar panel is 0.5 m width by 1.2 m length, so the drone sprays appropriate amount of water to remove dust and dirt from the solar panels for 23 seconds. Then, the appropriate amount of the lotus shield is sprayed on the surface of the solar panel for 10 seconds to form a protective layer for around 9 months. The spraying process is done vertically from top to base. Figure 5 shows the two steps of cleaning the solar panels, where (a) the drone starts pumping water on the solar panel and (b) the drone applies the lotus effect on the solar panel for protection.

It is worth shedding the light on the dataset that has been used in this work for training, testing, and validation, where due to the nonavailability of sufficient-size and good-quality solar panel image dataset in public dearest platforms, a primary dataset has been considered, so the dataset balanced good-sized and resolution dataset of solar panel images have been taken from the Taif University campus. The dataset includes defective and nondefective images of solar panels to achieve an effective training, CNN classification, and validation. Figure 6 shows a representative sample of the primary dataset. Clearly, many different samples of solar panels have been taken into consideration for better training, testing, and validation. These considerations include angles of view, different types of solar panels, and defective and nondefective solar panels. Our entry point to the simulation is this primary dataset that includes 350 samples which were divided into three subsets: a training set, testing set, and validating set, where 245 of them were dedicated to the training phase, 52 were dedicated to the testing phase, and 52 were dedicated to the validating phase. The dataset distribution of the total dataset was 70%, 15%, and 15% for training, testing, and validating, respectively.

Figure 7 illustrates the output results of applying the proposed MobileNetV2 for object detection and recognition missions. This mission can be done by coupling the camera on board the drone with the MobileNetV2 framework. In our case, the framework can distinguish between the solar panel and other elements either on the top of buildings (e.g., roof windows and dish antennas) or even immediately on the solar panels (e.g., tree papers and dust). Figure 8 shows the defective and nondefective solar panel output at the stage training and testing the proposed MobileNetV2 network. Additionally, the network was trained and tested on the different classes of objects including leaves, dust, pens, and tools. Thus, it shows the ability to recognize the object if it is a solar panel or not. If yes, it can show if it is

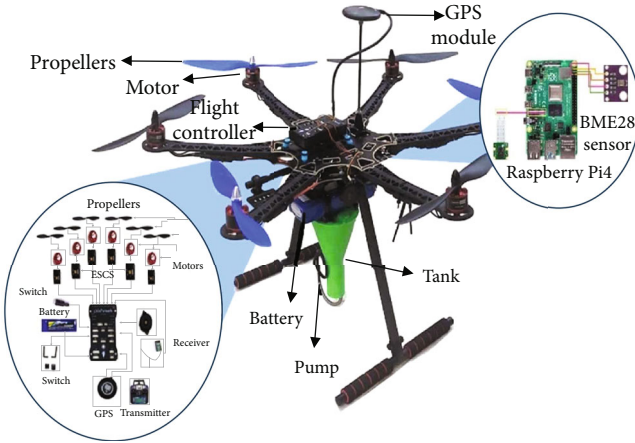


FIGURE 4: The final drone installation with its hardware elements.

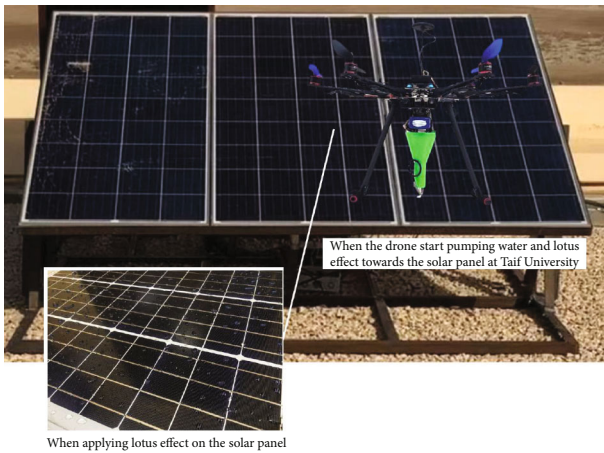


FIGURE 5: The two steps of cleaning the solar panels.

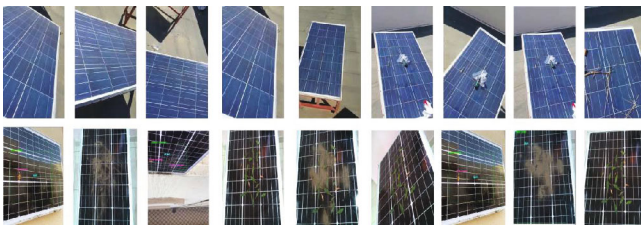


FIGURE 6: Representative samples of the primary dataset.

defective or nondefective. This is vital in the cleaning process without human interventions.

Various characteristics of the lotus effect have attracted us in this work to consider for cleaning solar panels and thus enhancing solar energy performance. Examples of these characteristics are preserving many materials for as long as possible, avoiding constant maintenance and cleaning work, decreasing the amount of water used in the cleaning process, and reducing cleaning time and effort. The panels have dye to absorb visible light 100%, where the nano glass minimizes the dust layer to occur overtime, and thus help the sun light

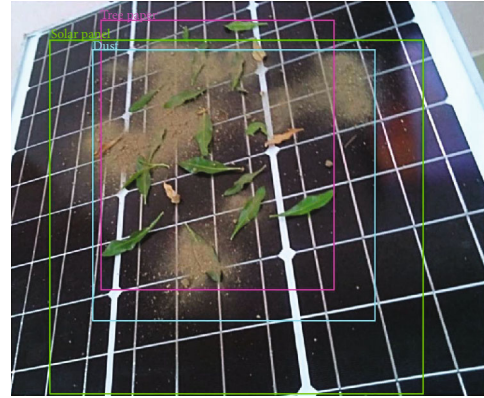


FIGURE 7: Results of solar panel detection and recognition.

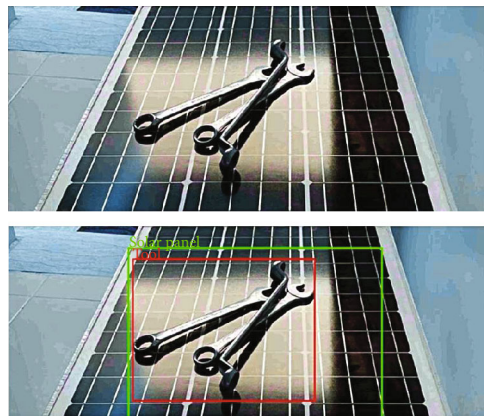


FIGURE 8: Defective and nondefective solar panel output.

to reach to the panels. Therefore, experimentally, applying lotus effect visor has increased power efficiency of solar panels up to 31%. Further, we found that 1 L of lotus effect liquid could cover around 70 m² of a solar panel sheet. Furthermore, the current version of the drone can fly up to 30 minutes and cover around 25 solar panels per trip, which represent around 30 m² of solar photovoltaic (PV) panels.

4.2. Validation Discussion. This section highlights two main performance indicators in the proposed AI framework by two indicators: first is the VGG-16 network loss and accuracy and second is the MobileNetV2 accuracy performance indicator, along with the confusion matrix. Figure 9 shows the VGG-16 network training and validation accuracy and loss. Clearly, the accuracy of 98.6% of the autonomous drone is a very good result that assures us for the drone intelligent navigation. Figure 10 displays the MobileNetV2 training accuracy against root mean square error (RMSE) and loss. This shows an accuracy of 99.1% with object detection and recognition. Figure 11 presents a confusion matrix, which is one of those well-recognized performance indicators in a classification. Thus, images of solar panels get classified with high accuracy into defective or nondetective classes. This is vital in the cleaning process without human interventions.

Table 2 shows the confusion matrix performance of the proposed MobileNetV2 framework. The table shows the

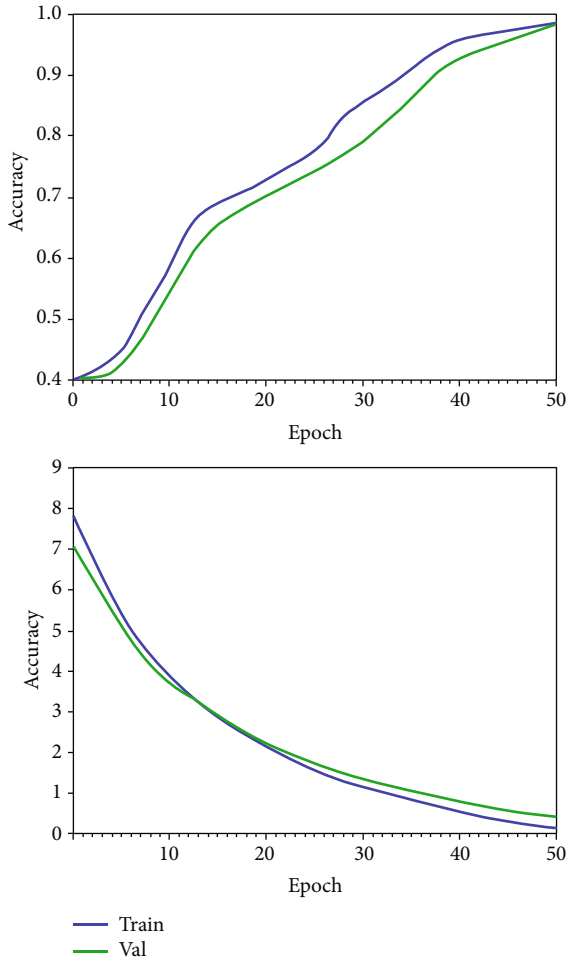


FIGURE 9: VGG-16 network loss and accuracy.

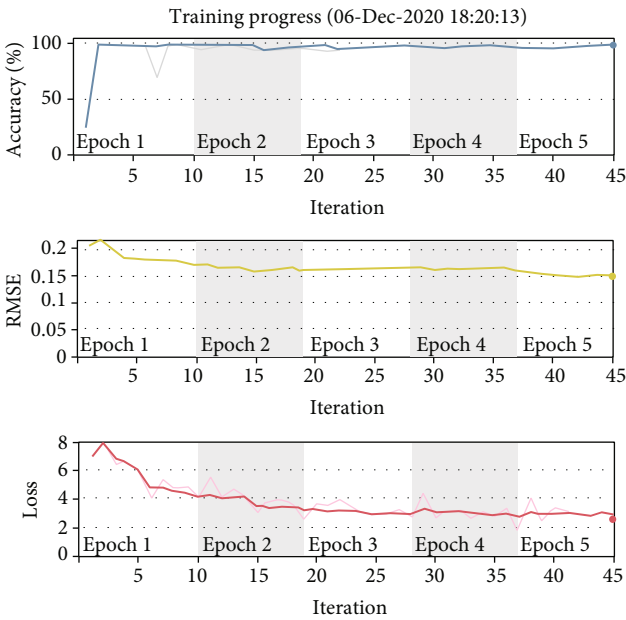


FIGURE 10: MobileNetV2 training accuracy.

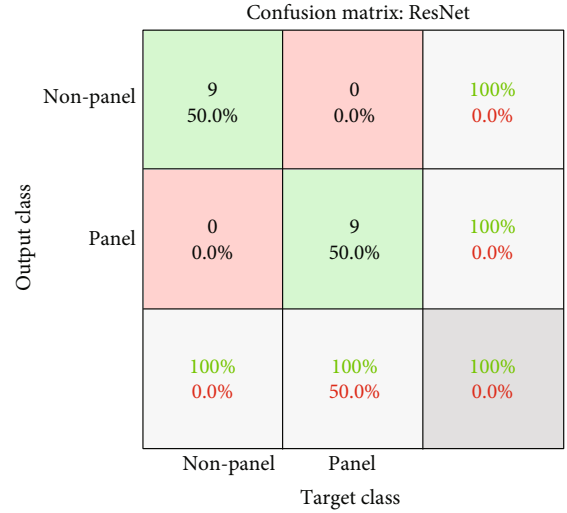


FIGURE 11: Confusion matrix of the MobileNetV2.

TABLE 2: Confusion matrix of the proposed framework.

Dataset	Sensitivity	Precision	Accuracy	F1-score
Training dataset	0.98	0.71	0.99	0.98
Validation dataset	0.99	0.74	0.99	0.99

four performance indicators of the confusion matrix after training, testing, and validation steps: indicator 1: sensitivity (out of all the positive classes, how much have been predicted correctly); indicator 2: precision (out of all the positive classes that have been predicted correctly, how many are actually positive); indicator 3: accuracy (out of all the classes, how much we predicted correctly); and indicator 4: the F1-score is used, which represents harmonic mean of precision and recall together. As can be seen in Table 1, the overall observation is that the proposed system is trained and classified with a high F1-score close to the value of 1. In turn, this confirms that the proposed MobileNetV2 framework is working well with a high level of accuracy.

To give our validation more depth, another performance indicator is used as verification is the system efficiency, as presented in Figure 12. The figure demonstrates a line graph of calculated input solar radiation (Prad) and the output as electrical power P_{el} , which refers to the measured current of the reach roof-mounted solar panels in relation to voltage characteristics at four stages dusty condition, partially dusty condition, clean condition, and clean condition with lotus effect. Clearly, there is a direct correlation between dust deposition accumulation on roof-mounted solar panel surface and output power of the energy system. Thus, as the dust deposition increases, a noticeable reduction of the output power and thus reduction in system efficiency were observed. Another observed point is that when applying a lotus effect on the roof-mounted solar panels, the system efficiency increases. That is because the lotus effect can slide off carrying dust and dirt, leaving the surface clean. This self-cleaning would serve as a protection layer and hence give

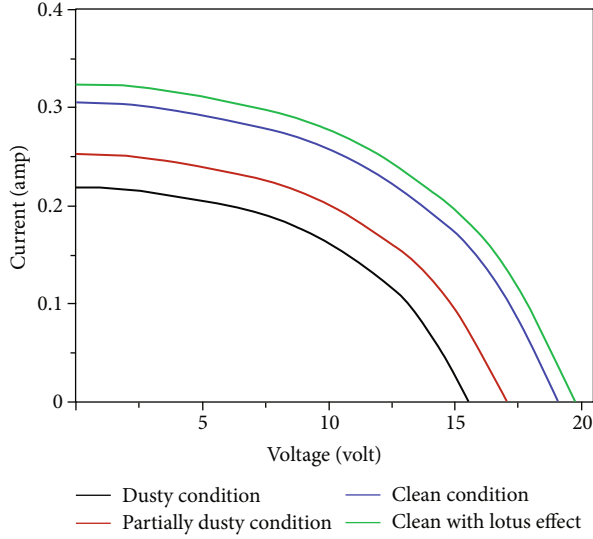


FIGURE 12: Measured current to voltage characteristics of the reached roof-mounted solar panels.

TABLE 3: Confusion matrix of the proposed framework against other techniques.

Technique	Accuracy (%)	Computation cost (ms)
SVM	98.47	216
Random forest	97.61	194
CNN ResNet 50	99.08	223
SqueezeNet	98.65	187
Proposed MobileNetV2	99.13	175

better system efficiency. Mathematically, there is around 31% improvement in the solar power efficiency between dusty condition and clean condition with the lotus effect.

4.3. *Comparison.* This section presents a comparison between the proposed MobileNetV2 against other neural network techniques that have been widely used for object detection and recognition missions, as shown in Table 3. The techniques that have been used are support vector machines (SVM), random forest, CNN ResNet 50, and SqueezeNet, which are considered examples of their groups. To note, the same dataset has been used for all techniques for the sake of unification and validation. Clearly, the table shows dissimilar neural network models concerning computation cost and accuracy rate. Where the proposed MobileNetV2 achieved better accuracy with less computation cost in comparison to other models. That is due to the total number of parameters and neural layers that affect the performance and computation cost.

5. Conclusion

This work seeks to help workers in the field of cleaning solar panels by providing unconventional methods of cleaning

solar panels, hence enhancing their efficiency. This paper is aimed at enhancing the performance of solar panels mounted on the top of buildings by designating an autonomous and intelligent drone, where the drone is equipped with payloads and paired with an AI framework to fulfil specific functions, not only to enhance performance of solar panels but also to solve issues on endangerment of human lives in such high places of installation. The proposed drone with its payloads and AI framework has successfully helped in detecting objects on the solar panels using the machine learning method before the drone utilizes water from the 3D designed conical tank to clean the solar panels and then sprays lotus effect on solar panels for long-term protection. The proposed framework shows great performance in training the drone to fly autonomously, as well as detect solar panels with high level of accuracy. Further, the system performance for both cleaning and dusty panels has been evaluated, where it shows around 31% improvement in the solar power efficiency. It is worth saying that the experiment of flying the drone was not easy and successful from the start, where it took multiple attempts as we have faced some issues. These issues would be an open research topic that can broaden the horizon of many applications to consider in the future. An example of these is the following:

- (i) Station keeping (drone stability) due to weather conditions (e.g., wind) is the key challenge, especially for lower altitudes, which might lose connectivity or require more power consumption to tracking and positioning
- (ii) Drone's load capabilities, which affect the quantity of water and lotus effect. So, it could be beneficial to use a larger drone and/or rely on a tethered drone

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to the Deanship of Scientific Research at Taif University, Kingdom of Saudi Arabia, for funding this project through Taif University Researchers Supporting Project number (TURSP-2020/265). The work of Amani Abdulrahman Albraikan was supported in part by Princess Nourah Bint Abdulrahman University Researchers Supporting Project Number (PNURSP2022R190). Furthermore, the authors are thankful to Aisha Abdullah, Azouf Alqethami, Aisha Hendawi, Wafaa Alotaibi, Maram Aljuaid, Kholod Almalki, Wejdan Alqethami, Reham Aljuaid, and Khadija Sherif for their cooperation.

References

- [1] A. N. Al-Shamani, K. Sopian, S. Mat, and A. M. Abed, "Performance enhancement of photovoltaic grid-connected system using PVT panels with nanofluid," *Solar Energy*, vol. 150, pp. 38–48, 2017.
- [2] F. A. Almalki and B. O. Soufiene, "Modifying Hata-Davidson propagation model for remote sensing in complex environments using a multifunctional drone," *Sensors*, vol. 22, no. 5, p. 1786, 2022.
- [3] A. Maxwell, G. A. Acosta, M. Ward, A. Hamm, and M. Hassanalian, "Cleaning mechanisms for solar panels of rovers and other extraterrestrial unmanned vehicles," AIAA Aviation Forum, 2019.
- [4] F. A. Almalki, "Comparative and QoS performance analysis of terrestrial-aerial platforms-satellites systems for temporary events," *International Journal of Computer Networks & Communications*, vol. 11, no. 6, pp. 111–133, 2019.
- [5] F. Almalki, S. H. Alsamhi, R. Sahal et al., "Green IoT for eco-friendly and sustainable smart cities: future directions and opportunities," *Mobile Networks and Applications*, 2021.
- [6] S. H. Alsamhi, F. Almalki, O. Ma, M. S. Ansari, and B. Lee, "Predictive estimation of optimal signal strength from drones over IoT frameworks in smart cities," in *IEEE Transactions on Mobile Computing*, p. 1, 2021.
- [7] F. A. Almalki, B. O. Soufiene, S. H. Alsamhi, and H. Sakli, "A low-cost platform for environmental smart farming monitoring system based on IoT and UAVs," *Sustainability*, vol. 13, no. 11, p. 5908, 2021.
- [8] S. Alsamhi, O. Ma, M. Ansari, and F. Almalki, "Survey on collaborative smart drones and Internet of things for improving smartness of smart cities," *IEEE Access*, vol. 7, pp. 128125–128152, 2019.
- [9] F. A. Almalki, M. Aljohani, M. Algethami, and B. O. Soufiene, "Incorporating drone and AI to empower smart journalism via optimizing a propagation model," *Sustainability*, vol. 14, no. 7, p. 3758, 2022.
- [10] F. A. Almalki, "Developing an adaptive channel modelling using a genetic algorithm technique to enhance aerial vehicle-to-everything wireless communications," *International journal of Computer Networks & Communications*, vol. 13, no. 2, pp. 37–56, 2021.
- [11] S. H. Alsamhi, F. A. Almalki, H. AL-Dois et al., "Multi-drone edge intelligence and SAR smart wearable devices for emergency communication," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6710074, 12 pages, 2021.
- [12] F. A. Almalki and S. B. Othman, "Predicting joint effects on CubeSats to enhance Internet of things in GCC region using artificial neural network," *Mobile Information Systems*, vol. 2021, Article ID 1827155, 16 pages, 2021.
- [13] M. Zhang, S. Feng, L. Wang, and Y. Zheng, "Lotus effect in wetting and self-cleaning," *Biotribology*, vol. 5, pp. 31–43, 2016.
- [14] J. Farrokhi Derakhshandeh, R. Alluqman, S. Mohammad et al., "A comprehensive review of automatic cleaning systems of solar panels," *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101518, 2021.
- [15] M. T. Grando, E. R. Maletz, D. Martins, H. Simas, and R. Simoni, "Robots for cleaning photovoltaic panels: state of the art and future prospects," *Revista Tecnología y Ciencia*, vol. 21, no. 35, pp. 137–150, 2019.
- [16] Z. He, Y. Zhang, and H. Li, "Self-inspection cleaning device for photovoltaic power plant based on machine vision," *IOP Conference Series: Earth and Environmental Science*, vol. 242, article 032020, 2019.
- [17] S. M. M. Uddin, M. R. Hossain, M. S. Rabbi, M. A. Hasan, and M. S. Rahman Zishan, "Unmanned aerial vehicle for cleaning the high rise buildings," in *International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 657–661, Dhaka, Bangladesh, 2019.
- [18] K. Saurav and S. Jung, "Robust detection of the coordinate of a solar panel using deep learning algorithm," in *19th International Conference on Control, Automation and Systems (ICCAS)*, pp. 1116–1119, Jeju, Korea (South), 2019.
- [19] K. Soman, "Rooftop detection using aerial drone imagery," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 281–284, Swisotel, Kolkata, India, January 2019.
- [20] M. Al-Housani, Y. Bicer, and M. Koç, "Experimental investigations on PV cleaning of large-scale solar power plants in desert climates: comparison of cleaning techniques for drone retrofitting," *Energy Conversion and Management*, vol. 185, pp. 800–815, 2019.
- [21] H. Ismail, A. Rahmani, N. Aljasmii, and J. Quadir, "Stitching approach for PV panel detection," in *Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1–4, Dubai, United Arab Emirates, 2020.
- [22] P. Pinceti, P. Profumo, E. Travaini, and M. Vanti, "Using drone-supported thermal imaging for calculating the efficiency of a PV plant," in *IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe*, pp. 1–6, Genova, Italy, 2019.
- [23] F. P. García Márquez and I. Segovia Ramírez, "Condition monitoring system for solar power plants with radiometric and thermographic sensors embedded in unmanned aerial vehicles," *Measurement*, vol. 139, pp. 152–162, 2019.
- [24] Á. Huerta Herraiz, A. Pliego Marugán, and F. P. García Márquez, "Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure," *Renewable Energy*, vol. 153, pp. 334–348, 2020.
- [25] F.-C. Chen and M. R. Jahanshahi, "NB-CNN: deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, 2017.
- [26] S. H. Han, T. Rahim, and S. Y. Shin, "Detection of faults in solar panels using deep learning," in *International Conference on Electronics, Information, and Communication (ICEIC)*, pp. 1–4, Jeju, Korea (South), 2021.
- [27] D. Weimer, B. Scholz-Reiter, and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection," *CIRP Annals*, vol. 65, no. 1, pp. 417–420, 2016.
- [28] T. Wang, Y. Chen, M. Qiao, and H. Snoussi, "A fast and robust convolutional neural network-based defect detection model in product quality control," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9-12, pp. 3465–3471, 2018.
- [29] S. Kim, W. Kim, Y.-K. Noh, and F. C. Park, "Transfer learning for automated optical inspection," in *International Joint Conference on Neural Networks (IJCNN)*, pp. 2517–2524, Anchorage, AK, USA, 2017.

- [30] R. Pierdicca, E. S. Malinverni, F. Piccinini, M. Paolanti, A. Felicetti, and P. Zingaretti, "Deep convolutional neural network for automatic detection of damaged photovoltaic cells," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. XLII-2, no. 2, pp. 893–900, 2018.
- [31] S. Deitsch, V. Christlein, S. Berger et al., "Automatic classification of defective photovoltaic module cells in electroluminescence images," *Solar Energy*, vol. 185, pp. 455–468, 2019.
- [32] B. Kim, J. Jung, H. Min, and J. Heo, "Energy efficient and real-time remote sensing in AI-powered drone," *Mobile Information Systems*, vol. 2021, 8 pages, 2021.
- [33] R. Pierdicca, M. Paolanti, A. Felicetti, F. Piccinini, and P. Zingaretti, "Automatic faults detection of photovoltaic farms: solAIr, a deep learning-based system for thermal images," *Energies*, vol. 13, no. 24, p. 6496, 2020.
- [34] D. Kim, J. Youn, and C. Kim, "Automatic fault recognition of photovoltaic modules based on statistical analysis of UAV thermography," *Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W6, pp. 179–182, 2017.
- [35] J. J. Vega Díaz, M. Vlaminck, D. Lefkaditis, S. A. Orjuela Vargas, and H. Luong, "Solar panel detection within complex backgrounds using thermal images acquired by UAVs," *Sensors*, vol. 20, no. 21, p. 6219, 2020.
- [36] T. Lee, S. Mckeever, and J. Courtney, "Flying free: a research overview of deep learning in drone navigation autonomy," *Drones*, vol. 5, no. 2, p. 52, 2021.
- [37] P. Naveen and B. Diwan, "Pre-trained VGG-16 with CNN architecture to classify X-rays images into normal or pneumonia," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 102–105, Pune, India, 2021.
- [38] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, "A novel image classification approach via dense-MobileNet models," *Mobile Information Systems*, vol. 2020, 8 pages, 2020.
- [39] N. Venkatesh, S. Yallappa, U. Hegde, and S. R. Stalin, "Fine-tuned MobileNet classifier for classification of strawberry and cherry fruit types," *Journal of Computer Science*, vol. 17, no. 1, pp. 44–54, 2021.

Research Article

Performance Optimization of 3-DOF Application Scene Based on 360-Degree Panoramic Technology with Depth Information

Qinghua Wang ¹, Linjuan Ma ², Xiaotong Huang ¹, Lan Lyu ¹ and Fuquan Zhang ^{3,4}

¹College of Mathematics and Data Science (Software College), Minjiang University, Fuzhou, China 350108

²School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China 100081

³College of Computer and Control Engineering, Minjiang University, Fuzhou, China 350108

⁴Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Chengdu, China 610021

Correspondence should be addressed to Fuquan Zhang; zfq@mju.edu.cn

Received 28 February 2022; Accepted 29 March 2022; Published 26 April 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 Qinghua 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.

High-quality 3D scenes often show poor rendering effect and insufficient operational performance in low-end devices. Therefore, how to make better use of 360-degree panoramic technology and replace the traditional pure model scene by visual deception is of great significance to the picture quality improvement and performance optimization of low-end equipment. In this paper, a three-degree-of-freedom (3-DOF) application scenario for indoor simulation is used as an example. Under the premise of 360-degree panoramic technology, depth information is obtained through laser ranging, converted into a two-dimensional form of depth image with the help of spherical coordinates, and combined with a transparent information image, the depth image is compressed using DEFLATE compression algorithm, and finally be stored. By optimizing the data through bilinear sampling and with the help of geometric knowledge, the occluded parts are removed or translucent according to the actual situation when the graphics are rendered, and the occlusion relationship is correctly handled in order to achieve a real integration of the virtual environment with the 3D objects. Through multiple data, it is shown that the method possesses significant performance improvement while ensuring geometric realism.

1. Introduction

With the development of digital media technology, the 3D rendering effect has been continuously improved and applied to various fields [1–10]. However, the generated 3D images require heavy computation, which result in significant increases in performance requirement by operating equipment. In order to make high-quality 3D images available on low-end devices, we often use 360-degree panoramic technology instead of the pure modeling scheme (PM) [11–16]. Nevertheless, because of the noninteractivity of virtual background, it is difficult to achieve real integration with the internal 3D object, with a result that it is generally being used only in distant scenes, or comes alone. Therefore, to further replace the 3D scene and increase the sense of reality, it is necessary to solve the occlusion relationship between this two.

As early as the augmented reality was proposed, people have carried out relevant studies on occlusion consistency [17–21], in which the technology that can be directly applied to 3-DOF application scenarios is the occlusion scheme eliminated by rendering sequence based on environmental modeling (OEM) [22]. Due to the invariant nature of the virtual scene, this technique reconstructs the environment with a low-model mesh, with depth testing and depth writing enabled, we adjust the rendering order so that the environment model is rendered first and the later rendered 3D objects are excluded from the back part of the environment and the opacity of the environmental low-model is set to 0. At present, the main problem of this solution is that the accuracy of the low model is not high and the masking of small objects is poor, while if the model accuracy is increased, the performance requirements of the equipment are significantly higher, resulting in the inability to play an

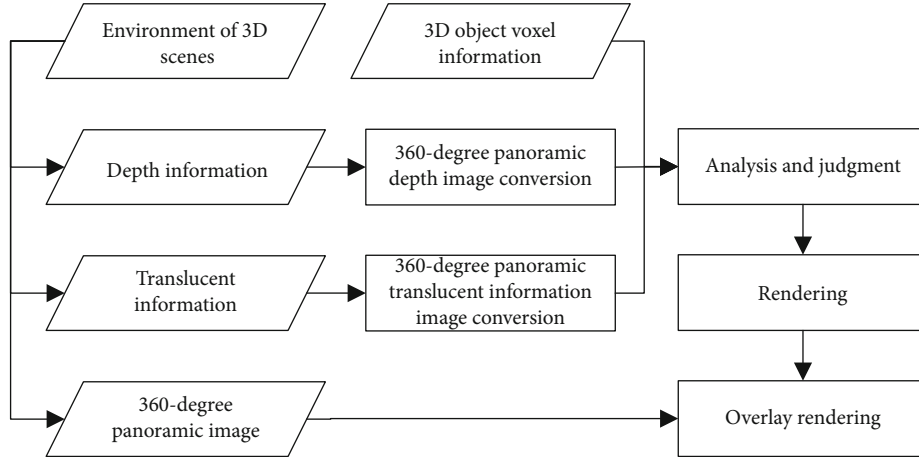


FIGURE 1: Flow chart of optimization method.

optimal role. Another common solution is a technique commonly used in 2D games which is implemented in 3D by layering the scene content into multiple 360-degree panoramas based on distance and proximity to achieve a sense of different levels of visualization. This approach is simpler, but the problem is also obvious: there is no transition effect when 3D objects span between layers.

In this paper, through the comparison and test of various solutions, we propose an occlusion scheme based on a 360-degree panoramic depth image (OPDI) for 3-DOF application scenarios of indoor simulation, based on the premise of optimizing performance and preserving spatial realism, using the depth information of 360-degree panoramic environment and eliminating the occluded parts of objects through graphical depth calculation, so that 3D objects can form a real visual sensation of occlusion with the environment in the only 360-degree panoramic image technology environment, which allows for the preservation of spatial realism while optimizing performance.

2. Method

This section specifically introduces the optimization scheme of 3-DOF application scenario using 360-degree panoramic technology with depth information in interior design simulation. The 360-degree panoramic image is acquired by the camera; furthermore, the 360-degree panoramic depth information is obtained by laser ranging and then saved in a 16-bit single-channel 360-degree panoramic grayscale image by spherical coordinate conversion and compression using DEFLATE compression algorithm and then based on the 360-degree panoramic depth image and scene translucent objects to draw a 360-degree panoramic transparency information image. Finally, when rendering the 3D object, the depth information and transparency information of the environment relative to the camera in each spatial voxel are retransformed to compare with the spatial voxel corresponding to the 3D object, so as to determine how to render and display. Figure 1 shows the specific process, where parallelograms represent the data, and rectangles represent the steps.

First of all, a 360-degree panoramic view of the required indoor environment needs to be collected according to the 360-degree panoramic camera, as shown in Figure 2.

In the same position of the camera, the laser ranging sensor rotates at high speed in the range of horizontal 360-degree and vertical 180 degrees, and the depth information of each angle is detected and collected in turn. The next step is to generate an associated 360-degree panoramic depth image. Unlike traditional 2D image methods, depth information will be built in a spherical coordinate system, which is more suitable to the properties of the depth data and then a lossless depth image is generated. Take the left-handed coordinate system as an example, the corresponding spherical coordinates are shown in Figure 3. Then, the information of each point on the spherical coordinate is mapped to the depth image, where the yaw angle $\theta(V_e)$ is the row coordinate of the depth image pixel, and the pitch angle $\omega(V_e)$ is the vertical coordinate of the depth image pixel. It is worth noting that for the convenience of the following calculation in this sphere coordinate, the yaw angle θ is calculated starting from +Z axis, and the pitch angle ω is calculated starting from -Y axis, while the distance value $\text{distance}(V_e)$ is converted to pixel value $P(P_{xi}, P_{yi})$. The mapping relationship is as follows:

$$\begin{aligned}
 P_{xi} &= \text{round}(\theta(V_e) \times \text{multiple}), \\
 P_{yi} &= \text{round}(\omega(V_e) \times \text{multiple}), \\
 P(P_{xi}, P_{yi}) &= \text{round}(\text{distance}(V_e)).
 \end{aligned} \tag{1}$$

In these formulas, the round function is used to round a number and multiple represents the quantization coefficient of the angle, which determines the resolution of the 360-degree panoramic depth image. Since $1/\text{multiple}$ is exactly the step size of the sensor, the obtained depth information density can be effectively controlled by adjusting the coefficient.

The generated depth images are saved in 16-bit single-channel PNG format grayscale images using DEFLATE compression derived from LZ77, and the data are



FIGURE 2: 360-degree panoramic image.

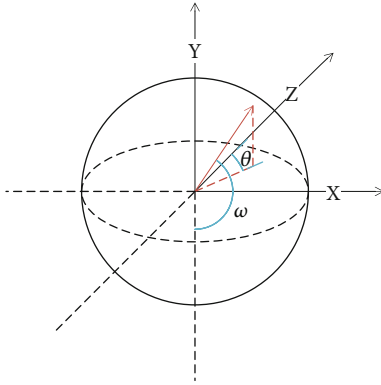


FIGURE 3: Ball coordinates.

compressed using DEFLATE compression algorithm after processing by the Lempel-Ziv (LZ77) [23] algorithm to ensure lossless compression while reducing spatial redundancy. The final result is shown in Figure 4.

After acquiring the 360-degree panoramic depth image, the next step is to perform the most important occlusion removal work in this paper, which will be done in the graphics rendering, where the model spatial coordinates of the 3D object vertices are obtained through the vertex shader and converted to world spatial coordinates and then further into camera spatial coordinates position_{camera}(x_{camera} , y_{camera} , z_{camera}).

$$\text{position}_{\text{camera}} = \text{position}_{\text{model}} \times \text{matrix}_{\text{mw}} \times \text{matrix}_{\text{wc}}, \quad (2)$$

where $\text{matrix}_{\text{mw}}$ is the transformation matrix from the model space to the world space and $\text{matrix}_{\text{wc}}$ is the transformation matrix from the world space to the camera space. After transformation, the results are transmitted to the patch shader, in which the vertex coordinate information is automatically linear interpolated, thus ensuring that the current camera space coordinates of each voxel of the object are obtained. Then, the distance value $\text{distance}(V_o)$ is calculated according to the camera spatial coordinates of each voxel. The formula is as follows:

$$\text{distance}(V_o) = \sqrt{x_{\text{camera}}^2 + y_{\text{camera}}^2 + z_{\text{camera}}^2}. \quad (3)$$

The next step is to obtain information about the direction of each corresponding voxel in the 360-degree panoramic depth image. According to the spatial coordinates of each voxel, the yaw angle $\theta(V_o)$ and the pitch angle $\omega(V_o)$ relative to the center position are calculated according to the following formula:

$$\alpha = \arctan\left(\frac{x_{\text{camera}}}{z_{\text{camera}}}\right) \times \frac{180}{\pi},$$

$$\theta(V_o) = \begin{cases} 180 - \text{abs}(\alpha), & (\text{IF } z_{\text{camera}} \leq 0 \text{ AND } x_{\text{camera}} > 0), \\ 180 + \text{abs}(\alpha), & (\text{ELSE IF } z_{\text{camera}} < 0 \text{ AND } x_{\text{camera}} \leq 0), \\ 360 - \text{abs}(\alpha), & (\text{ELSE IF } z_{\text{camera}} \geq 0 \text{ AND } x_{\text{camera}} < 0), \\ \text{abs}(\alpha), & (\text{ELSE}), \end{cases}$$

$$\omega(V_o) = \arctan\left(\frac{y_{\text{camera}}}{\sqrt{x_{\text{camera}}^2 + z_{\text{camera}}^2}}\right) \times \frac{180}{\pi} + 90. \quad (4)$$

It is worth noting that the value range of function \arctan is $[-\pi/2, \pi/2]$; therefore, we also need to make interval judgment of angle to ensure that $\theta(V_o)$ is in $[0, 360]$ interval. Since we defined $\omega(V_o)$ is in $[0, 180]$ interval as to store data conveniently before, here, we also need to add 90 to it in order to comply with the previous setting of ω starting with $-Y$ axis.

Based on yaw angle $\theta(V_o)$ and pitch angle $\omega(V_o)$ obtained from 3D model voxels, mapping to the $[0, 1]$ interval which served as the UV value of the pixel point of the depth image, the corresponding distance $P(UV_{xi}, UV_{yi})$ is read out from $_MainTex$ texture using the sampling function tex2d .

$$UV_{xi} = \frac{\theta}{360},$$

$$UV_{yi} = \frac{\omega}{180}, \quad (5)$$

$$P(UV_{xi}, UV_{yi}) = \text{tex2d}(_MainTex, UV(xi, yi))$$

Due to the interval of sampling steps, the final effect will produce a certain degree of sawtooth. Therefore, according to the step size and application requirements when

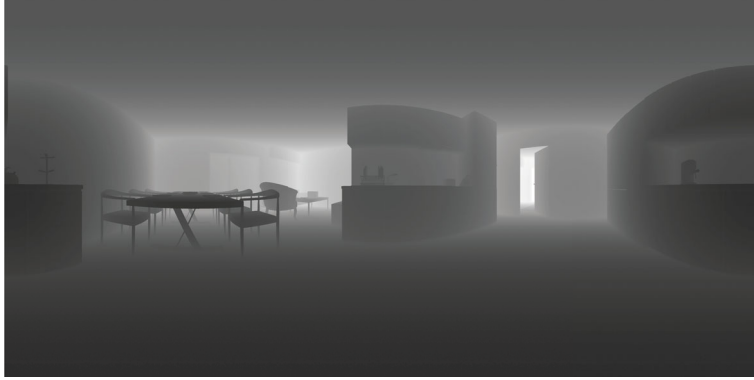


FIGURE 4: 360-degree panoramic depth image.



FIGURE 5: 360-degree panoramic transparent information image.

collecting the 360-degree panoramic depth map, the 360-degree panoramic depth image can be selectively bilinear sampling processing, so as to smooth the data. Finally, the distance value in the depth image is obtained according to the UV value of the corresponding voxel and compared with that voxel to determine whether the voxel needs to be rejected, where Max is the maximum value that can be stored in the depth image.

$$\text{IsDiscard}_{(1)} = \begin{cases} \text{true}, & (\text{IF } P(UV_{xi}, UV_{yi}) < \text{distance}(V_o) \text{ AND } \text{distance}(V_o) < \text{Max}), \\ \text{false}, & (\text{ELSE}). \end{cases} \quad (6)$$

$\text{IsDiscard}_{(1)}$ can now do most of the scene masking effects, but in reality, there are some cases where full masking is not desired, for example, glass and other transparent material environment objects which need to translucent processing rather than simply eliminate, so we also need a 360-degree panoramic image of transparent information. It is worth mentioning that through the drawing software, a copy is made on the depth image, and the part that needs to be transparent is filled with different grayscale colors according to the intensity of transparency, while white is completely opaque, and the resolution can be reduced appropriately according to the actual situation, so a 360-degree panoramic environmental transparency information image is obtained, as shown in Figure 5.

We can use it to calculate the opaque value Alpha of 3D objects to achieve the superposition effect with the 360-degree panoramic environment. The data acquisition method of 360-degree panoramic transparency information image is the same as acquiring depth information, where $A(UV_{xi}, UV_{yi})$ is the opacity value of the corresponding point on the 360-degree panoramic transparency information image, and when $A(UV_{xi}, UV_{yi}) = 1$, it means completely opaque, factor dynamically controls the adjustment of transparency intensity, function saturate limits the value range to $[0, 1]$ interval, and at the same time, we need to redefine the elimination condition $\text{IsDiscard}_{(2)}$ on the basis of $\text{IsDiscard}_{(1)}$.

$$\text{Alpha} = \begin{cases} 1, & (\text{IF } \text{IsDiscard}_{(1)} = \text{true} \text{ AND } A(UV_{xi}, UV_{yi}) \neq 1), \\ 1, & (\text{ELSE}), \end{cases}$$

$$\text{IsDiscard}_{(2)} = \begin{cases} \text{true}, & (\text{IF } \text{IsDiscard}_{(1)} = \text{true} \text{ AND } A(UV_{xi}, UV_{yi}) = 1), \\ \text{false}, & (\text{ELSE}). \end{cases} \quad (7)$$

3. Experiment

In order to verify the superiority of the scheme in this paper, we conducted three groups of control experiments on the virtual scene in Unity Engine 2020.3.11f1c1: they are OPDI,

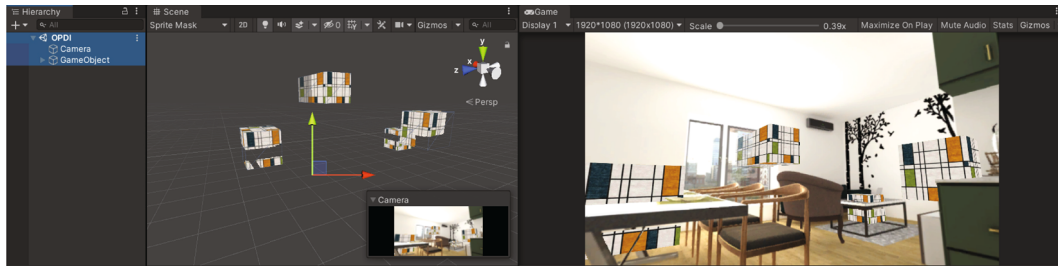


FIGURE 6: OPDI.

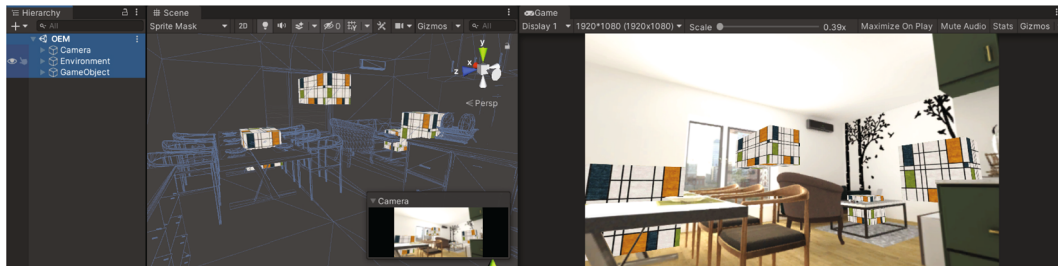


FIGURE 7: OEM.

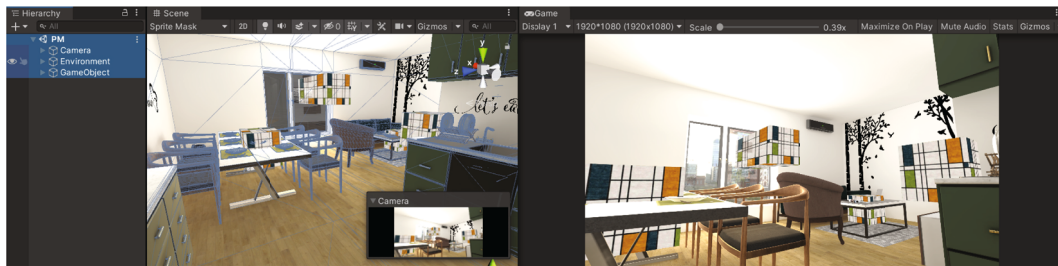


FIGURE 8: PM.



FIGURE 9: Translucent effect image of OPDI.

OEM, and PM. In the experiment, the Game panel was rendered with a resolution of 1920px * 1080px, and an indoor environment (model from Unity Asset Store) was selected as the scene; the scene of three scenarios was tested for the

corresponding masking effect with a rectangular cube as the 3D object.

In order to observe clearer, the display type of Scene Panel in Figures 6–8 is Sprite Mask. Through the Hierarchy

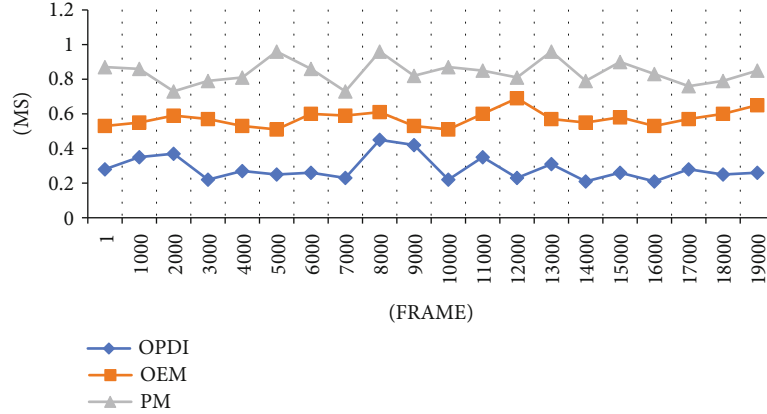


FIGURE 10: CPU usage (Rendering).

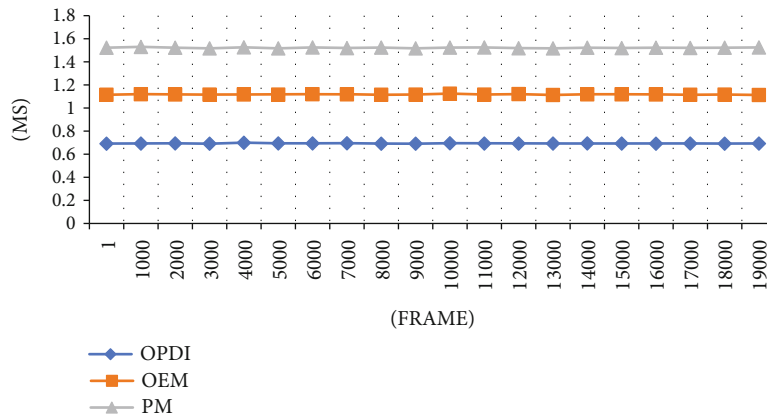


FIGURE 11: GPU usage (Camera.Render).

TABLE 1: Memory usage.

Experiment name	Memory usage
OPDI	1.92 GB
OEM	2.08GB
PM	4.61GB

TABLE 2: Experiment related environment.

OS	Windows 10 64-bit (DirectX 12)
CPU	Intel® Core™ i7-8750H CPU @ 2.20 GHz
GPU	Nvidia GeForce GTX 1060 (6 GB)
RAM	32 GB (DDR4 2666 MHz)

TABLE 3: Rendering.

Experiment name	Draw calls	Triangles	Vertices
OPDI	6	1.7 k	5.1 k
OEM	73	12.5 k	23.4 k
PM	79	56.0 k	50.7 k

Panel and the Scane Panel, we can clearly see the differences of these schemes in terms of model requirements, in which the OPDI does not use any environmental

model, while the model compression rate of OEM is about 75% of the model of PM. The Game Panel on the right side shows the occlusion effect under different schemes, and there is almost no difference in vision. The OPDI also reproduces the translucent effect almost realistically, as shown in Figure 9.

With the help of Unity Profiler Panel, the performance consumed by the three schemes is statistically analyzed. Figures 10 and 11 and Table 1 show the differences of CPU usage (Rendering), GPU usage (Camera.Render), and memory usage in different schemes during a period of running. The runtime configuration is shown in Table 2.

By comparison, it can be found that no matter in CPU usage (Rendering) or GPU usage (Camera. Render), OPDI takes the shortest processing time, and memory usage is also the least, followed by OEM, and the last is PM.

By analyzing the data in Table 3, triangles, vertices, and draw calls, all three of which are major factors affecting performance, where triangles and vertices represent the number of triangles and vertices to be drawn, and too much slice by slice and vertex by vertex computation will seriously affect the GPU usage. Draw calls, on the other hand, represent the number of times the CPU submits data to the GPU and then issues rendering commands to the GPU, so it will directly determine the CPU usage. Therefore, reducing the number of these three has a significant effect on improving

performance. The depth image-based occlusion scheme provided in this paper does exactly exploit this property while ensuring a sense of spatial realism as a way to greatly improve performance. Moreover, the scheme proposed in this paper has no significant change in the number of draw calls, triangles, vertices, and memory required by the scene in any environmental scenario; however, through the above data analysis, the relatively more complicated pure model-native scenario scheme, however, can optimize the performance substantially.

4. Conclusion

In this paper, taking the 3-DOF application scenario of indoor simulation as an example, a relatively novel performance optimization scheme is proposed under the condition of ensuring the spatial realism. Using 360-degree panoramic image instead of real 3D environment, mapping the distance of 360-degree panoramic environment relative to the camera to 2D image with the help of spherical coordinate conversion, and bilinear sampling data with the help of data alignment features in the image so as to optimize it, and comparing and eliminating the distance of each voxel of 3D object relative to the camera with the corresponding information saved in the 360-degree panoramic depth image to form the corresponding occlusion effect. Similarly, with the help of a 360-degree panoramic transparent information image, the occlusion of glass and other materials has a translucent effect. Through the method proposed in this paper, both the picture quality and the spatial realism are ensured, and the performance optimization is further achieved. According to the experimental results, the proposed scheme is obviously superior to the current mainstream 3D scene optimization schemes in terms of visual effect and performance. At the same time, the scheme is not only suitable for 360-degree panoramic image but also for 360-degree panoramic video. Therefore, it will have a broader application scenario, for example, in order to solve the problem of insufficient performance and poor display effect of 360-degree panoramic games with nonmobile perspective using high-quality rendering on traditional mobile platforms, the 360-degree panoramic video can also be used to achieve some visual effects of scene view movement, which is of certain significance for the expansion of virtual reality technology.

Data Availability

The data we used is available and the performance optimization scheme proposed in this paper can be used in 3-DOF application scenarios. And part of them is available to you from the corresponding author upon request (zfq@mju.edu.cn).

Conflicts of Interest

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

Acknowledgments

Our work is supported by the Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Project No.: 21DMAKL01; supported by the first batch of industry-university cooperation collaborative education project funded by the Ministry of Education of the People's Republic of China, 2021, Project No.: 202101071001; supported by Minjiang University school-level scientific research project funding, Projects No.: MYK17021, MYK18033, and MYK21011; and supported by Minjiang University Introduced Talents Scientific Research Start-up Fund, Projects No.: MJY21030.

References

- [1] N. F. Aristin, B. Budijanto, D. Taryana, and I. N. Ruja, "3D map of dry land use based aerial image as learning media in era of education 4.0," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 15, no. 7, p. 171, 2020.
- [2] O. López Chávez, L.-F. Rodríguez, and J. O. Gutierrez-Garcia, "A comparative case study of 2D, 3D and immersive-virtual-reality applications for healthcare education," *International Journal of Medical Informatics*, vol. 141, article 104226, 2020.
- [3] H. Siala, E. Kutsch, and S. Jagger, "Cultural influences moderating learners' adoption of serious 3D games for managerial learning," *Information Technology & People*, vol. 33, no. 2, pp. 424–455, 2020.
- [4] J. Kim, "Digital and postdigital 3D animation in the contemporary Chinese art scene: Miao Xiaochun and Lu Yang," *Journal of Chinese Cinemas*, vol. 11, no. 3, pp. 227–242, 2017.
- [5] F. Zhang, G. Ding, L. Ma, Y. Zhu, Z. Li, and L. Xu, "Research on stage creative scene model generation based on series key algorithms," in *International Conference on Smart Vehicular Technology, Transportation, Communication and Applications*, pp. 170–177, Mount Emei, China, 2018.
- [6] F. Zhang, G. Ding, Q. Lin, L. Xu, Z. Li, and L. Li, *Research of Simulation of Creative Stage Scene Based on the 3DGans Technology*14.
- [7] P. Derudas and Å. Berggren, "Expanding field-archaeology education: the integration of 3D technology into archaeological training," *Open Archaeology*, vol. 7, no. 1, pp. 556–573, 2021.
- [8] T. K. Coşkun and G. F. Deniz, "The contribution of 3D computer modeling education to twenty-first century skills: self-assessment of secondary school students," *International Journal of Technology and Design Education*, vol. 31, 2021.
- [9] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, p. 40, 2019.
- [10] A. Laksono and Aditya, "Utilizing a game engine for interactive 3D topographic data visualization," *ISPRS International Journal of Geo-Information*, vol. 8, no. 8, p. 361, 2019.
- [11] K. A. Ritter and T. L. Chambers, "Three-dimensional modeled environments versus 360 degree panoramas for mobile virtual reality training," *Virtual Reality*, vol. 25, 2021.
- [12] F. Chiariotti, "A survey on 360-degree video: coding, quality of experience and streaming," 2021, <https://arxiv.org/abs/2102.08192>.

- [13] S. Gilbert, W. Boonsuk, and J. W. Kelly, *Virtual Displays for 360-Degree Video*, Proceedings of SPIE - The International Society for Optical Engineering, Burlingame, California, USA, 2012.
- [14] R. Eiris, M. Gheisari, and B. Esmaili, "PARS: using augmented 360-degree panoramas of reality for construction safety training," *International Journal of Environmental Research and Public Health*, vol. 15, no. 11, p. 2452, 2018.
- [15] R. Eiris, J. Wen, and M. Gheisari, "iVisit: digital interactive construction site visits using 360-degree panoramas and virtual humans," in *Construction Research Congress 2020*, pp. 1106–1116, American Society of Civil Engineers, Tempe, Arizona, 2020.
- [16] C. Snelson and Y.-C. Hsu, "Educational 360-degree videos in virtual reality: a scoping review of the emerging research," *TechTrends*, vol. 64, no. 3, pp. 404–412, 2020.
- [17] A. Holynski and J. Kopf, "Fast depth densification for occlusion-aware augmented reality," *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1–11, 2018.
- [18] H. Kim and K. Sohn, *Hierarchical Depth Estimation for Image Synthesis in Mixed Reality*, A. J. Woods, M. T. Bolas, J. O. Merritt, and S. A. Benton, Eds., Proceedings of SPIE - The International Society for Optical Engineering, Santa Clara, CA, 2003.
- [19] D. E. Breen, R. T. Whitaker, E. Rose, and M. Tuceryan, "Interactive occlusion and automatic object placement for augmented reality," *Computer Graphics Forum*, vol. 15, no. 3, pp. 11–22, 1996.
- [20] D. R. Walton and A. Steed, "Accurate real-time occlusion for mixed reality," in *23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10, ACM, Gothenburg Sweden, 2017.
- [21] A. K. Hebborn, N. Hohner, and S. Muller, "Occlusion matting: realistic occlusion handling for augmented reality applications," in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 62–71, Nantes, 2017.
- [22] H. Kim and K. Sohn, "3D reconstruction from stereo images for interactions between real and virtual objects," *Signal Processing: Image Communication*, vol. 20, no. 1, pp. 61–75, 2005.
- [23] M. Sudan and D. Xiang, "A self-contained analysis of the Lempel-Ziv compression algorithm," 2019, <https://arxiv.org/abs/1910.00941>.

Research Article

Artificial Intelligence of Things-Based Optimal Finite-Time Terminal Attractor and Its Application to Maximum Power Point Tracking of Photovoltaic Arrays in Smart Cities

En-Chih Chang , Chun-An Cheng , and Rong-Ching Wu 

Department of Electrical Engineering, I-Shou University, No. 1, Sec. 1, Syuecheng Rd., Dashu District, Kaohsiung City 84001, Taiwan

Correspondence should be addressed to En-Chih Chang; enchihchang@isu.edu.tw

Received 6 January 2022; Revised 5 March 2022; Accepted 17 March 2022; Published 7 April 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 En-Chih Chang 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 combination of artificial intelligence of things (AIoT) and photovoltaic power generation can save energy and reduce carbon emissions and further promote the development of smart cities. In order to obtain the maximum power output from photovoltaic (PV) arrays, we can use optimal maximum power point tracking (MPPT) technique with AIoT sensing to improve system efficiency. The optimal MPPT technique is the finite-time terminal attractor (FTTA) based on the gradient particle swarm optimization (GPSO), which can be applied to track the maximum power of a PV array system. The FTTA not only provides fast finite-time convergence but also attenuates steady-state errors, making it ideal for nonlinear system applications. The GPSO is used to search the control parameters of the FTTA, which is able to find the global best solution. This avoids unmodeled dynamic behavior of the system excited by the quiver, which slows down the control convergence and prematurely traps the system into a local optimum. The MATLAB computer software is used to simulate the proposed PV maximum power point tracking system. The results show that more accurate and better tracking control of the PV array can be produced under partial shading conditions and then improve the steady-state and transient performance.

1. Introduction

With advances in technology, solar energy is becoming the most cost-effective long-term investment [1–3]. Like any equipment, solar systems must be managed effectively to optimize energy and power production. With solar smart cloud monitoring, users can get real-time information on PV power generation through the Internet and collect environmental information from the site to ensure the efficiency of the PV array system. Therefore, how to maximize the performance of solar cells has been one of the most important development projects in various countries around the world and is also the most important issue in solar energy-related technology. With the aim of attaining PV array systems with maximum power tracking, a DC-DC power converter including maximum power tracking control method modulates the solar cell output to deliver maximum power [4–7]. There are many research

works presenting all kinds of maximum power point tracking (MPPT) alternatives, such as incremental conductivity, perturbation observation, most valuable player methodology, and fuzzy logic [8–11]. There are couplings among sunlight brightness as well as atmospheric climate and the changes in PV maximum output power. The majority of the MPPT algorithms fails to perform a rigorous evaluation of both convergence and reliability and even cannot preserve stable pursuit of the maximum power point quickly. The sliding mode control (SMC) has simplified architecture, allows effortless design, and gives the system robustness against changes in internal parameters as well as external perturbations [12–18]. There are lots of useful SMC publications for controlling PV array systems [19–22]. However, photovoltaic array systems controlled by the SMC are subject to uncertainties. The state convergence time of such a system is not limited and quiver arises, thus compromising system performance in the steady-state

and transience. The quivering can be particularly problematic in practice, as it not only implies excessive energy consumption but also can provoke unmodeled high-frequency plant dynamics. There have been suggested solutions to ameliorate the quiver problem, such as estimator and adaptive methods. Even though they have improved the quiver problem and also enhanced the transience behavior of the system in the presence of unspecified disturbances, the mathematically complex and computationally time-consuming exist [23–25]. With explicit convergence and stability analysis, the finite-time terminal attractor (FTTA) not only provides the design approach for the robustness of the conventional sliding law, but more specifically, it provides a finite system state convergence time, i.e., a finite time for the system trajectory to reach the sliding mode zone in the presence of uncertainties. This methodology further strengthens the dynamical characteristics of the system and minimizes steady-state errors as well as quiver problems [26–29]. However, what is more important to note is that even though the FTFA allows the PV system to achieve the desired control effect, PV arrays may be partially shaded by buildings, wood, pollution, etc., degrading power output and leading to energy loss. There will be no regular fluctuation in the PV output power which resides in multiple local extremes. If the conventional MPPT methods described previously (e.g., disturbance observation, incremental conductivity, most valuable player methodology, and fuzzy logic) are used, they will be traced to the local extremes (local maximum power points) rather than the global extremes (global maximum power points). There have been several approaches tried to tackle multiple local extrema, such as the grey wolf optimizer and the brute force algorithm [30–32]. The grey wolf optimizer is fast, but it is constrained to local searching and incapable of searching globally. The brute force algorithm gives a superior ability to seek the best solution, but it has the weakness in needing a longer seeking time which stagnates easily on a certain solution. The gradient particle swarm optimization (GPSO) algorithm is based on the concept of gradient mechanism, showing simplified computation and enhanced population diversity that can demonstrate global domain search capability and has been widely used to solve many optimization problems [33–37]. It can improve the drawback of traditional PSO algorithm which tends to prematurely converge into local extrema [38–42]. The GPSO is therefore used to calculate the voltage reference value of the maximum power point of the PV array in the event of partial shading. In this paper, a GPSO is used to find the global extrema of the PV array under partial shading, and then, a FTFA is employed for its tracking control, providing good control power output to maintain the highest conversion efficiency of the PV array system. As a result, the FTFA based on the GPSO will improve the quiver problem, lessen the steady-state error, shorten the system state convergence time, solve the multipeak phenomenon (local maximum power point), and then render the PV MPPT system to have good steady state and dynamic response at the same time. The proposed controller with simplicity and high speed convergence as well as easy programming permits higher accuracy and stable following control. Computer simulations display that the proposed controller will provide the PV MPPT system to improve steady-state performance and

transient tracking speed in the presence of partial shading or under highly uncertainty conditions. The proposed system is also compared with a conservative sliding mode controlled PV MPPT system to show the superior performance and theoretical applicability of the proposed system.

2. Modeling of PV MPPT System

A solar energy system usually consists of a PV array, a DC-DC converter, a DC-AC inverter, and the attached load. With changes in brightness as well as temperature, there will be a reference value for the voltage corresponding to the maximum power point of the PV system. Therefore, an interleaved Boost DC-DC converter (Figure 1) is used to adjust the PV array voltage (for the maximum power point voltage). The dynamic equation of the interleaved Boost DC-DC converter is derived from the state-space averaging method for the PV system as follows:

$$\begin{cases} \dot{i}_{db} = -2L_{dbL}^{-1}[v_{dbo} \cdot (1 - u) - i_{db} \cdot R_{dbL} - v_{dbi}], \\ \dot{v}_{dbo} = C_{db}^{-1}[i_{db} \cdot (1 - u) - v_{dbo} \cdot R_{db}^{-1}], \end{cases} \quad (1)$$

where $i_{db} = i_{db1} + i_{db2}$, v_{dbi} is the PV array output voltage, i_{dbo} indicates the PV array output current, i_{dbo} denotes the output load current, v_{dbo} means DC output voltage, u symbols control input, $L_{dbL} = L_{dbL1} = L_{dbL2}$ and $R_{dbL} = R_{dbL1} = L_{dbL2}$, here R_{dbL1} and R_{dbL2} refer to the equivalent series resistance of inductors L_{db1} and L_{db2} , respectively. Let v_{dem} be the voltage reference of the maximum power point calculated by the GPSO algorithm, and in order to make v_{dbo} follow v_{dem} , the FTFA closed-loop control technology is necessary to achieve such a result. In other words, in a PV system, the error in the output voltage can be defined as the state variable $\tilde{x}_2 = v_{dbo} - v_{dem}$, which is regulated from the state variable $\tilde{x}_1 = i_{db}$ feed current. Our goal is to design the control law u properly. If it is well designed, (1) will be stable and the error \tilde{x}_2 quickly converges to the balance point. The PV output voltage will be the same as the required reference voltage. Even if the PV MPPT system is partially shaded or malfunctioning or under nonmatching uncertainty, the tracking control can still be fast, accurate, and robust.

Then, a single-phase PWM DC-AC inverter is used to convert the generated DC power into AC power for transmission to the grid. A typical single-phase PWM inverter is shown in Figure 2, where four semiconductor switches, LC filters, and loading (resistive loading or capacitive input rectifier loading) are considered as the plant to be handled. Owing to the diversity of the loading connected to the inverter, it is not possible to develop a general model that is representative of every type of load. In this case, a nominal load is defined to derive a linear model, where load changes and model uncertainties are regarded as disturbances. Furthermore, in view of the assumption that the switching frequency is much higher than the modulation frequency of the inverter, the dynamics of the single-phase PWM DC-AC inverter is mainly handled by its LC filter and the attached load, which can be modeled as a second-order linear system. Based on Kirchhoff's voltage and current laws, the state equations of the inverter can be described as follows:

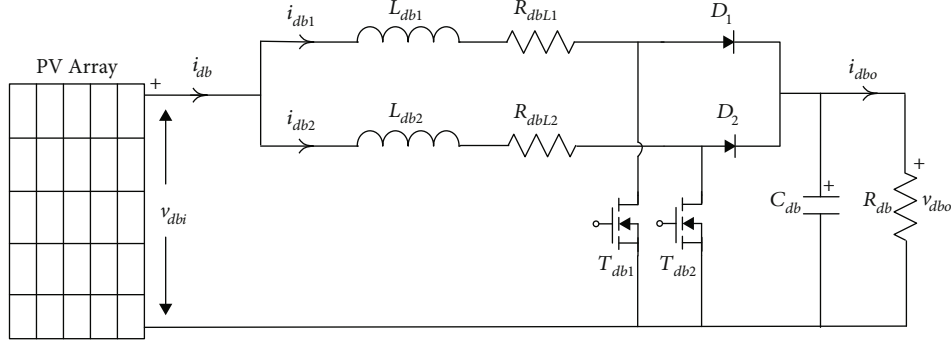


FIGURE 1: Interleaved Boost DC-DC converter.

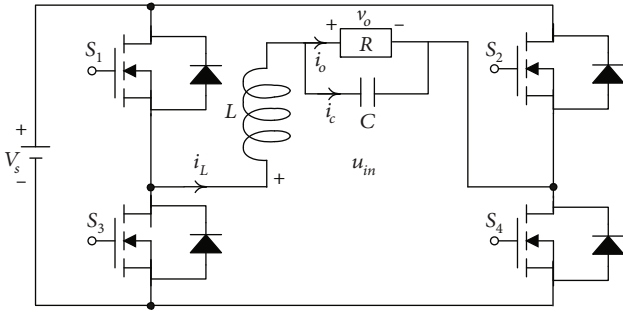


FIGURE 2: Single-phase PWM DC-AC inverter.

$$L\dot{i}_L + v_o = u_{in}, \quad (2)$$

$$i_L = C\dot{v}_o + \frac{v_o}{R}, \quad (3)$$

where u_{in} is equal to $V_s \cdot d$ (time-averaging technique), where d indicates the duty ratio of switching ranging from negative one to positive one and V_s represents the DC link voltage.

The differential equation is obtained by replacing (3) in (2) in the following.

$$\ddot{v}_o + \frac{1}{RC}\dot{v}_o + \frac{1}{LC}v_o = \frac{1}{LC}u_{in}. \quad (4)$$

The state variables are specified $x_1 = v_o$ and $x_2 = \dot{v}_o$, whereupon (4) can be formulated as the result of the state-space form below:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{1}{LC} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{LC} \end{bmatrix} u_{in}. \quad (5)$$

The design problem of the inverter is equivalent to a typical tracking control issue since the output voltage of the inverter is a sinusoidal waveform. Allow v_{ref} to be the desired sinusoidal waveform that v_o will have to follow. Through the definition of the error state variables $e_1 = x_1 - v_{ref}$ and $e_2 = \dot{x}_1 - \dot{v}_{ref}$, it is possible to acquire the tracking control error state equation as follows:

$$\begin{bmatrix} \dot{e}_1 \\ \dot{e}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} u_{in} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w, \quad (6)$$

where $a_1 = 1/LC$, $a_2 = 1/RC$, and $b = 1/LC$ and $w = -a_1 v_{ref} - a_2 \dot{v}_{ref} - \ddot{v}_{ref}$ denotes the disturbance, with the variable e_1 being a difference measurement between v_o and v_{ref} . The system (6) will become stable where e_1 eventually converges to zero with the control signal u_{in} being completely designed so that the output of the inverter is kept at the same level as the desired v_{ref} . In order to allow the error states converged to zero, the control law u_{in} can be designed via discrete proportional-integral-differential scheme. The significance of the inverter in the system needs to be specifically pointed out as follows. The inverter is the key equipment of a PV generation system, mainly serving as the interface between the PV arrays as well as the power grid. The absence of an inverter will not create a completely operational and autonomous PV generation system for residential or commercial use. In this system configuration, there are two power-converting stages. The DC-DC converter handles maximum power point tracking (MPPT) as well as adjusts the DC loading voltage. While the grid-connection occurs, the power is produced by PV arrays which are converted by the DC-AC inverter into high-quality AC output. Also, it is worth observing that the sampling rate ($1/\text{sampling time}$) acts as a significant deciding point for the performance of the control. It is advisable to adopt a high sampling rate (which leads to a less total harmonic distortion, even at rectified loads) in order to achieve a superior level of control. It implies the possibility of using a digital control system with the maximum frequency, alternatively an analogue design reflecting the algorithm of control [43–45].

3. Control Design

The problem statement for nonlinear systems using the classical TA is briefly summarized, and then, an improved technique is designed. Consider the following classical TA sliding surface as

$$\Lambda = \dot{x}_1 + \eta x_1^{m_1/m_2}, \quad (7)$$

where x_1 is the system variable, η signifies positive constant,

and $0 < m_1/m_2 < 1$. An SMC law $u = u_{\text{upper}}(x)$, $u_{\text{lower}}(x)$ can be used, representing $\Lambda > 0$, $\Lambda < 0$, which is expressed as driven Λ to the sliding mode $\Lambda = 0$ during finite time. As such, it is possible to control the dynamics of the system as

$$\Lambda = 0 = \dot{x}_1 + \eta x_1^{m_1/m_2}. \quad (8)$$

There is a finite time T_f for getting from the starting state $x_1(0)$ to zero that arises from the following:

$$T_f = \eta^{-1} \frac{|x_1(0)|^{1-m_1/m_2}}{(1-m_1/m_2)}. \quad (9)$$

In this case, it is revealed that the system state is converged to zero for a finite time duration. Taking into account the Jacobian matrix $= \partial \dot{x}_1 / \partial x_1$, there is a convergence of the system states to zero around the balancing point $x_1 = 0$ at finite time. The eigenvalues of the Jacobian matrix are derived to be negatively infinite, as x_1 tends to zero. As the speed of the system trajectory towards the balancing point becomes infinite, this causes time reachability to be limited. Yet, the equivalent control section suffers from the issues listed below. In the case of $\dot{x}_1 \neq 0$, as well as $x_1 = 0$ and $0 < m_1/m_2 < 1$, the control law probably gives rise to a singularity. An unknown control signal is generated by this singularity, which induces an unstable feedback system. (ii) For the condition $0 < m_1/m_2 < 1$, it is possible to have an imaginary number $x_1^{m_1/m_2-1}$.

Therefore, the FTFA is proposed to improve the conservative TA, and its sliding function based on the system dynamics (1) is represented as

$$\Lambda = \tilde{x}_2 + \rho_1^{-1} \|\tilde{x}_2\|^{\kappa+1} + \rho_2^{-1} \dot{\tilde{x}}_2^m, \quad (10)$$

where $\kappa_1 > 0$, $\rho_1 > 0$, $\rho_2 > 0$, and $1 < m < 2$. Simultaneously, it is recommended that the sliding mode reaching law, i.e.,

$$\dot{\Lambda} = -\sigma_1 \Lambda - \sigma_2 |\Lambda|^k \text{sigm}(\Lambda), \quad (11)$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, $0 < k < 1$, and $\text{sigm}(\Lambda) = [2/(1 + e^{-\Lambda/\varepsilon})] - 1$ is a smooth sigmoidal function with a positive constant ε .

Recapitulating (1), the dynamics can be written as $\dot{\tilde{x}} = p(\tilde{x}(t)) + q(\tilde{x}(t))u$; here, p and q denote the functions. Then, it is possible to derive the control law of FTFA to be

$$u = -q^{-1} \left[K_e \tilde{x} + \rho_2 m^{-1} \dot{\tilde{x}}_2^{2-m} \frac{(1 + (\kappa+1) \|\tilde{x}_2\|^\kappa)}{\rho_1} \right] - q^{-1} \left[\sigma_1 \Lambda + \sigma_2 |\Lambda|^k \text{sigm}(\Lambda) \right], \quad (12)$$

where K_e represents the feedback gain of equivalent control, which creates the desired sliding mode while system uncertainties equal to zero. Equation (12) leads to the system state will be compelled towards $\Lambda = 0$, which converges in a finite amount of time.

$$v_{j+1} = W v_j + \ell_1 \text{RAN}_1 (\chi_{pb,j} - \chi_j) + \ell_2 \text{RAN}_2 (\chi_{gb,j} - \chi_j), \quad (13)$$

$$\chi_{j+1} = \chi_j + v_{j+1}, \quad (14)$$

where v_{j+1} represents present flying speed, χ_j indicates present position, χ_{pb} shows individual best position, χ_{gb} stands for global best position, ℓ_1 and ℓ_2 signify learning factor, RAN_1 and RAN_2 are random number amidst zero and one, and W denote inertia weight. Then, let the function $G(\chi)$ advance in the direction of its fastest descent, namely, the direction of the negative gradient $-\nabla G(\chi)$. It will gradually move towards the optimum so that the negative gradient arises from the condition of $(j-1)$ generation to j generation as follows:

$$-\nabla G(\chi) = \frac{[G(\chi_{j-1}) - G(\chi_j)]}{v_j}, \quad (15)$$

where the symbol ∇ stands for gradient operator. One may choose carefully, W is positive in case $G(\chi_{j-1})$ is greater than equal to $G(\chi_j)$; otherwise, W is negative. The proposed algorithm converges fast as well as the searching of the gradient at the global best position. Because there is a constant or minor variation in its fitness value depending on a prescribed iteration number, the computational burden of the proposed algorithm diminishes [46–48]. It is important to note that the voltage, current, and power of the PV array vary with the strength of daylight as well as climate, which impacts both the power output and energy efficiency of the system. As the environmental temperature rises, there will be a reduction in the maximum power of the PV array. However, increasing irradiation levels result in potentially greater maximum PV power. To allow the PV array system to be more efficient, a MPPT function becomes imperative. Once the solar cell module approaches the maximum power point, the derivative of power with respect to voltage becomes zero. Thus, one always operates the PV array at the maximum power point through power converters together with the MPPT algorithm. More precisely, the design of the MPPT controller can be summarized as follows. (i) The overall number of particles and particle locations as well as velocities are initialized and the fitness values can be estimated. (ii) Determine each particle voltage relative to the PV array output power, as well as initialize the group best voltage. (iii) Renew the particles and execute the gradient search in accordance with (13)–(15). (iv) Allow the voltage of all particles in the group to be renewed for one-round iteration g ; afterwards, the g -th particle voltage should be renewed as well as estimate the output power of the PV array working at such a voltage. (v) Update the best voltages of the particles themselves and the group. (vi) Check whether the convergence criterion has been fulfilled. In case of yes, the best solution can be obtained; otherwise, step (iii) is continued.

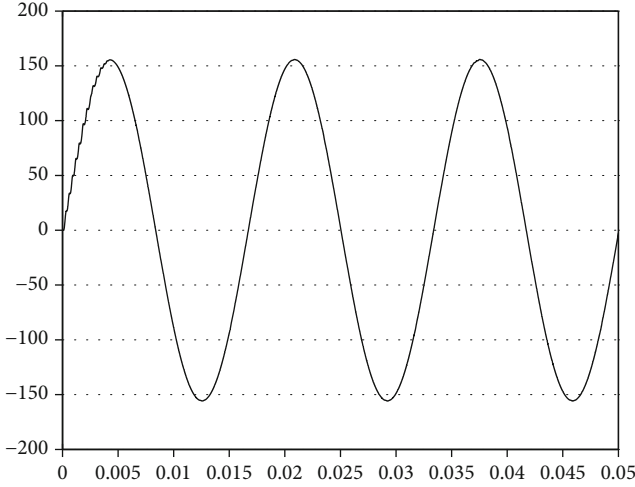


FIGURE 3: Output voltage of a PV array system for the conservative SMC under resistive load (vertical: 50 V/div; horizontal: 5 ms/div).

Proof. The definition of a Lyapunov candidate gives the following $V = \Lambda^2/2$, and based on the dynamical system trajectory along from the control law (10) with the use of the Lyapunov candidate, time derivative of V becomes $\dot{V} = \Lambda \dot{\Lambda} = \Lambda(\dot{\tilde{x}}_2 + \rho_1^{-1} \|\tilde{x}_2\|^{\kappa+1} + \rho_2^{-1} \dot{\tilde{x}}_2^m)'$. Because m is fractional satisfying $1 < m < 2$, and when $\dot{\tilde{x}}_2 \neq 0$, $\dot{\tilde{x}}_2^{m-1} > 0$ is valid. For $\rho_2 > 0$, therefore $\dot{V} \leq 0$ holds. Under the condition $\dot{\tilde{x}}_2 \neq 0$, it is proven that the system fulfills the Lyapunov stability condition which allows it to quickly arrive at the sliding surface within a finite time. Yet, there is quiver or steady-state error in the FTTC. This is caused by drastic variations or great nonlinearities in the system load, which prevent the eventual system output from following the reference sine wave, rendering the tracking behavior imprecise. Aiming to acquire the global best solution, the GPSO algorithm can be used. Equations (13) and (14) represent the model of particle evolution, which can then update the velocity and location of each particle as it flies to its target. \square

4. Results and Discussion

In order to verify the effectiveness of the proposed controller, the system is tested under steady-state response for full resistive loads and rectifier loads and transient response for step load changing. Figure 3 shows the output voltage (%THD = 0.05%) of the PV array system with a resistive load $R = 12 \Omega$ for the conservative SMC. It can be seen that the output voltage is very close to a sine wave. Figure 4 shows the output voltage (%THD = 0.02%) of the proposed controller for the PV array system with a resistive load $R = 12 \Omega$. Similarly, the output-voltage waveform is very close to a sine waveform. Figure 5 shows the output voltage of a PV array system controlled by the conservative SMC at a trigger angle, changing from no load to full load ($R = 12 \Omega$). The figure shows that the transient voltage drop does not recover quickly and the controller is not able to compensate well. Figure 6 shows the output voltage of the PV array system using the proposed controller at trigger angle, changing from no load to full load

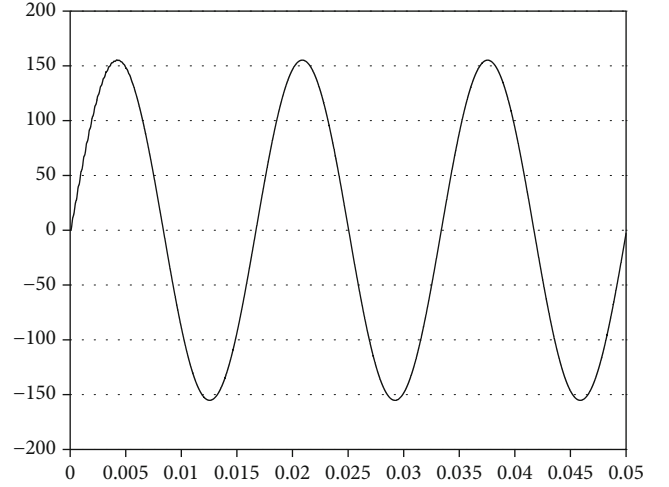


FIGURE 4: Output voltage of a PV array system for the proposed controller under resistive load (vertical: 50 V/div; horizontal: 5 ms/div).

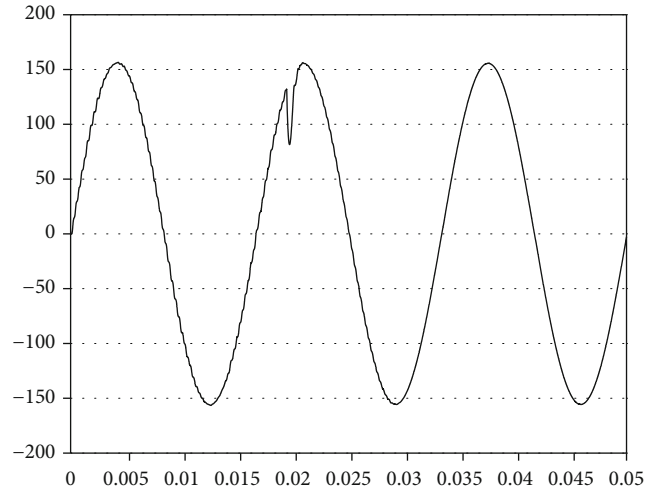


FIGURE 5: Output voltage of a PV array system for the conservative SMC under step load changing (vertical: 50 V/div; horizontal: 5 ms/div).

($R = 12 \Omega$). The graph shows that the transient voltage drop is recovered within a very short period of time with excellent compensation capability. Figure 7 shows the output voltage of the PV array system under rectifier-type load with the conservative SMC. It can be seen that the output voltage is a distorted sine waveform with a high %THD value of 12.36%. Figure 8 shows the output waveform of the PV array system controlled by the proposed controller under rectified load. The output-voltage waveform is very close to the required sinusoidal reference voltage (low %THD value of 0.08%). It can be seen that the proposed PV array system has a better performance than the conservative sliding mode controlled PV array system. A comparison is listed in Table 1 of the voltage drops and THD values of the conservative SMC and the proposed controller for transient loading and steady-state loading. The proposed controller reveals small voltage drop

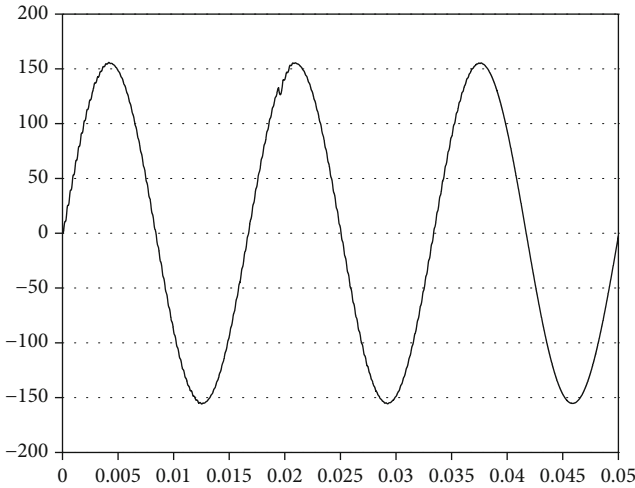


FIGURE 6: Output voltage of a PV array system for the proposed controller under step load changing (vertical: 50 V/div; horizontal: 5 ms/div).

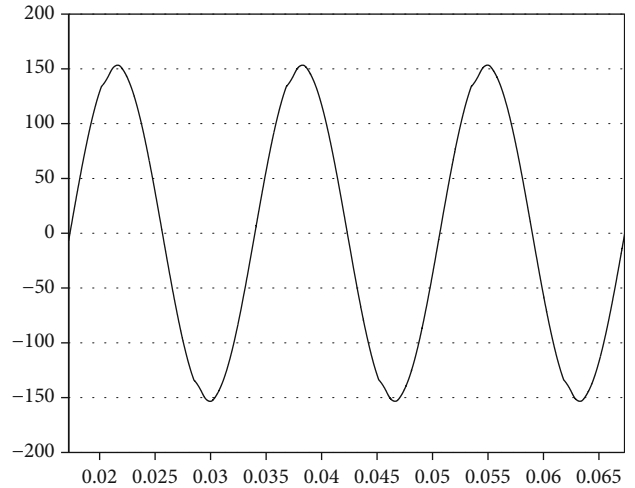


FIGURE 8: Output voltage of a PV array system for the proposed controller under rectifier load (vertical: 50 V/div; horizontal: 5 ms/div).

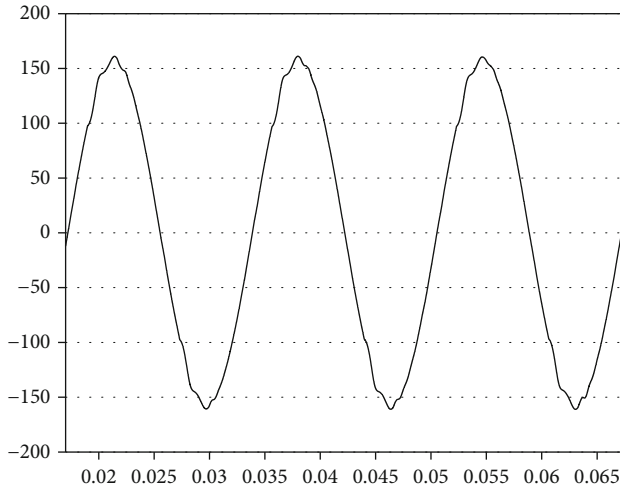


FIGURE 7: Output voltage of a PV array system for the conservative SMC under rectifier load (vertical: 50 V/div; horizontal: 5 ms/div).

and low THD values with the voltage waveforms close to the required sinusoidal reference voltages. However, in the case of rectified loads, the output voltage of the PV array system controlled by the conservative SMC has a distortion rate of more than 5%, while the proposed PV array system has a distortion rate of less than 5%, which is better than the 519 harmonic control standard set by the American Institute of Electrical and Electronics Engineers. The results are compared with existing work hereunder. A SMC is designed for active and reactive power control of the renewable energy generations, while the PSO algorithm allows optimizing the SMC parameters. The system can generate real power within 1% of the required power at a fast time, indicating stable power. Although the proposed method is capable of producing acceptable transience and steady-state actions, the speed of the system state driving to the sliding surface remains to be refined [49]. The single-phase inverter based on the SMC is

TABLE 1: Output voltage drop and %THD.

	Proposed controller	Conservative SMC
Step loads Voltage drop	$6.67V_{\max}$	$51.3V_{\max}$
Full resistive load %THD	0.02%	0.05%
Rectifier load %THD	0.08%	12.36%

connected to the grid with the purpose of establishing optimized PV delivery. Meanwhile, the overall circuit structure uses a boost converter with a PSO algorithm optimized fuzzy logic scheme, which yields the MPPT. The presented inverter provides quick dynamic response; nevertheless, the steady state should be investigated under severely nonlinear loading perturbations [50]. A double feedback controller realized by the SMC and proportional-integral control has been applied to the microsource inverter, where a PSO algorithm is employed for the optimization of the control parameters. The sliding-mode reaching law does not precisely strike the requisite sliding manifold, so the resulting waveforms still suffer from considerable distortion [51]. A brushless DC motor actuated water pumping system dependent on the PV supply has been proposed. This system offers several features such as no additional DC-DC converter hardware, sliding mode controller implementing sensorless speed controller, and hybrid whale optimization-perturbation and observation algorithm getting the maximum available power either in partially shaded and normally operating conditions. Because the applied technology fails to remove the chattering effect, there is a steady-state error around the state trajectory [52]. The passive-based control has been developed for grid-connected inverters. The PSO algorithm and Kalman filter observer give the simplification of the passive-based control parameters design and the diminution of sensors number, respectively. It has revealed an improvement in the total harmonic distortion and dynamic behavior both during steady state and transient loading; however, a little oscillation

in the output waveform appears due to the potential for premature convergence with conventional PSO [53]. In spite of the fact that the proposed system's performance does not greatly overtake the THD results from the latest reports, the proposed controller is unsophisticated, straightforward to comprehend, and quick to converge, thus affording better target tracking control. It can be a valuable referral for readers concerned with renewable energy converters as well as AIoT applications.

5. Conclusions

With the proposed controller, it is possible to employ a gradient particle swarm optimization algorithm to detect the global maximum power point of the photovoltaic array in the presence of partial shading while using the unique advantages of the finite-time terminal attractor to offer fast convergence of the system state under uncertain interferences, quiver, and reduction of steady-state errors for tracking control. The simulated operating conditions of the photovoltaic arrays, such as voltage, current, and power, are also stored, collated, monitored, and analysed by the IoT platform to enhance the reliability and integrity of the system. The controller developed in this manner can create the maximum power output from the solar panel and retain the highest energy conversion efficiency in the event of partial shading. The proposed controller has good steady-state and transient response behavior in terms of total harmonic distortion and momentary voltage drop under rectifier loading and step load varying conditions, displaying the waveforms are close to the required sinusoidal reference voltage. However, the output voltage of the conservative SMC of the PV array system suffers from higher than 5% THD and greater voltage drop for the same testing load conditions. As a result, the proposed PV array system actually yields promising performance when subjected to partial shading circumstances. This paper has innovative highlights which can be summarized as follows. (i) The proposed controller can be used not only for interleaved Boost DC-DC converter but also directly for Cuk, Sepic, and Zeta DC-DC converters, which provides the advantage of extended applications. (ii) The proposed controller incorporating AIoT allows the information analysis of the voltage, current, and power in the PV system, which will be the future trend of smart city development. (iii) The proposed controller is capable of yielding high-quality AC supply, which can be fully adopted in the future for various renewable energy sources, such as wind and hydrogen energies.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgments

The authors gratefully acknowledge the financial support of the Ministry of Science and Technology of Taiwan, R.O.C., under project number MOST 110-2221-E-214-021.

References

- [1] M. Saad and A. M. Eltamaly, *Advanced Technologies for Solar Photovoltaics Energy Systems*, Springer, Switzerland, 2021.
- [2] A. J. Yaman and H. Eklas, *Photovoltaic Systems: Fundamentals and Applications*, Springer, Switzerland, 2022.
- [3] M. Moh, K. P. Sharma, R. Agrawal, and V. G. Diaz, *Smart IoT for Research and Industry*, Springer, Switzerland, 2022.
- [4] K. Vinod, K. B. Ranjan, J. Dheeraj, and B. Ramesh, *Power Electronics, Drives, and Advanced Applications*, CRC Press, Boca Raton, FL, USA, 2020.
- [5] H. Yang, *Modeling and Control of Power Electronic Converters for Microgrid Applications*, Springer, Switzerland, 2022.
- [6] F. L. Luo and H. Ye, *Power Electronics: Advanced Conversion Technologies*, CRC Press, Boca Raton, FL, USA, 2018.
- [7] E.-C. Chang, "Applying robust intelligent algorithm and Internet of things to global maximum power point tracking of solar photovoltaic systems," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8882482, 10 pages, 2020.
- [8] A. R. Gautam, L. Umanand, and R. B. Subba, "Current control of boost converter for PV interface with momentum-based perturb and observe MPPT," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 4071–4079, 2021.
- [9] B. Shamik, S. K. P. Dattu, S. Susovan, and M. Sukumar, "Steady output and fast tracking MPPT (SOFT-MPPT) for P&O and InC algorithms," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 1, pp. 293–302, 2021.
- [10] P. Imran, S. Immad, M. Saad, S. Adil, T. Mohd, and A. Basem, "Most valuable player algorithm based maximum power point tracking for a partially shaded PV generation system," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 1876–1890, 2021.
- [11] B. Badreddine, A. Fahad, H. Nouredine, K. Sami, and S. M. Ghoneim, "Fractional-fuzzy PID control approach of photovoltaic-wire feeder system (PV-WFS): simulation and HIL-based experimental investigation," *IEEE Access*, vol. 9, pp. 159933–159954, 2021.
- [12] V. I. Utkin, "Variable structure systems with sliding modes," *IEEE Transactions on Automatic Control*, vol. 22, no. 2, pp. 212–222, 1977.
- [13] M. Ayaykumar and B. Bijan, *Emerging Trends in Sliding Mode Control: Theory and Application*, Springer, Singapore, 2021.
- [14] U. Itkis, *Control Systems of Variable Structure*, Wiley, New York, 1976.
- [15] C. Y. Chan, "Servo-systems with discrete-variable structure control," *Systems Control Letters*, vol. 17, no. 4, pp. 321–325, 1991.
- [16] H. Sira-Ramirez, "Sliding regimes in general non-linear systems: a relative degree approach," *International Journal of Control*, vol. 50, no. 4, pp. 1487–1506, 1988.
- [17] Z. Doulgeri, "Sliding regime of a nonlinear robust controller for robot manipulators," *IEE Proceedings-Control Theory and Applications*, vol. 146, no. 6, pp. 493–498, 1999.
- [18] G. Bartolini, L. Fridman, A. Pisano, and E. Usai, *Modern Sliding Mode Control Theory*, Springer-Verlag, Berlin, 2008.
- [19] R. Chinnappan, P. Logamani, and R. Ramasubbu, "Fixed frequency integral sliding-mode current-controlled MPPT boost converter for two-stage PV generation system," *IET Circuits, Devices & Systems*, vol. 13, no. 6, pp. 793–805, 2019.

- [20] B. Banerjee and W. W. Weaver, "Generalized geometric control manifolds of power converters in a DC microgrid," *IEEE Transactions on Energy Conversion*, vol. 29, no. 4, pp. 904–912, 2014.
- [21] H. F. Feshara, A. M. Ibrahim, N. H. El-Amary, and S. M. Sharaf, "Performance evaluation of variable structure controller based on sliding mode technique for a grid-connected solar network," *IEEE Access*, vol. 7, pp. 84349–84359, 2019.
- [22] M. Alsumiri, "Residual incremental conductance based non-parametric MPPT control for solar photovoltaic energy conversion system," *IEEE Access*, vol. 7, pp. 87901–87906, 2019.
- [23] P. Peltoniemi, P. Nuutinen, and J. Pyrhonen, "Observer-based output voltage control for DC power distribution purposes," *IEEE transactions on power electronics*, vol. 28, no. 4, pp. 1914–1926, 2013.
- [24] X. B. Wu, Q. Liu, M. L. Zhao, and M. Y. Chen, "Monolithic quasi-sliding-mode controller for SIDO buck converter with a self-adaptive free-wheeling current level," *Journal of Semiconductors*, vol. 34, no. 1, 2013.
- [25] L. Shen, D. D. Lu, and C. Li, "Adaptive sliding mode control method for DC-DC converters," *IET Power Electronics*, vol. 8, no. 9, pp. 1723–1732, 2015.
- [26] X. D. Du, X. Fang, and F. Liu, "Continuous full-order nonsingular terminal sliding mode control for systems with matched and mismatched disturbances," *IEEE Access*, vol. 7, pp. 130970–130976, 2019.
- [27] V. C. Nguyen, A. T. Vo, and H. J. Kang, "A finite-time fault-tolerant control using non-singular fast terminal sliding mode control and third-order sliding mode observer for robotic manipulators," *IEEE Access*, vol. 9, pp. 31225–31235, 2021.
- [28] H. Z. Hou, X. H. Yu, L. Xu, R. Kamal, and Z. W. Cao, "Finite-time continuous terminal sliding mode control of servo motor systems," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5647–5656, 2020.
- [29] P. D. Ge, Y. Zhu, C. G. Tim, and F. Teng, "Resilient secondary voltage control of islanded microgrids: an ESKBF-based distributed fast terminal sliding mode control approach," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1059–1070, 2021.
- [30] S. K. Cherukuri and S. R. Rayapudi, "Enhanced grey wolf optimizer based MPPT algorithm of PV system under partial shaded condition," *International Journal of Renewable Energy Development*, vol. 6, no. 3, pp. 203–212, 2017.
- [31] A. Yahiaoui, F. Fodhil, K. Benmansour, M. Tadjine, and N. Cheggaga, "Grey wolf optimizer for optimal design of hybrid renewable energy system PV- diesel generator-battery: application to the case of Djanet city of Algeria," *Solar Energy*, vol. 158, pp. 941–951, 2017.
- [32] N. Zhao, C. Roberts, and S. Hillmansen, "The application of an enhanced brute force algorithm to minimise energy costs and train delays for differing railway train control systems," *Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit*, vol. 2288, no. 2, pp. 158–168, 2014.
- [33] Y. Y. Jia, J. Q. Wang, and Q. Y. Xiao, "Multiple groups of gradient particle swarm optimization and its application in optimal operation of reservoir," in *Proceedings of the 2014 10th International Conference on Natural Computation (ICNC)*, pp. 622–626, Xiamen, China, 2014.
- [34] B. T. Zhang and R. Yahya, "Including design sensitivity into particle swarm optimization: adaptive gradient-based techniques," in *Proceedings of the 2015 Asia-Pacific Microwave Conference (APMC)*, pp. 1–3, Nanjing, China, 2015.
- [35] T. A. Loau, C. P. Jagdish, and S. Alex, "Solving economic dispatch problem under valve-point loading effects and generation constraints using a multi-gradient PSO algorithm," in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Rio de Janeiro, Brazil, 2018.
- [36] N. Subhprattim, S. Aditya, S. Ritankar, B. Suharta, K. S. Jamuna, and K. S. Subir, "Minimizing wirelength with bend reduction using gradient descent PSO hybrid in VLSI global routing," in *Proceedings of the 2021 Devices for Integrated Circuit (DevIC)*, pp. 401–405, Kalyani, India, 2021.
- [37] D. L. Cavalca and A. S. Fernandes, "Gradient-based mechanism for PSO algorithm: a comparative study on numerical benchmarks," in *Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–7, Rio de Janeiro, Brazil, 2018.
- [38] U. Bartłomiej, K. Arkadiusz, and M. G. Lech, "Particle swarm optimization of the multioscillatory LQR for a three-phase four-wire voltage-source inverter with an LC output filter," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 484–493, 2015.
- [39] V. Mummadi and R. S. Anmol, "Optimized power stage design of low source current ripple fourth-order boost DC-DC converter: a PSO approach," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 3, pp. 1491–1502, 2015.
- [40] K. E. Parsopoulos and M. N. Vrahatis, *Particle Swarm Optimization and Intelligence: Advances and Applications*, Information Science Reference, Hershey, PA, USA, 2010.
- [41] K. Mohammad, T. Shamsodin, A. Cretu, H. Seyedkazem, and P. Edris, "PSO-based modeling and analysis of electrical characteristics of photovoltaic module under nonuniform snow patterns," *IEEE Access*, vol. 8, pp. 197484–197498, 2020.
- [42] H. C. Shi, H. Q. Wen, Y. H. Hu, and L. Jiang, "Reactive power minimization in bidirectional DC-DC converters using a unified-phasor-based particle swarm optimization," *IEEE Transactions on Power Electronics*, vol. 33, no. 12, pp. 10990–11006, 2018.
- [43] Y. W. Tzou and S. L. Jung, "Full control of a PWM DC-AC converter for AC voltage regulation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 4, pp. 1218–1226, 1998.
- [44] H. S. Kim and S. K. Sul, "A novel filter design for output LC filters of PWM inverters," *Journal of Power Electronics*, vol. 11, no. 1, pp. 74–81, 2011.
- [45] M. Blachuta, B. Robert, and G. Rafal, "Sampling rate and performance of DC/AC inverters with digital PID control—a case study," *Energies*, vol. 14, no. 16, pp. 5170–5222, 2021.
- [46] M. M. Noel and T. C. Jannett, "Simulation of a new hybrid particle swarm optimization algorithm," in *Proceedings of the 36th Southeastern Symposium on System Theory (SSST04)*, pp. 150–153, Atlanta, GA, USA, 2004.
- [47] M. M. Noel, "A new gradient based particle swarm optimization algorithm for accurate computation of global minimum," *Applied Soft Computing*, vol. 12, no. 1, pp. 353–359, 2012.
- [48] G. Cao, M. K. Lai, and A. Fakhru, "Enhanced particle swarm optimisation algorithms for multiple-input multiple-output system modelling using convolved Gaussian process models," *International Journal of Intelligent Systems Technologies and Applications*, vol. 17, no. 3, pp. 1–17, 2018.

- [49] S. Yousaf, A. Mughees, M. G. Khan, A. A. Amin, and M. Adnan, "A comparative analysis of various controller techniques for optimal control of smart nano-grid using GA and PSO algorithms," *IEEE Access*, vol. 8, pp. 205696–205711, 2020.
- [50] A. Borni, A. Bouchakour, L. Zaghba, A. Thameur, A. Lakhdari, and N. Bessous, "Optimization of the fuzzy MPPT controller by PSO for the single-phase grid-connected photovoltaic system controlled by sliding mode," in *Proceedings of the 2018 6th International Renewable and Sustainable Energy Conference (IRSEC)*, pp. 1–7, Rabat, Morocco, 2018.
- [51] J. Bian, C. Z. Zang, X. H. Li, B. Hu, and J. Liu, "The rapid development of three-phase grid-forming micro-source inverter based on SMC and PI control," in *Proceedings of the 2017 China International Electrical and Energy Conference (CIEEC)*, pp. 1–6, Beijing, China, 2017.
- [52] S. G. Malla, P. Malla, J. M. R. Malla et al., "Whale optimization algorithm for PV based water pumping system driven by BLDC motor using sliding mode controller," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, pp. 1–12, 2022.
- [53] F. B. Zheng, W. M. Wu, B. L. Chen, and E. Koutroulis, "An optimized parameter design method for passivity-based control in a LCL-filtered grid-connected inverter," *IEEE Access*, vol. 8, pp. 189878–189890, 2020.

Research Article

A Study on the Optimization Simulation of Big Data Video Image Keyframes in Motion Models

Jianbang Guo,¹ Peng Sun ,² and Sang-Bing Tsai ³

¹Athletics College, Beijing Sport University, Beijing 100084, China

²Physical Education College, Anqing Normal University, Anhui 246133, China

³Regional Green Economy Development Research Center, School of Business, Wuyi University, China

Correspondence should be addressed to Peng Sun; 18640131611@163.com

Received 11 January 2022; Revised 20 February 2022; Accepted 25 February 2022; Published 16 March 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 Jianbang Guo 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, the signal of athletic sports video image frames is processed and studied according to the technology of big data. The sports video image-multiprocessing technology achieves interference-free research and analysis of sports technology and can meet multiple visual needs of sports technology analysis and evaluation through key technologies such as split-screen synchronous comparison, superimposed synchronous comparison, and video trajectory tracking. The sports video image-processing technology realizes the rapid extraction of key technical parameters of the sports scene, the panoramic map technology of sports video images, the split-lane calibration technology, and the development of special video image analysis software that is innovative in the field of athletics research. An image-blending approach is proposed to alleviate the problem of simple and complex background data imbalance, while enhancing the generalization ability of the network trained using small-scale datasets. Local detail features of the target are introduced in the online-tracking process by an efficient block-filter network. Moreover, online hard-sample learning is utilized to avoid the interference of similar objects to the tracker, thus improving the overall tracking performance. For the feature extraction problem of fuzzy videos, this paper proposes a fuzzy kernel extraction scheme based on the low-rank theory. The scheme fuses multiple fuzzy kernels of keyframe images by low-rank decomposition and then deblurs the video. Next, a double-detection mechanism is used to detect tampering points on the blurred video frames. Finally, the video-tampering points are located, and the specific way of video tampering is determined. Experiments on two public video databases and self-recorded videos show that the method is robust in fuzzy video forgery detection, and the efficiency of fuzzy video detection is improved compared to traditional video forgery detection methods.

1. Introduction

Vision is one of the most important ways for humans to perceive information about the external world. By imitating the human visual perception system, humans have created various imaging tools (e.g., cameras, depth sensors, and surveillance cameras) to obtain video image data, and these imaging devices give machines the ability to perceive the external world. These imaging devices give machines the ability to perceive the external world, while how to further analyze and understand the acquired images and video data needs to be implemented by computer vision-related algorithms [1]. Computer vision combines applied mathematics and statistics, digital signal processing, and other related

theoretical foundations to analyze the image and video data acquired by visual imaging devices to achieve machine understanding of the objective world. The research of computer vision is of great importance for the realization of machine intelligence. As an indispensable and important part of computer vision, visual target tracking not only has received great attention in the academic field but also has a wide application prospect in the fields of national defense and military, transportation, video surveillance, human-computer interaction, and automatic driving [2]. On the other hand, visual target tracking techniques can be used to localize and describe the motion trajectory of targets in videos and are the basis for higher-level tasks such as video understanding and behavior recognition. The application

scenario of target tracking requires that the tracking algorithm needs to be real-time in nature; otherwise, the algorithm would have no practical application value. In addition, the accuracy and robustness of tracking are also important metrics for tracking algorithms [3]. However, due to the diversity and complexity of realistic scenarios, designing a robust real-time target tracking algorithm has been an extremely challenging task. The main challenges of visual target tracking come from two aspects. Haar is a feature description that has evolved over time from the three simple features of Haar Basic to Haar-Like and now Haar Extended, where the feature template contains both white and black rectangles and defines the feature value of the template as the sum of the white rectangle pixels minus the sum of the black rectangle pixels. The Haar feature values reflect the greyscale variation of the image. Variation of the tracking target is as follows: scale change, nonrigid deformation, and fast motion. Although target tracking algorithms have been greatly developed in recent years and most tracking algorithms can cope with limited scenes and specific objects, there is still room for improvement in complex scenarios.

All this shows that video coding and transmission technology are constantly developing and progressing. Video information is technically required to face a wide variety of demands in practical applications, which puts forward higher requirements for video coding technology, and this indicates that video coding technology will usher in new opportunities and challenges. Current video compression techniques are based on hybrid coding frameworks to improve coding efficiency through motion compensation, predictive coding, transform quantization, and entropy coding techniques [4]. The convergence of acquisition, computation, and cognitive techniques has made intelligent coding possible. Thus, there is now also intelligent coding based on neural networks that apply intelligent techniques to traditional coding frameworks, and there is also research being conducted on feature-based coding, i.e., texture feature cocoding. Facing more complex scenes and more demands for compression of video information data, the improvement of traditional coding techniques and the introduction of new coding techniques are also hot topics of research. In addition, the video can generally achieve acceptable quality after compression by standard techniques, but in the face of complex video content, flexible video scenes, real-time dissemination, and other complex situations cannot guarantee good video quality, which requires the use of other video coding optimization and control techniques. Among them, rate-distortion optimization and bit rate control techniques are very effective optimization and control techniques. Distortion optimization is to weigh the coding consuming bits as well as information distortion, expecting the least number of bits to be consumed with limited distortion [5]. The most intuitive role in encoders is to guide the encoder to select the optimal coding parameters from multiple coding candidate configurations according to a specific strategy to achieve optimal coding performance. The code rate control technique is the study of setting the appropriate quantization parameters for the coded image group, the coded image, and even the coding unit for various specific code rate

requirements under the prerequisite conditions so that the coded code rate can conform to the initially set code rate while ensuring that the output video image quality is as stable as possible. The coding optimization technique is not part of the coding standard, but in the actual application scenario, the coding content is very complex, and different users have different needs, so the coding optimization technique is crucial. There are currently two main categories of methods for face detection: knowledge-based and statistics-based. Knowledge-based methods are as follows: mainly use a priori knowledge to view faces as a combination of organ features and detect faces based on the features of organs such as eyes, eyebrows, mouth, and nose and the geometric position relationship between them. The main methods include template matching, face features, shape and edge, texture features, and color features. Statistical-based methods are as follows: The face is regarded as a whole pattern—a two-dimensional pixel matrix—and the face pattern space is constructed from a statistical point of view through a large number of face image samples to determine whether a face exists based on a similarity measure. The main methods include principal component analysis and feature faces, neural network methods, support vector machines, hidden Markov models, and AdaBoost algorithms.

More detailed and comprehensive game information is provided to the sports team at the game site or after the game, to achieve the purpose of scientific research and monitoring to study and evaluate the individual athletes' technical and tactical performance and team technical and tactical ability and to improve the athletes' technical and tactical level and team athletic ability. This is one of the most common and frequently used methods and means of video image-multiprocessing technology in sports training practice. At present, video image-processing technology has been widely used in any of the competitive sports teams in China and has also been supported and recognized by most coaches, athletes, and scientific research staff, which has played a very important role in helping and guaranteeing the improvement of athletes' technical and tactical ability and level. In the face of a large number of sports video image-multiprocessing technology means and forms, it is very necessary to conduct a more scientific and reasonable in-depth research and summary of the sports video image-processing technology system, to better serve the scientific sports practice in the field of sports in China, to use sports video image-processing technology to better guide and evaluate the technical and tactical ability of athletes, scientific monitoring and evaluation of modern sports science training and competition practice activities, and at the same time help athletes and coaches to improve their technical and tactical analysis and research ability, improve their ability to monitor and evaluate technical and tactical skills, and better improve the role and benefits of sports technical video images in the scientific process of modern sports training.

2. Current Status of Research

Video image-processing technology in a broad sense refers to various technologies related to video image processing

in general, including the physical processing technology of video images and related video image-processing software and hardware development and applications [6]. At present, people mainly study digital video image-processing technology, and the main application is modern computer technology and video image-multiprocessing technology [7]. This includes a complete, orderly and tightly organized, and programmed set of systematic work using computers and other electronic devices, such as video image acquisition, video compression and coding, video clip decomposition and synthesis, video image format conversion and unification, video annotation and identification, video storage and transportation, image nonlinear editing and generation, video image display and output, video image transformation and enhancement, recovery (restoration) and reconstruction of video images, segmentation of video images and detection of targets, representation and description of video images, extraction and measurement of video image feature frames, correction and translation of video image sequences, reconstruction and restoration of 3D scenes, development and creation of video image databases, indexing and extraction of video image databases, classification representation and recognition of video images, video image model building and matching, interpretation and understanding of video image transitions scenes, and judgmental decision making and behavior planning of video images. An outline can be described as a curve that combines all consecutive points with the same color or intensity. They show the shape of the objects contained in the picture. Contour detection may be a useful technique for shape analysis and object detection and recognition. Contour detection is not the only algorithm for image segmentation; there are many others such as semantic segmentation, the Hough transform, and K -means segmentation, which are currently state-of-the-art. As one of the most promising and active research directions in the field of computer vision at home and abroad, video human behavior recognition, with the in-depth development of artificial intelligence and computers, has been able to initially realize the use of key nodes, images, videos, and other data to identify human behavior [8]. Computer vision is a key AI research area, and researchers are making increased contributions to it. In recent years, major internet companies, as well as universities, have paid close attention to the development trend of human behavior recognition technology in the international frontier, mainly focusing on three directions: dynamic region detection, modeling, and classification recognition [9].

One of the research directions that is becoming of increasing importance and interest in the field of computer vision is video stream-based behavior analysis and understanding [10]. The core research focus is on the use of visual pattern recognition image signal processing techniques and other related techniques to perform processing of video stream target sequences for target detection, target classification, target tracking, and thus the behavioral analysis of human behavior in video surveillance regions [11]. The basis of video surveillance image-processing systems is target detection, for which many more mature detection algorithms are available and which is a low-level computer vision

problem [12]. Target classification must accurately classify moving objects in the scene so that moving objects in the scene can be further tracked and analyzed. And one of the main bottlenecks that currently restrict the development of video surveillance image-processing systems is target tracking, which is one of the most basic functions in video surveillance image-processing systems [13]. And one of the research hotspots that has been widely focused on in recent years is how to analyze and identify human behavior, which focuses on how to analyze and identify human behavioral motion patterns, which can be regarded as the problem of classifying and matching time-varying motion data, i.e., matching test sequences with precalibrated reference data sequences of standard operations [14].

This method focuses on two adjacent image frames and performs a differential operation between consecutive frames to obtain a feature representation of human behavior. The advantage of this approach is that it preserves the temporal features of human behavior in the video, but this depends to some extent on the manual segmentation of human contours, is sensitive to color, lighting, contrast, and occlusion issues, and is only applicable to video scenes with limited space. Video behavior recognition algorithms based on local features of motion behavior do not require presegmentation processing of video images. Common local features include edges, corners, curves, and regions with special properties. The step length and speed of a walk can change continuously. The step length of a walk can change from time to time, some steps can be short and some can be long, and the speed can change from slow to fast or fast to slow and can be done in different background environments. Thus, there are many types of behavioral expressions and many variations of each behavior, which pose some problems for research in human behavior recognition.

3. Analysis of Frame Signal Processing for Athletics' Video Images with Big Data

3.1. Big Data Processing Analysis of Athletics' Video Images. Current digital multimedia forensic techniques can be divided into two main categories: active forensics and passive forensics. Active forensics refers to the use of information hiding technology, which embeds imperceptible verification information, such as digital watermark, digital signature, and perceptual hashing, in the process of image or video generation in advance [15]. The party receiving the image or video determines the reliability of the received data by analyzing whether this embedded information has been corrupted. Active forensic techniques have good detection effects but, at the same time, have significant limitations in practical applications; many imaging devices cannot embed signals such as watermarks, and the embedded information may be removed or modified and reembedded. Passive forensics, also known as blind forensics, does not require information to be embedded in the multimedia data in advance as inactive forensics. Passive forensics directly analyzes the information of the received multimedia data itself, because multimedia data will inevitably leave some traces of tampering in the process of tampering, although

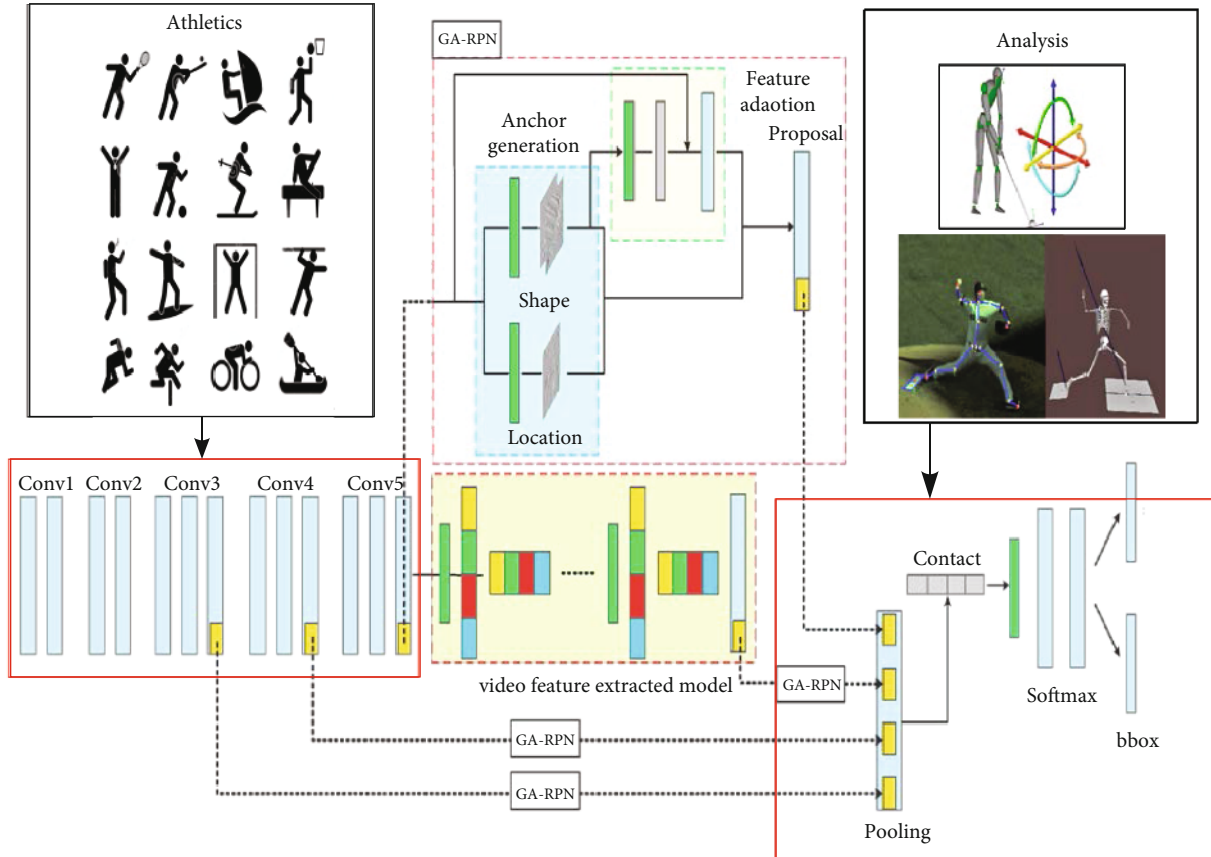


FIGURE 1: Schematic diagram of the three levels of video image-processing engineering.

these traces cannot be detected by the eye; however, extracting certain statistical features of itself can verify these traces. Based on such a principle, passive forensic techniques can detect the authenticity of a video based on whether the encoding features, statistical features, or other characteristics of the video change before and after tampering [16–18].

Digital multimedia forensic techniques can be divided into two main categories: digital image forensics and digital video forensics. Among them, the technology of digital image forensics has been developed earlier and is becoming increasingly mature. And digital video forensics technology requires higher computational complexity of algorithms for digital video-tampering forensics because of the large amount of video data. There are also more ways of video tampering than image tampering and complex video codecs, which makes the complexity of digital video-tampering detection greater. Nowadays, digital video forensics is also being developed gradually. Video tampering refers to the use of video editing software, and the content of the video is maliciously modified to achieve the effect of disguise or “create something out of nothing.” Digital video forensics is the use of relevant video-tampering detection algorithms to verify the integrity and authenticity of a video, as shown in Figure 1.

In the practical application of video image-processing technology for research work, the researcher generally does not need to have a very deep understanding and mastery of complex video image-processing processes, formulas,

and principles; especially, the field of track and field scientific research is a mainly innovative application of software and hardware systems and equipment related to video image-processing technology, to achieve the purpose of improving the athletic ability of outstanding athletes, to achieve for the field of track and field scientific research within the purpose of scientific and practical work. The video image-processing technology in a narrow sense only refers to the general term for the technology of direct processing and processing of video images. AdaBoost stands for “Adaptive Boosting,” which is adaptive in the sense that the weights of the incorrectly classified samples of the previous basic classifier are increased, while the weights of the correctly classified samples are decreased and used again to train the next basic classifier. At the same time, a new weak classifier is added in each round of iterations until some predetermined sufficiently small error rate is reached or a prespecified maximum number of iterations is reached before the final strong classifier is determined. It mainly includes the acquisition, compression, decomposition, synthesis, coding, storage, and transportation of video images, and the display and output of video images, and also the transformation, enhancement, recovery, and reconstruction of video images, the segmentation of video images, the detection, expression, and description of targets, the extraction and measurement of features, the correction of sequence images, the reconstruction and recovery of 3D scenery, and the establishment and classification of video

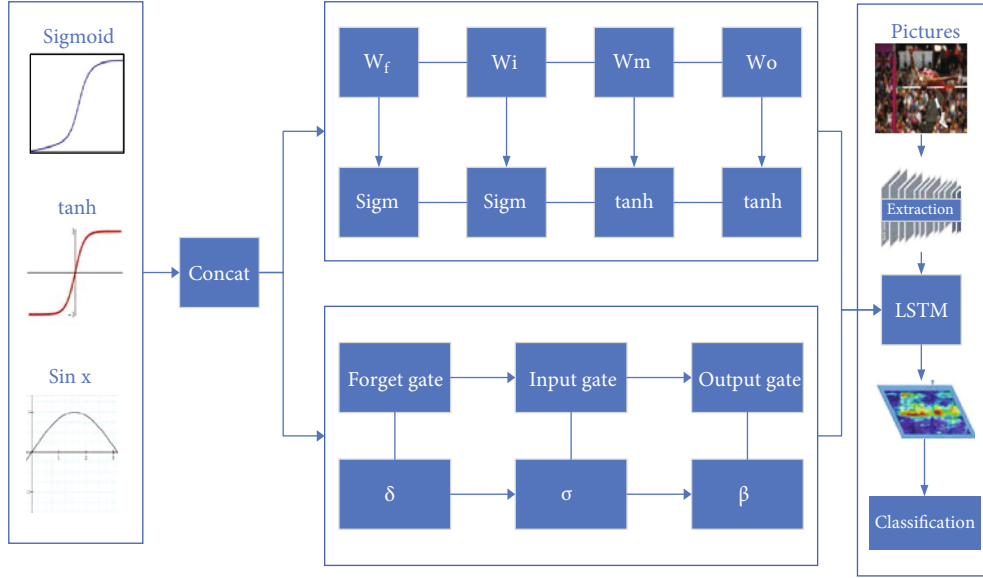


FIGURE 2: Structure of the LSTM cell.

image databases. These are all processing multiple techniques for processing the physical scenes of video images to achieve the corresponding work purpose. The convolutional layer consists of a set of convolutional kernels (each neuron as a kernel). These kernels are associated with a small region of the image, which is called the receptive field. It works by segmenting the image into small pieces (receptive fields) and convolving them with a specific set of weights (multiplying the elements of the filter (weights) with the corresponding receptive field elements).

$$F_l^k = (I_{x,y} \times K_l^k). \quad (1)$$

Once features are extracted, only their approximate positions relative to other features need to be retained, and their exact positions become less important. Pooling layers can summarize similar information within a neighborhood of the perceptual field and output the dominant response within that region. The use of pooling operations helps to extract features that are invariant to translational shifts and small distortions. In addition, pooling can also help improve generalization by reducing overfitting. In addition, reducing the size of the feature map can also regulate the complexity of the network.

$$Z_l = f_p F_{x,y,l}^k. \quad (2)$$

It is a decision function that helps in learning complex patterns. Choosing the right activation function can speed up the learning process. The activation function for the convolutional feature map is defined.

$$T_l^k = f_A F_{x,y,k}^l. \quad (3)$$

For CNN feature learning of single-frame images, today's behavior recognition methods are generally based

on image 2D convolutional neural networks, which tend to ignore the information connections between consecutive frames. As a result, a lot of information is lost in the action-processing aspect inside the video [17]. Therefore, the full utilization of 3D convolutional networks has become one of the important directions for behavior recognition research. Based on 3D convolutional feature extraction on a 3D cube composed of consecutive video frames, 3D convolutional networks can capture the feature information of video images in both spatial and temporal dimensions, and more importantly, the speed of network operation, which benefits from the one-time processing of multiple frames, is largely improved. However, 3D convolutional networks are not very accurate, require too much hardware processor, and are relatively less cost-effective, so the dual-stream network idea is also utilized in the current state-of-the-art recognition methods using 3D convolutional networks, which are fully utilized for optical stream images from the point of view of performance enhancement for behavioral recognition methods, as shown in Figure 2.

Performing the modeling of temporal features in terms of human behavior in the video is achieved by making full use of the temporal correlation between adjacent frames, which uses recurrent neural networks as an important underlying condition. However, the recognition accuracy in this regard is still some distance away from the expected value. The keyframe is the frame where the key action in the movement or change of the object is located and is the most valuable information in the video that is represented most intuitively, representing the maximum complete representation of all the information contained in the video based on the use of as few video image frames as possible. This is the keyframe extraction technology focused on the research direction that is also the principle of keyframe extraction.

$$\gamma_t = f_{imp}(a(X, t), \beta^2(x_{t-1})), \quad (4)$$

$$a(X, t) = \frac{1}{\gamma_t} (\gamma_{t-1} a(X, t) - \gamma_{t+1} \beta^2(x_{t-1})). \quad (5)$$

The keyframe extraction method based on shot detection is the more classical keyframe extraction method; the main idea of the method is to extract the first and last frames or middle frames in each shot after segmentation as the keyframe of that shot after segmenting the video sequence into multiple shots. The method is extremely computationally small and simple to implement, but the number of extracted keyframes is determined. The method wants to use the extracted keyframes to represent the main content of the original shot, which is only applicable to the case where the content of the shot changes little so that the limitation is greater. The first and last frames of the middle frames of each shot do not represent the main content of the shot completely when the visual content has certain changes, and the extracted keyframes lose their meaning because they do not represent the whole video. The method also results in poorly extracted keyframes when the result of the shot boundary detection deviates from the actual shot, indicating that the keyframe extraction method relies too much on the shot boundary detection.

$$\gamma_t = \int_{\Omega} \{f\phi(I_0(x) + I_1(X - u(x) - \varphi(u, \Delta u))\} dx \quad (6)$$

Specifically, the segmented video recognition network framework proposed in this paper is aimed at illustrating how to maximize the use of visual information in the entire video sequence for video-level prediction. The main part of this recognition network consists of two parts: a spatial flow convolutional neural network and a temporal flow convolutional neural network. Instead of operating on individual frames or stacks of frames, the segmented video recognition network performs this operation on a series of short video segments sparsely sampled from the entire video; each short video segment in this sequence generates its preliminary prediction of the behavior category and uses the agreement between the short video segments as a video-level prediction for the entire video. During network training, this dual-stream network updates the model parameters by iteration.

$$L(y, G) = \sum_{i=1}^c y_i \left(G_i + \ln \sum_{j=1}^c \exp G_j \right), \quad (7)$$

$$\frac{\partial L(y, G)}{\partial W} = \frac{L}{G} \left(\sum_{i=1}^c y_i \left(G_i + \ln \sum_{j=1}^c \exp G_j \right) \right). \quad (8)$$

Various coding methods have been introduced in the development of video coding technology, and the methods of data compression based on statistical properties of images are known as predictive coding. This type of method uses the spatial and temporal correlation of the image signal to predict the image currently being encoded using the encoded image information and then encodes and transmits the difference between the predicted value and the true value. The method uses mathematical transformations such as

the discrete Fourier transform and discrete cosine transforms, to convert the image described in the spatial domain into coefficients in the transform domain to reduce the amount of data. The combination of different types of coding methods results in the so-called hybrid coding. This has slowly evolved into the current hybrid coding framework. Although there is a tendency for video standards to compete, they all adopt a hybrid coding framework. Background subtraction is one of the more widely used methods in current motion target detection technology. The basic idea is similar to that of interframe differencing, in that it uses the differential operation of different images to extract the target region. However, unlike interframe differencing, background subtraction does not subtract the current frame from the adjacent frames but rather subtracts the current frame from a continuously updated background model to extract the motion target in the differential image.

3.2. Experiments on Frame Signal Processing of Athletic Sports Video Images. Video coding has evolved to the point where encoders are required to perform real-time encoding of ultrahigh-definition videos such as 4 K and 8 K [18]. The maximum coding unit has been extended to a 256×256 size as the proportion of the area of flat image content increases in high-resolution video image content. The coding standards currently under investigation also support more flexible block division structures to achieve higher coding efficiency. The basic idea of intraframe prediction is to exploit the correlation of adjacent pixels, which in video coding refers to the reconstructed pixels of the encoded blocks around the current block, to effectively remove the video time-domain redundancy, because for an image, the luminance and chromaticity values between two adjacent pixels are often relatively close to each other, that is, the color changes gradually and does not change abruptly into a completely different color at once. Current intraframe prediction uses block-based multidirectional intraframe prediction. There are 33 angular prediction modes defined in H265, plus planar and DC modes, for a total of 35 in-frame prediction modes. The number of angle prediction modes has been increased to 65. In video coding, the angle of the predicted direction is not a geometric angle but is expressed in terms of the number of pixels.

According to the research needs, in the practice of scientific research work, the video image needs to be moved superimposed for comparison, but at present, this video image-processing function is difficult to carry out clear superimposition of moving background video images because the background is cluttered and disorderly, and this technology still needs to be further developed and applied [19, 20]. At the same time, because the accuracy and clarity of modern video image acquisition still need to be improved and developed, the video image is often enlarged after processing, and the image quality and clarity declined, so there is a need to do further development and improvement of the accuracy of the video image enlargement demonstration processing, as shown in Figure 3.

Visual understanding of human behavior is a core capability for building AI systems, and so recognition of behavior

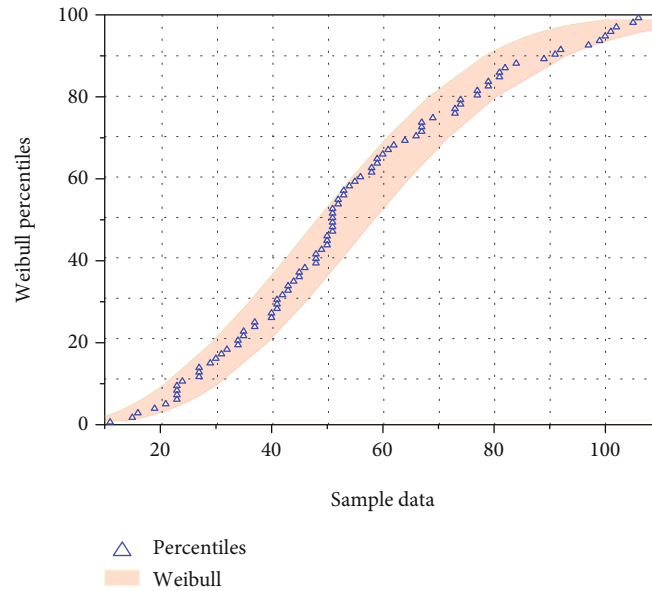


FIGURE 3: Behavior recognition-related dataset.

in the video is now the focus of most researchers, to classify the input video data containing specific actions. However, this setup does not have much practical application, as real-world videos are usually unedited and most actions will occur accompanied by other actions rather than alone. As a result, recent researchers have gradually turned their attention to the task of detecting temporal behavior in the unedited video, which goes beyond classification to find the boundaries of the behavior performed in unedited video data, i.e., the start and end moments. The development of temporal behavior detection has led to the emergence of many real-world applications, such as highlight moments in sports videos, or more advanced tasks such as the automatic generation of video subtitles.

Ideally, the selection unit can remove the interference of background noise. However, in practical applications, limited by the number of convolutional channels and the size of the training dataset, the features learned offline are insufficient to adapt to arbitrary tracking targets without online learning of target information. Therefore, this paper proposes an adaptive learning method to obtain more discriminative features online. Experiments show that online updating does not bring significant performance gains but rather aggravates the computational burden. The authors of this paper argue that fine-tuning the motion regression part is not reasonable because it mainly predicts motion by comparing two similar features, while the purpose of online updating is to make the changing targets have similar features and suppress the interference of background noise. In addition, it was found in the experiments that updating the parameters of the fully connected layer tends to lead to overfitting due to the large number of parameters contained in the fully connected layer. In the case of known blur kernels, the recovery of blurred images can be transformed into the problem of finding the deconvolution of blurred images

and blur kernels, using a nonblind deblurring algorithm such as the Richardson-Lucy algorithm (RL) algorithm to solve clear images in an iterative loop. The RL algorithm is one of the most widely used image-deblurring algorithms, and the method has been improved with continuous improvements and works better for most blurred images. The Poisson noise statistical properties are used to recover low-quality video frame images when the blur kernel is known, as shown in Figure 4.

In this paper, 100 videos with significant jitter, blurring, and poor video effects were downloaded from video websites to judge the video quality based on the image sharpness metric, and these videos were found to be lower than the sharpness metric on the experimental video library. Since the authenticity and reliability of the videos on the video sites were uncertain and could not be used as experimental videos, the videos from the SULFA and OV video libraries were subjected to video blurring using Adobe Premiere Pro CC software. The videos after blurring are close to the clarity metrics of the videos on these websites, so they can be used as low-quality video libraries for experiments [21–23].

The motion search approach is to develop a strategy for finding the best matching block from the starting point of the search. In the beginning, a global search algorithm was used to ensure the accuracy of the matching. The global search predefines a search region, compares the coding unit with all the candidate blocks in the reference frame region, and determines the best matching block according to the motion estimation criterion, and the displacement between these two blocks is estimated. But this method inevitably brings great computational complexity. Therefore, various fast search methods have been investigated. Although the process of these fast algorithms varies, they all mainly are aimed at improving search efficiency by trying to avoid locations that are unlikely to be the best matching blocks.

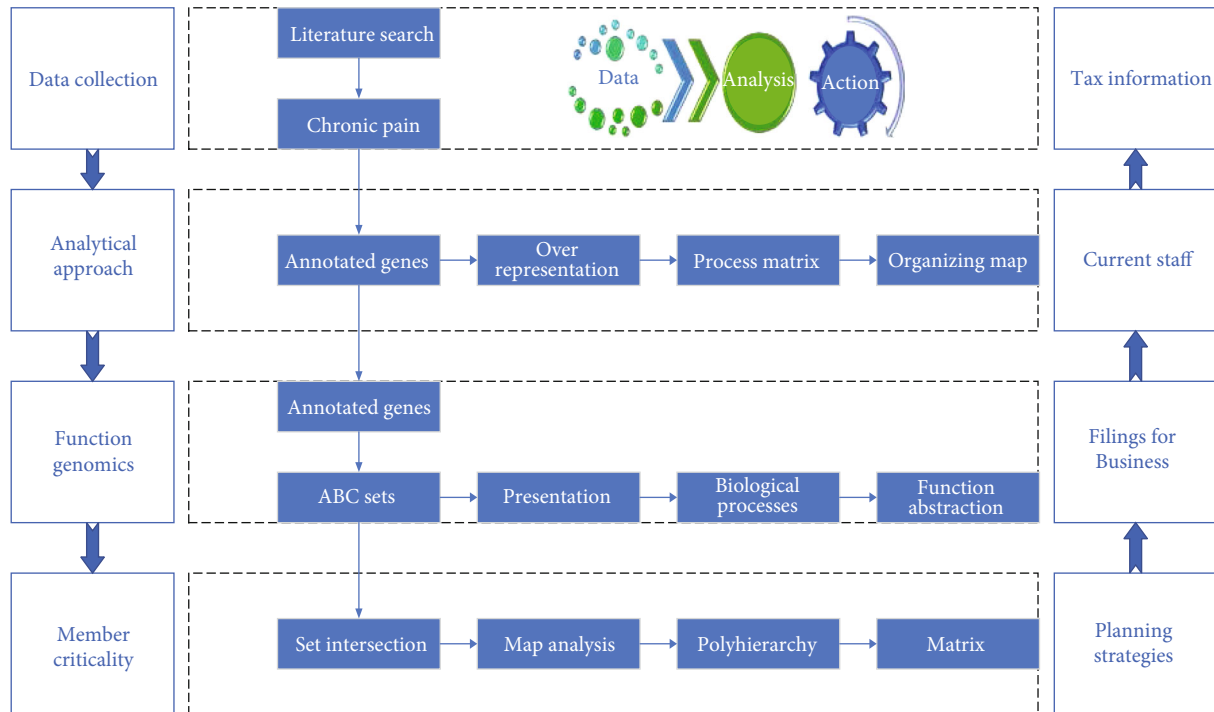


FIGURE 4: Experimental flow chart.

4. Analysis of Results

4.1. Performance Results of Big Data Graphics-Processing Algorithms. The chunked motion compensation encodes the motion vector of each block on a block-by-block basis. The block-to-block motion vectors show a strong null-domain correlation, and usually, the motion vector of one block shows the same characteristics as the motion vectors of the surrounding blocks. If only the motion information of one block is saved in detail, the information of the adjacent blocks only needs to refer to the information of that block, and the differences between them are recorded, and this reference relationship is saved at the same time to effectively reproduce the original information. To improve the accuracy of motion compensation, the reference frame pixel values are usually interpolated at the subpixel level, when the motion vector is also noninteger. However, there are drawbacks in this way, and the most common one is the block effect, as the chunked motion compensation divides the whole into multiple subblocks, forcibly cutting off the overall continuity, resulting in the boundary parts of each subblock going in different directions due to their different coding methods, breaking the boundary continuity and presenting a square content overall. When the block effect is serious, the decoded image will look like a mosaic effect, which seriously affects the visual quality. Another disadvantage is that when the high-frequency component is large, it will cause a ringing effect. A reasonable secondary coding scheme was designed, reasonable initial quantization parameters were set in the single-trip coding process, and the reference relations of coding blocks as well as bit consumption were counted. A video-tampering detection method with fused

audio is proposed for video interframe tampering that is often accompanied by simultaneous audio tampering. First, the ENF signal of the audio is extracted and analyzed to determine the suspicious tampering point of digital audio according to the phase continuity and consistency. Then synchronously locate the location of the suspicious anomalies of video frames, while calculating the video frame similarity using GMSD coarse detection to extract the suspicious anomalies.

Quantization coding is a means of further compressing the data after transform coding and predictive coding, mainly by mapping a signal interval to a signal to reduce the amount of information to be recorded. Also, predictive coding and transform coding did not bring distortion to the image, and quantization as a lossy compression technique is the main source of distortion generated by video coding. To reduce the distortion caused by quantization while maintaining better video quality, reasonable quantization methods need to be designed. The quantization methods are mainly uniform quantization, nonuniform quantization, and adaptive quantization. Among them, uniform quantization is a linear quantization, which is simple and easy to implement, but the quantization effect is not good because the distribution of the quantized objects is not considered. Take the residual signal as an example; most of the residuals are concentrated around 0. At this time, nonuniform quantization can be used to quantify the finer regions with concentrated distribution and the coarse-grained quantization for the more loosely distributed regions. This can get a better quantization effect. This method combines quantization with rate-distortion optimization to choose the least costly quantization value among

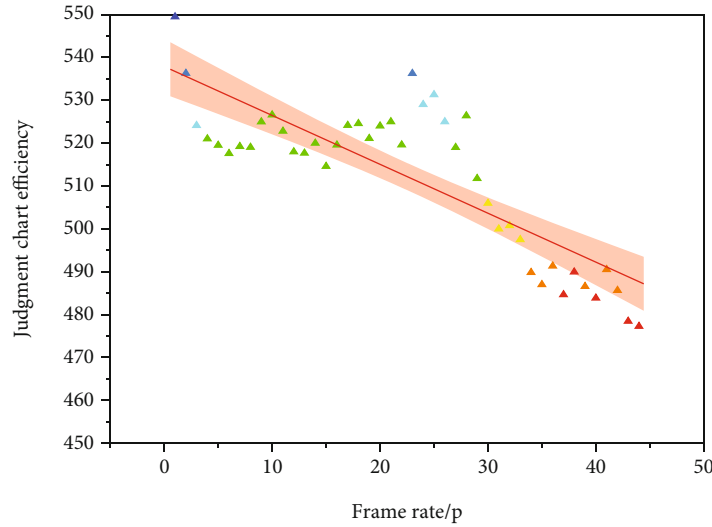


FIGURE 5: Algorithm performance.

multiple selectable quantization values by rate-distortion optimization, as shown in Figure 5.

Among the selected test sequences, there are also differences in the performance gains of different sequences, which indicates a certain tendency of the algorithm in this paper. Analysis of the sequence image content reveals that the performance gain achieved is greater in sequences with characteristics such as fixed scenes and flat object motion. This is because some of the image contents in the sequence exist for a long time and are referenced by subsequent images for a long time, and the performance gain of the high reference value coding unit is accumulated in the image propagation chain, so the performance gain is more obvious. In contrast, in sequences where the background changes and the object motion are drastic, the image content with reference value disappears when the propagation process is in progress, thus bringing lower performance gain, as shown in Figure 5.

From the data in Figure 6, the algorithm can achieve a more significant performance gain, with gains of 3.59%, 2.69%, and 2.99% on the Y , U , and V components in the GOP length 8 configurations. This indicates that the coding quality of the coding unit can be reasonably improved according to its reference value, which can effectively improve the quality of the reference coding unit and thus achieve overall quality improvement. Moreover, the performance gains achieved with the configuration of GOP length 16 are more significant, with the average performance gains of 7.21%, 12.17%, and 12.06% on the three components of Y , U , and V . This indicates that the larger the GOP length, the more significant the hierarchical structure, the stronger the reference dependence between frames and between coding blocks, and therefore, the more performance is brought by the algorithm that corrects the Lagrange multipliers by reference values.

4.2. Experimental Results of Video Graphics Processing for Track and Field Sports. The channel injects bit resources into the decoding buffer at a constant rate, previously referred to

as the constant code rate. To accomplish a good code rate control process, the following requirements are imposed on it. Since the first frame of the encoding is usually an I-frame and requires more resources to be allocated, if the allocation is done at the beginning, the resources will be exhausted quickly, so a moderate number of resources will be injected into the buffer now, which is also known as delayed encoding. The buffer size should also be set reasonably and should not be too large or too small. The allocated resources for each image frame are reasonably interfered with to facilitate continuous system operation.

The kinematic parameters such as velocity, acceleration, displacement, angular and angular velocity, rotation, and rotation of different technical movements are obtained through certain calculations and analysis, and then the skeletal muscle biomechanics reproduction software system is used to simulate and stimulate the technical movements of the human body under the domination of skeletal muscle in real situations to achieve the purpose of studying and simulating the skeletal muscle biomechanical movements, as shown in Figure 6.

This requires firstly a large number of 3D video tests and analyses of sports technical movements to establish a database of skeletal muscle-dominated technical movement forms, then the corresponding technical movement analyses to reproduce video images of skeletal muscle-dominated sports technical movements, and also the inverse derivation of human skeletal muscle or joint force characteristics from the kinematic parameters of different kinematic parameters of skeletal muscle involved in the movement kinetic tests and studies for a more scientific and rational study and analysis of sports technical movements. Figure 7 shows the simulation and reproduction of the skeletal muscle dynamics of the volleyball snapping technical action and the human walking action, as well as the derivation and analysis curves of the inverse dynamics of the action of the sole on the ground.

Using athletes' individual sports technology panorama, you can combine two athletes' technology panorama,

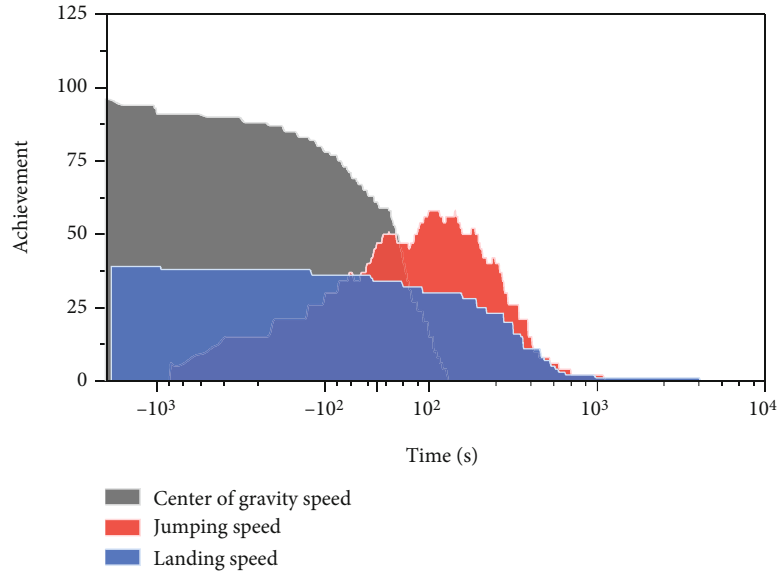


FIGURE 6: Movement time and movement speed.

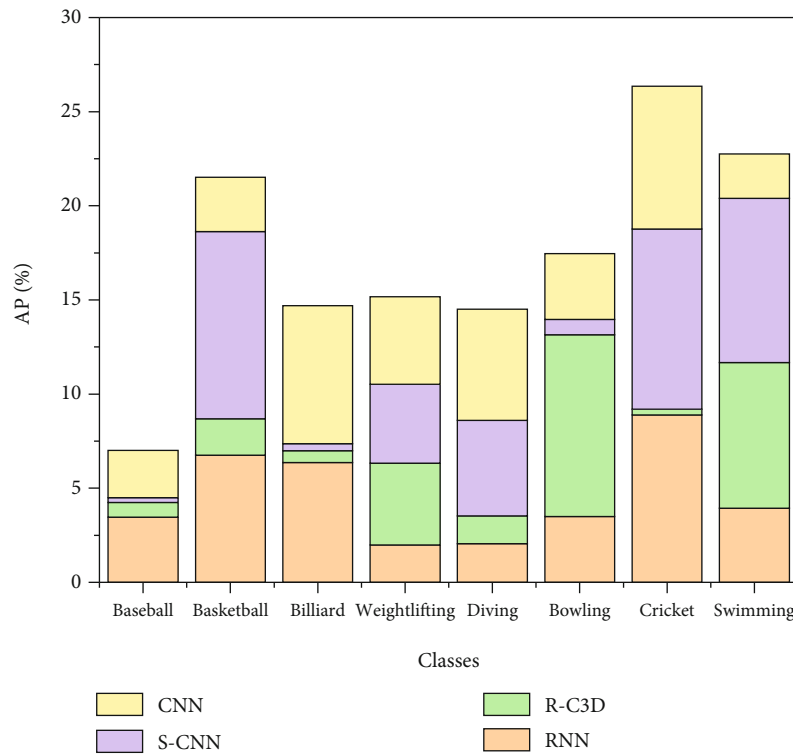


FIGURE 7: Average accuracy of motion categories.

synthesize a technology picture using image-processing software, and align the key images of the athletes' same time phase for the position, forming a panorama technology comparison picture including two athletes. It is convenient for coaches and athletes to make targeted observations and comparative analyses to find the technical gaps in different key positions, which can be studied and imitated intuitively to have the most intuitive

improvement of individual technical movements. For such pictures, more advanced image-processing software can be used to fit the background, which can be compared more clearly in the same background, and such background removal and fitting of video image-processing software have now appeared, but the processing is slightly cumbersome and requires more simple and effective processing, as shown in Figure 7.

A new stable and effective hierarchical feature network was proposed to accurately detect behaviors and their temporal coordinates in uncut videos. The network is divided into two parts: a behavior classification network and a coordinate regression network. The behavior classification network is mainly used for video behavior determination, predicting the motion score for each frame and generating initial proposals based on the score distribution. The coordinate regression network, on the other hand, uses coarse granularity to divide the initial proposals into cell levels, achieves fast computation by recycling cell-level features, and uses temporal coordinate regression to refine the boundaries of the proposed regions eventually to stable and accurate motion boundaries.

5. Conclusion

Starting from the working principle of code rate control, several code rate control methods and their respective applicable scenarios are introduced. The advantages of quadratic coding are analyzed, and a code rate control algorithm based on quadratic coding is proposed. With the help of decoding buffers, the code rate allocation errors that occur in the real-time coding process are analyzed. The work of code rate control is converted to eliminate the code rate allocation error. A reasonable secondary coding scheme is designed, reasonable initial quantization parameters in the single-trip coding process are set, and the reference relations of coding blocks as well as bit consumption are counted. A video-tampering detection method with fused audio is proposed for video interframe tampering that is often accompanied by simultaneous audio tampering. First, the ENF signal of the audio is extracted and analyzed to determine the suspicious tampering point of digital audio according to the phase continuity and consistency. Then synchronously locate the location of the suspicious anomalies of video frames, while calculating the video frame similarity using GMSD coarse detection to extract the suspicious anomalies. Therefore, for the offline training process, this paper borrows from the data augmentation methods in other fields of computer vision and proposes to generate many complex background images with semantic information using a limited training set by linearly overlaying different images, which enhances the discriminative ability of different motions. The online multiscale block-filter network reorders the improved output and outputs an online confidence score for each similar object. The online multiscale block-filter network utilizes online negative sample mining for online learning, which mitigates tracking drift caused by similar interfering objects. The effectiveness of the proposed algorithm is verified by conducting extensive comparison experiments on authoritative datasets in the tracking domain.

Data Availability

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

Conflicts of Interest

No competing interests exist concerning this study.

References

- [1] H. Ba, "Medical sports rehabilitation deep learning system of sports injury based on MRI image analysis," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 5, pp. 1091–1097, 2020.
- [2] Y. Zhang, M. Zhang, Y. Cui, and D. Zhang, "Detection and tracking of human track and field motion targets based on deep learning," *Multimedia Tools and Applications*, vol. 79, no. 13–14, pp. 9543–9563, 2020.
- [3] E. M. Saoudi and S. Jai-Andalousi, "A distributed content-based video retrieval system for large datasets," *Journal of Big Data*, vol. 8, no. 1, pp. 1–26, 2021.
- [4] C. Cuevas, D. Quilon, and N. García, "Techniques and applications for soccer video analysis: a survey," *Multimedia Tools and Applications*, vol. 79, no. 39, pp. 29685–29721, 2020.
- [5] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, A. Bouridane, and A. Beghdadi, "A combined multiple action recognition and summarization for surveillance video sequences," *Applied Intelligence*, vol. 51, no. 2, pp. 690–712, 2021.
- [6] A. A. Khan, J. Shao, W. Ali, and S. Tumrani, "Content-aware summarization of broadcast sports videos: an audio-visual feature extraction approach," *Neural Processing Letters*, vol. 52, no. 3, pp. 1945–1968, 2020.
- [7] F. A. Khan, M. Nawaz, M. Imran, A. U. Rahman, and F. Qayum, "Foreground detection using motion histogram threshold algorithm in high-resolution large datasets," *Multimedia Systems*, vol. 27, no. 4, pp. 667–678, 2021.
- [8] B. Li and X. Xu, "Application of artificial intelligence in basketball sport," *Journal of Education, Health and Sport*, vol. 11, no. 7, pp. 54–67, 2021.
- [9] K. Rangasamy, M. A. As'ari, N. A. Rahmad, N. F. Ghazali, and S. Ismail, "Deep learning in sport video analysis: a review," *Telkomnika*, vol. 18, no. 4, pp. 1926–1933, 2020.
- [10] J. W. Yang, "Target tracking and recognition of a moving video image based on convolution feature selection," *International Journal of Biometrics*, vol. 13, no. 2–3, pp. 180–194, 2021.
- [11] S. Agrawal and P. Natu, "Segmentation of moving objects using numerous background subtraction methods for surveillance applications," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 3, pp. 2553–2563, 2020.
- [12] A. D. Smith, "Event detection in educational records: an application of big data approaches," *International Journal of Business and Systems Research*, vol. 15, no. 3, pp. 271–291, 2021.
- [13] C. Guntuboina, A. Porwal, P. Jain, and H. Shingrakhia, "Deep learning based automated sports video summarization using YOLO," *Electronic Letters on Computer Vision and Image Analysis*, vol. 20, no. 1, pp. 99–116, 2021.
- [14] S. Asadianfam, M. Shamsi, and A. R. Kenari, "TVD-MRDL: traffic violation detection system using MapReduce-based deep learning for large-scale data," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2489–2516, 2021.
- [15] S. Banerjee, H. H. Chopp, J. G. Serra, H. T. Yang, O. Cossairt, and A. K. Katsaggelos, "An adaptive video acquisition scheme for object tracking and its performance optimization," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17227–17243, 2021.

- [16] T. Dekel and N. Snavely, "Unveiling unexpected training data in internet video," *Communications of the ACM*, vol. 64, no. 8, pp. 69–79, 2021.
- [17] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.
- [18] L. X. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of "bring your own device" policy," *Journal of Organizational and End User Computing (JOEUC)*, vol. 31, no. 2, pp. 65–82, 2019.
- [19] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, "Vision-based human activity recognition: a survey," *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 30509–30555, 2020.
- [20] T. Li, J. Sun, and L. Wang, "An intelligent optimization method of motion management system based on BP neural network," *Neural Computing and Applications*, vol. 33, no. 2, pp. 707–722, 2021.
- [21] Y. Hu, "Image segmentation based on velocity feature vector for moving target extraction," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11983–11991, 2020.
- [22] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification: the consensus, the best practice and the grey areas," *Journal of Organizational and End User Computing (JOEUC)*, vol. 31, no. 1, pp. 39–60, 2019.
- [23] L. Sun, Y. Li, College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China, and University Medical Center Groningen, University of Groningen, Groningen 9713 GZ, Netherlands, "Multi-target pig tracking algorithm based on joint probability data association and particle filter," *International Journal of Agricultural and Biological Engineering*, vol. 14, no. 3, pp. 199–207, 2021.

Research Article

Image Real-Time Detection Using LSE-Yolo Neural Network in Artificial Intelligence-Based Internet of Things for Smart Cities and Smart Homes

Zheng Zhi-Xian¹ and Fuquan Zhang^{2,3} 

¹Fujian Chuanzheng Communications College, Fuzhou Fujian, 350007, China

²College of Computer and Control Engineering, Minjiang University, Fuzhou, China 350108

³Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Chengdu, China 610021

Correspondence should be addressed to Fuquan Zhang; zfq@mju.edu.cn

Received 6 January 2022; Accepted 22 February 2022; Published 9 March 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 Zheng Zhi-Xian and Fuquan 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.

In this paper, a novel visual image real-time detection LSE-Yolo neural network is presented, which is in artificial intelligence-based Internet of Things for smart cities and smart homes. Despite the great achievements that have been acquired in image detection, the issue of visual image real-time detection combined with privacy data protection to serve for smart cities and smart homes has been overlooked. The technique we applied in our study is referred to as visual object detection, which can contribute to more healthy and comfortable life. When several studies have been carried out to test the validity, it is suggested that our proposed LSE-Yolo neural network has better performance in image real-time detection based on AIoT for smart cities and smart homes. And it is similar to state-of-the-art. The fruitful work has made great contributions to our present understanding of the visual image detection serving for smart cities and smart homes.

1. Introduction

Recently, with the improvement of residents' living standards and consumption power, the traditional life has been changed to the life with science and technology, which can make it more healthy, fast, convenient, and comfortable. To meet the needs of modern smart cities and homes, visual system in computer vision applied in smart cities and smart homes has become a hot topic in artificial intelligence and Internet of Things (AIoT). Therefore, image detection in AIoT has attracted much attention from the academia, which can serve for smart cities and smart homes. There have been several studies highlighting object detection for smart cities and smart homes [1–8] in recent years. And object detection plays an important role in the visual system application in AI-based for smart cities and homes. The common object detection algorithms can be classified into two categories: one is traditional detectors and the other is deep learning-based detectors. One of the typical traditional

detectors is HOG detector, which can be regarded as an important improvement over scale-invariant feature transform and shape context, and it was firstly proposed by Dalal et al. in 2005. And DPM proposed by Felzenszwalb et al. [9] in 2008 was the culmination of traditional target detection methods. In deep learning-based detectors, there are two major classes: one-stage detectors [10–12] and two-stage detectors [13, 14], respectively. In one-stage detectors, features from the network to predict object classification and location directly. While in two-stage detectors, a proposal is required; that is, a preselection box containing the objects to be detected, and then, fine-grained object detection is carried out. In recent years, it has become a trend that object detection algorithms served for smart cities and homes. In literature, Khan et al. [3] in 2017 presented the detection of people through computer vision in the Internet of Things scenarios to improve the security in smart cities, smart towns, and smart homes. In 2019, Garcia et al. [1] proposed to perform object detection mechanism based

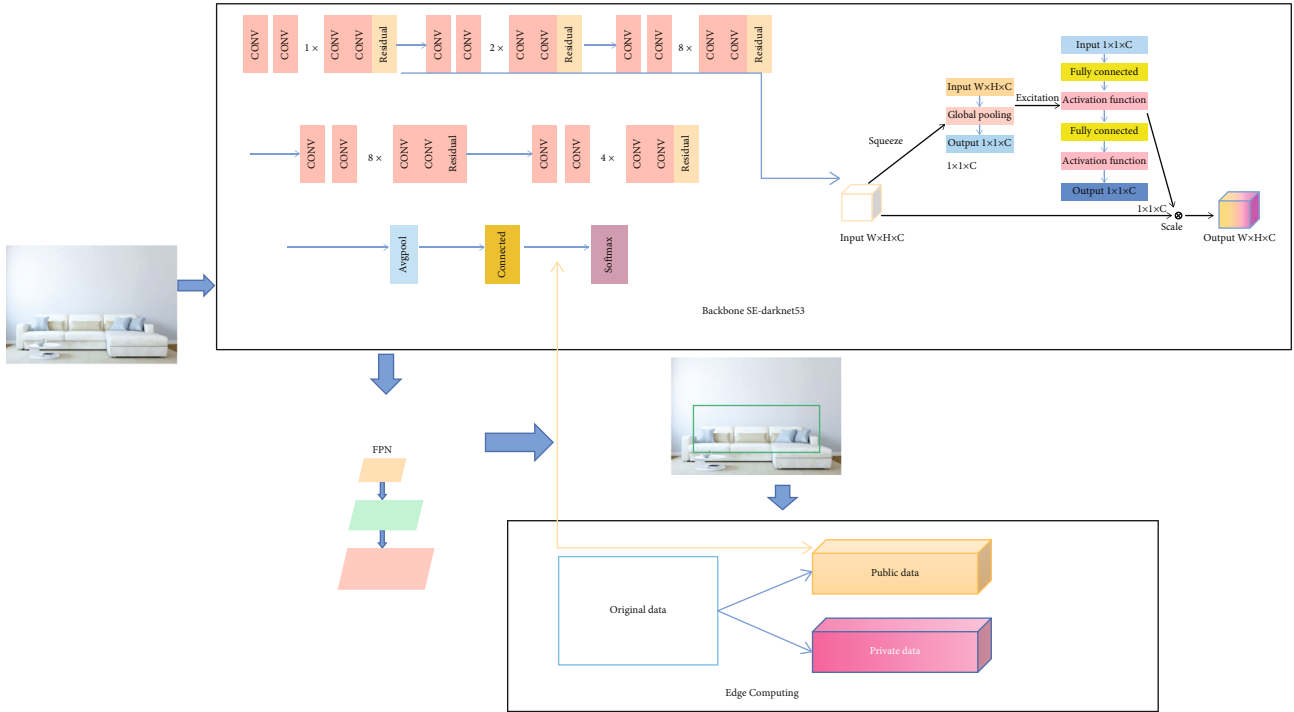


FIGURE 1: The architecture of our LSE-Yolo visual image real-time detection for smart cities and smart home.

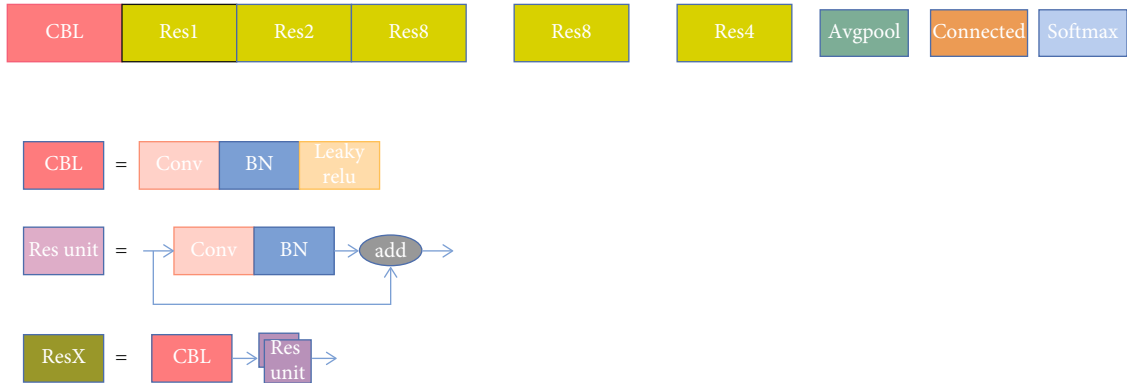


FIGURE 2: The architecture of darknet53.

on deep learning algorithm using embedded IoT devices for smart home appliance control in CoT. Mehmood et al. [2] in 2021 proposed deep learning-based hybrid approach for the development of an IoT-based intelligent home security and appliance control system in the smart cities. Park J et al. [4] in 2018 presented to use deep neural networks for activity recognition with multisensor data in a smart home. Meanwhile, Hu and Ni [7] proposed IoT-driven automated object detection algorithm for urban surveillance systems in smart cities. In 2019, Nayak et al. [5] proposed to use deep learning video-based real-time intrusion detection system for smart city applications. Then in 2020, Wang et al. [6] presented the algorithm for target detection in smart city combined with depth learning and feature extraction. Xu et al. in 2021 [8] presented feature-enhanced occlusion perception object detection for smart cities. Wang et al. [15] in

2019 proposed an end-to-end three-dimensional (3D) object detection method based on sparse convolution neural network and feature fusion for autonomous driving in smart cities. In 2020, Mettupally and Menon [16] presented a smart ecosystem for parking detection using deep learning and big data analytics for smart city. Great progress has been made in the object detection algorithm applied in smart cities and homes; some limitations are the following: (1) lack of sufficient data; it is hard to obtain a sound object detection model served for smart cities and smart homes. (2) Due to the limitation in common deep neural networks, it is difficult to be deployed in the mobile device and real-time detection carried out to contribute to smart cities and smart homes. (3) Neglecting the privacy data protection of the users, it is tough to realize the smart city- and smart home-based AIoT. In order to overcome the problems and limitations stated

TABLE 1: The structure of darknet53.

	Type	Filters	Size	Output
	Convolutional	32	3×3	512×512
	Convolutional	64	$3 \times 3/2$	128×128
	Convolutional	32	1×1	
1×	Convolutional	64	3×3	
	Residual			128×128
	Convolutional	128	$3 \times 3/2$	64×64
	Convolutional	64	1×1	
2×	Convolutional	128	3×3	
	Residual			64×64
	Convolutional	256	$3 \times 3/2$	32×32
	Convolutional	128	1×1	
8×	Convolutional	256	3×3	
	Residual			32×32
	Convolutional	512	$3 \times 3/2$	16×16
	Convolutional	256	1×1	
8×	Convolutional	512	3×3	
	Residual			16×16
	Convolutional	1024	$3 \times 3/2$	8×8
	Convolutional	512	1×1	
4×	Convolutional	1024	3×3	
	Residual			8×8
	Avgpool		Global	
	Connected		1000	
	Softmax			

above, in our work, a new neural network for visual image system-based AIoT is presented to serve for smart cities and smart homes to perform real-time detection. To solve the problem of the insufficient data, supervised data augmentation is adopted in our network including single-sample data augmentation and multiple data augmentation, in which the data can be enlarged on the basis of existing data. And more lightweight neural network is proposed in our study, which can be installed easily and no dependencies. So it can be deployed on the mobile device to carry out real-time detection-based AIoT for the smart cities and smart homes. In addition, to protect the privacy data, edge computing data are collected and calculated locally or on edge nodes without being uploaded to the cloud, and important and sensitive information does not need to be transmitted through the network, thus effectively avoiding the problem of privacy leakage. Therefore, the proposed LSE-Yolo neural network can realize the smart city- and smart home-based AIoT. The contributions in our study are the following:

- (1) A new neural network for visual image systems-based AIoT is presented to serve for smart cities and smart homes to perform real-time detection

- (2) Supervised data augmentation is adopted in our network to address the problem of the insufficient data
- (3) To protect the privacy data, edge computing is used to classify the data into public data and private data to avoid the privacy leakage

This paper is divided into 4 sections as follows: Section 2 introduces the materials and methods, that is, the architecture of our LSE-Yolo neural network. The results and discussion of our work are shown in Section 3. Finally, Section 4 summed up this paper.

2. Materials and Methods

Figure 1 shows the architecture of the novel LSE-Yolo neural network for visual image systems to perform detection-based AIoT to serve for smart cities and smart homes. In the image input, it would be conveyed to the model to perform the operation. The input image is the sofa in a home which is sampled by the mobile device randomly. Firstly, it would go through the backbone network, that is, the darknet53, which contains 53 convolutional layers to extract the features and the information. And the Squeeze-and-Excitation module is added after each residual layer to perform the Squeeze part and Excitation operation, which can solve the loss problem caused by different importance of different channels of feature map in the process of convolutional pooling. Then in the FPN module, it would be carried out by upsampling by high-level features and top-down linking by low-level features, and predictions are made for each layer. Finally, the output of the image which be detected and taken as the sofa will proceed to the database to perform edge computing to be classified the public data or the private data for the smart cities and smart homes. And the classified public data can be transferred to the model to improve the performance of the model, while important and sensitive private data does not need to be transmitted through the network, thus effectively avoiding the problem of privacy leakage. Therefore, it is seen that our LSE-Yolo visual image system for smart cities and smart home is considerable, which can make people's life more convenient and comfortable. To sum up, the proposed neural network can perform real-time detection along with privacy data protection and sufficient data.

2.1. Network Architecture. In our LSE-Yolo neural network, the darknet53 is used as our backbone network as shown in Figure 2, which is the basic unit of our network. There are several advantages of darknet53. Firstly, it can be installed easily and just take a few minutes to finish installing. Secondly, it has no dependencies and does not rely on any libraries. Thirdly, its structure is clear, and the source code of the backbone network is easy to view and modify. Finally, it has python interface and is easy to transplant. The architecture of the darknet53 is shown in Table 1, which is mainly composed of a series of 1×1 and 3×3 convolution layers. It is seen that 53 convolution layers are contained in the darknet53, where the fully connected layer is regraded as convolution layer but the residual layer is not taken as

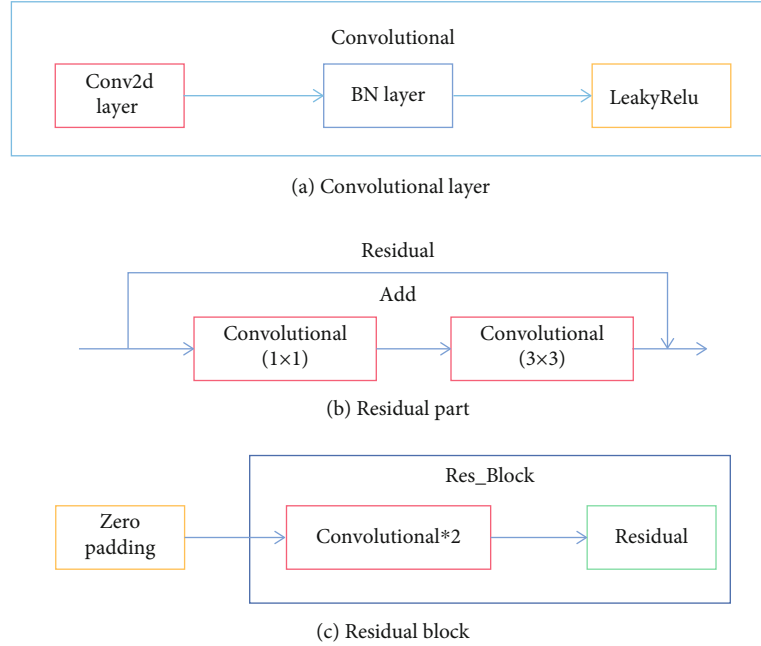


FIGURE 3: The structure of residual block.

convolution layer. It is found that there is a residual layer after two convolution carried out, in which the original information is retained and the extracted features are integrated by adding residuals.

As shown in Figure 3(a), each convolutional layer is followed by a normalization layer and an activation layer. In Figure 3(b), it is demonstrated that in the residual part, $1 \times 1 + 3 \times 3$ is used to deepen the network depth to enhance feature sampling. Therefore, in Figure 3(c), it can be regarded as residual block when two convolutional module and a residual module are included. Moreover, each residual block is preceded by a zero padding.

If we train a normal network using a standard optimization algorithm with no residuals, no shortcuts or jump connections, say gradient descent, or some other popular optimization algorithm, it is found that with the deepening of network depth, training errors will first decrease and then increase. In theory, as the network deepens, the performance of the network should be more sound. While in fact, for a normal network with no residual modules, as the network deepens, there are more and more training errors. Therefore, darknet53 has benefited from the residual module; it can help solve the problem of gradient vanishing and gradient explosion, allowing to train deeper networks while maintaining good performance. In our network, to address the loss problem, the Squeeze-and-Excitation module is added after each residual layer. Also, the FPN module is added to fuse the high-level features and low-level features. Meanwhile, supervised data augmentation is adopted in our network to address the problem of the insufficient data and edge computing to protect the privacy data.

2.2. Attention Mechanism SE Module. Visual attention mechanism is a special brain signal processing mechanism of human vision, which can quickly scan the global image

to obtain the target area that needs to be focused on and greatly improves the efficiency and accuracy of visual information processing. And the attentional mechanism in deep learning is essentially similar to the human-selective visual attention mechanism. Its goal is also to select from the information that is more critical to the current task and goal. In our work, the Squeeze-and-Excitation (SE) module is adopted in our LSE-YOLO neural network, to solve the loss problem caused by different importance of different channels of feature map in the process of convolutional pooling. In Figure 4, the SE module is explained, where Squeeze part and Excitation part are contained. In the module, W and H represent the width and height of the feature map and C represents the number of channels. Therefore, the size of the input feature map is $W \times H \times C$. In the Squeeze part, the squeeze operation can be regarded as a global average pooling, where the feature map can be compressed into $1 \times 1 \times C$ vector. Then in the Excitation part, it consisted of two fully connected layers to perform Excitation operation, where SERatio is a scaling parameter that is intended to reduce the number of channels and thus the computation effort. In the first fully connected layer, there are $C * \text{SERatio}$ of neurons, where the input is $1 \times 1 \times C$ and the output is $1 \times 1 \times C * \text{SERatio}$. In the second fully connected layer, the input is $1 \times 1 \times C * \text{SERatio}$, and the output is $1 \times 1 \times C$. Finally, the scale operation was carried out to multiply the channel weights. The output of the parameters number and the calculated quantity is $2 * C * C * \text{SERatio}$.

2.3. Neck Network FPN. In our work, the feature pyramid network (FPN) is added in the Neck part of LSE-Yolo to solve the multiscale problem in object detection, in which the performance of small object detection is greatly improved without increasing the calculation amount of the original model through simple network connection changes.

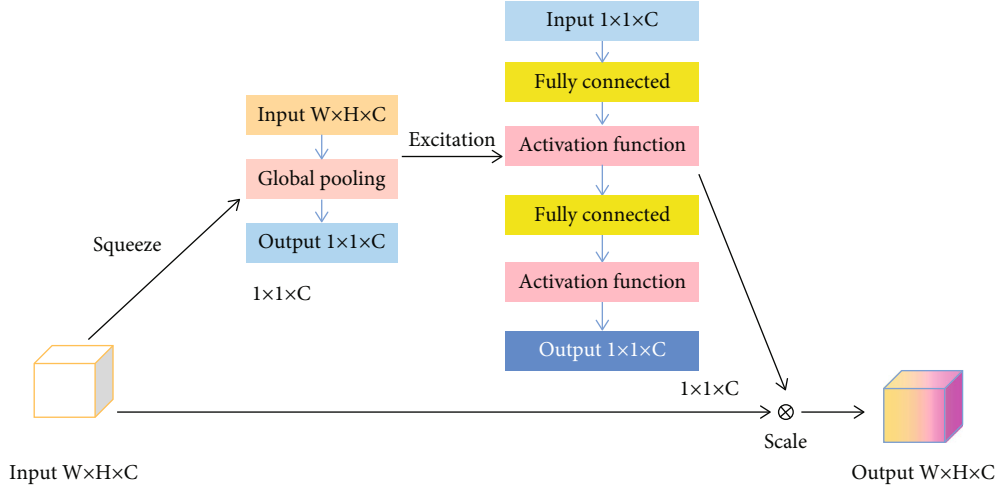


FIGURE 4: The architecture of SE module.

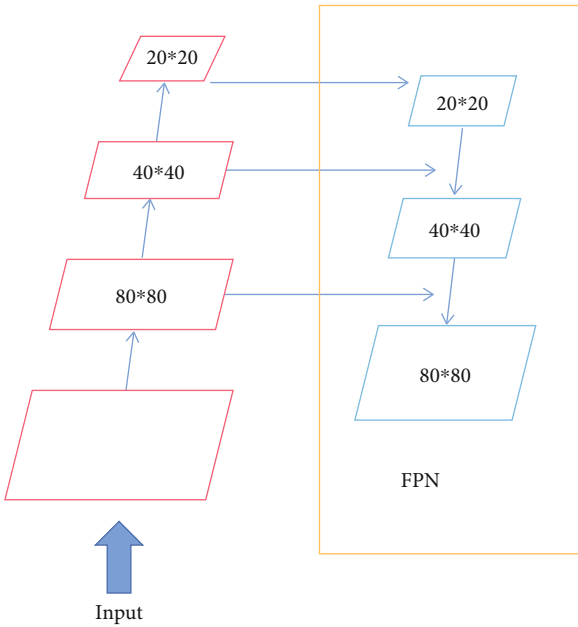


FIGURE 5: The architecture of FPN.

In Figure 5, the structure of FPN is demonstrated, upsampling by high-level features and top-down linking by low-level features, and predictions are made for each layer. In the bottom-up process, it is the common forward propagation process of neural network, and the feature map is usually smaller after the convolution kernel calculation. When in the top-down process, it is to carry out upsampling of more abstract and semantic high-level feature maps. And in horizontal connection, the result of upsampling is merged with the feature map of the same size generated from the bottom-up. The features of the two layers connected horizontally are the same in spatial dimensions to underlie position details. In addition, the horizontal connection can reduce the number of feature maps. In a word, the FPN added in the Neck part of our network can utilize both high resolution of low-level features and high semantic informa-

tion of high-level features. And the prediction results can be achieved and performed separately at each fused feature layer through combining the features of these different layers.

2.4. Loss Function. In our work, the loss function in prediction box used is IOU loss and GIoU loss when BCE_loss is used in object loss and class loss.

As shown in Formula (1), it is explained that the IoU function and B^{gt} represent the target box when B is the prediction box. It can be used in regression tasks of bounding box due to the characteristic of nonnegative scale invariance, identity, symmetry, and triangle inequality.

$$L_{IoU} = 1 - \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}. \quad (1)$$

The GIoU loss function is shown in Formula (2), where C represents the minimum bounding box for B and B^{gt} . It aims to solve the problem that the value of IoU loss is unified as 1 when there is no overlapping area between the detection box and the ground truth.

$$L_{GIoU} = 1 - IoU + \frac{|C - B \cap B^{gt}|}{|C|}. \quad (2)$$

When the calculation of mutual information is taken into account, BCE_loss which is suitable for classification is used in object loss and class loss as shown in Formula (3), which represents the loss corresponding to the N sample. ω is the super parameter and y is the actual label.

$$\text{Loss}_{Xiyi} = -\omega[yi \log xi + (1 - yi) \log (1 - xi)]. \quad (3)$$

2.5. Data Augmentation. In our work, data augmentation is used to reduce the overfitting of the network, and the training samples of MS COCO dataset are shown in Figure 6. By transforming the training images, a network with stronger generalization ability can be obtained, which can better

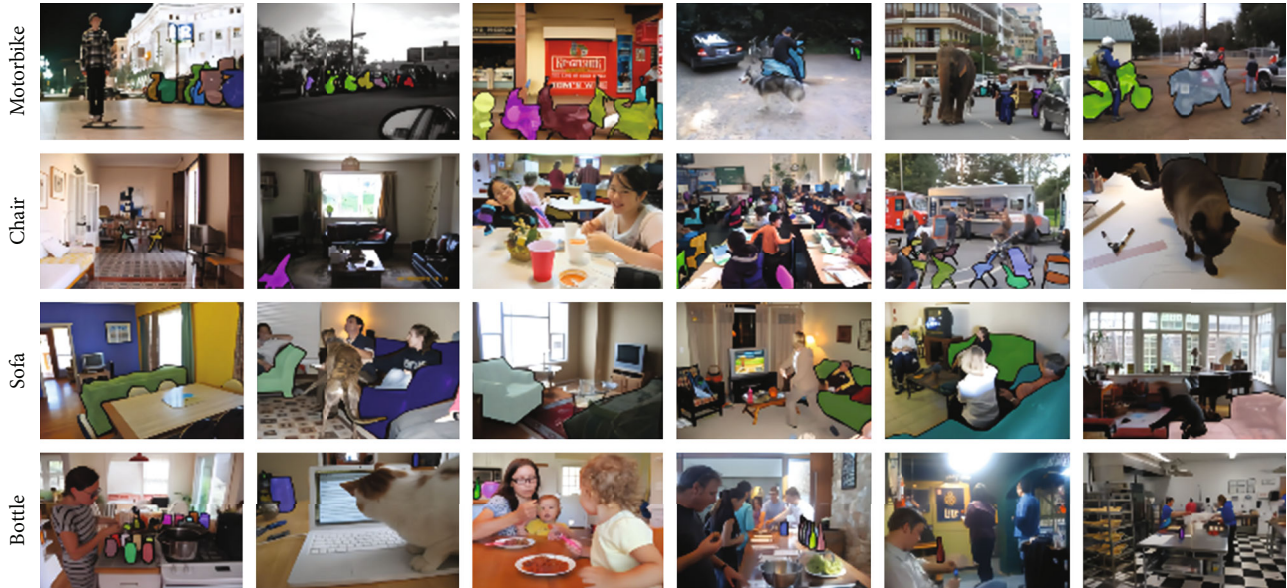


FIGURE 6: Samples of MS COCO dataset.

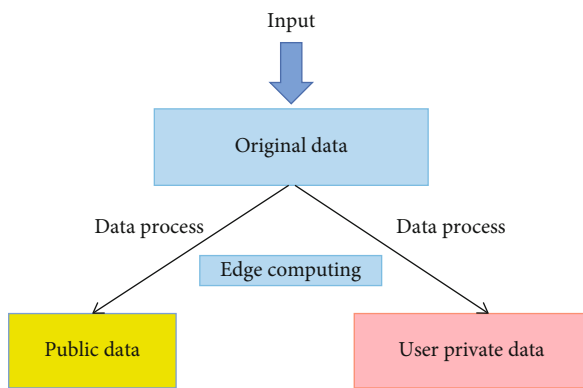


FIGURE 7: The architecture of data privacy protection.

TABLE 2: The comparison of AP and parameters (state-of-the-art) of different models for MSCOCO.

Methods	AP	Parameters
Yolov3+NMS	41.7	67.27 M
Yolov3 baseline	38.5	63.00 M
Yolov4 baseline	43	31 M
Yolov5 baseline	44.5	21.4 M
Ours	45.2	26.3 M

adapt to the application scenarios. It can be divided into supervised data augmentation and unsupervised data augmentation methods. In supervised data augmentation, it can be fall into single sample data augmentation and multiple data augmentation methods, while unsupervised data augmentation can be divided into two directions: generating new data and learning augmentation strategies. In our network, supervised data augmentation is adopted including

single sample data augmentation and multiple data augmentation, in which the data can be enlarged on the basis of existing data. In single sample data augmentation, that is, when enhancing a sample, all operations are carried out around the sample itself, including geometric transformation classes and color transformation classes. Also, there are several common approaches to transform the image geometrically including flipping, rotation, clipping, deformation, scaling, and other operations. In addition, Mosaic and Mixup are also used as multiple data augmentation in our network, which contribute to the effect of the small object detection, the robustness, and the stability of the model.

2.6. Data Privacy Protection. In our work, the edge computing is added to protect the data privacy for smart cities and smart homes in Figure 7. It can be taken as an operation performed by using the edge strip near the data source, and the efficiency can be improved, namely, proximity computing. The cloud computing model that relies on a single centralized processing mode for the construction of smart cities or smart homes cannot cope with all the problems, and it needs the integration of multiple computing modes to solve the problems. There are several advantages of edge computing compared with other traditional cloud computing. Firstly, it can protect the localized data. The data in the Internet of Things is so vital to the users' lives that uploading it to a cloud computing center increases the risk of exposing users' private data. As shown in Figure 7, edge computing data are collected and calculated locally or on edge nodes without being uploaded to the cloud, and important and sensitive information does not need to be transmitted through the network, thus effectively avoiding the problem of privacy leakage. Secondly, it can reduce the cloud data to transfer and energy consumption. Thirdly, it can realize real-time computing.

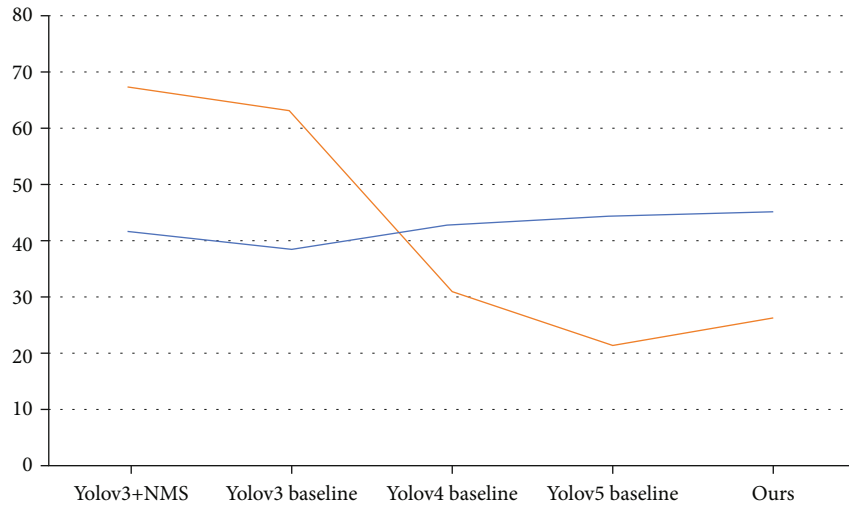


FIGURE 8: The line chart of the comparison of AP and parameters of different models.

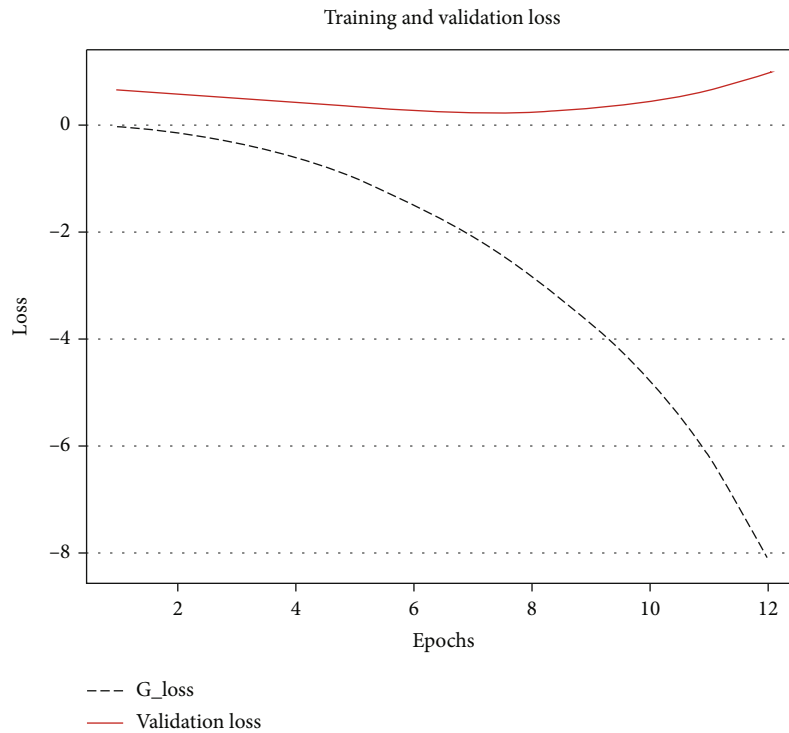


FIGURE 9: The training and validation loss.

3. Results and Discussion

In Table 2, the results obtained from the studies of the proposed neural network and other common neural network are shown. The experiment is based on Tesla p100, and the dataset is MSCOCO dataset. It is found that our proposed LSE-Yolo neural network in artificial intelligence-based Internet of Things for smart cities and smart homes has better performance than other common neural networks. In our work, the AP is 45.2, which is higher compared to the neural network, and the parameters is 26.3 M, which is lower

than Yolov5. And the line chart can be shown in Figure 8. In Figure 9, the training and the validation loss are demonstrated. In a word, the comprehensive of the proposed LSE-Yolo is highly consistent with the prediction of the theoretical model, which is suitable for smart cities and smart homes. Despite the great advantages mentioned above, there are still some problems which have not been addressed in our proposed models. It counts that the tradeoff between the accuracy and the speed when applied in smart cities and homes is still a challenge. Therefore, it deserved to be further studied to solve the above problems.

4. Conclusions

In conclusion, it is stated that our proposed LSE-Yolo neural network in artificial intelligence-based internet of things for smart cities and smart homes is fruitful. It can not only bring us more healthy and comfortable life but also make great contributions to meeting the needs of the modern smart cities and smart homes. However, there are still some limitations in our study, which is how to realize the tradeoff between the accuracy and the speed of the model. In addition, it should be noted that the data for smart cities and smart homes is still a challenge. And 3D object deep detection method which should be developed and combined with augmented reality for smart cities and smart homes is still the problem to be solved. Therefore, the limitations above remain to be solved in the future study, which aims to make it more convenient and fast for the smart cities and smart homes.

Data Availability

The data we used is available and can be accessed to perform image detection system using neural network in artificial intelligence-based Internet of Things for smart cities and smart homes study. And part of them are available to you from the corresponding author upon request (zfq@mju.edu.cn).

Conflicts of Interest

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

Acknowledgments

Our work is supported by the 2021 Fujian Middle-aged and Young Teacher Education Research Project (Science and Technology), Project Nos. JAT210719 and JAT210704; supported by Digital Media Art, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music, Project No. 21DMAKL01; supported by the First Batch of Industry-University Cooperation Collaborative Education Project funded by the Ministry of Education of the People's Republic of China, 2021, Project No. 202101071001; supported by Minjiang College 2021 School-Level Scientific Research Project Funding, Project No. MYK21011.

References

- [1] C. G. Garcia, D. Meana-Llorian, B. G-Bustelo, J. M. C. Lovelle, and N. Garcia-Fernandez, "Midgar: Detection of people through computer vision in the Internet of Things scenarios to improve the security in smart cities, smart towns, and smart homes," *Future Generation Computer Systems*, vol. 76, pp. 301–313, 2017.
- [2] F. Mehmood, I. Ullah, S. Ahmad, and D. H. Kim, "Object detection mechanism based on deep learning algorithm using embedded IoT devices for smart home appliances control in CoT," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2019.
- [3] S. Khan, S. Nazir, and H. U. Khan, "Smart object detection and home appliances control system in smart Cities," *Computers, Materials and Continua*, vol. 67, no. 1, pp. 895–915, 2021.
- [4] J. Park, K. Jang, and S. B. Yang, "Deep neural networks for activity recognition with multi-sensor data in a smart home," in *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, pp. 155–160, Singapore, 2018.
- [5] R. Nayak, M. M. Behera, U. C. Pati, and S. K. Das, "Video-based real-time intrusion detection system using deep-learning for smart city applications," in *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pp. 1–6, Goa, India, 2019.
- [6] F. Wang, Z. Xu, Z. Qiu, W. Ni, J. Li, and Y. L. Luo, "Algorithm for target detection in smart city combined with depth learning and feature extraction," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8885670, 7 pages, 2020.
- [7] L. Hu and Q. Ni, "IoT-driven automated object detection algorithm for urban surveillance systems in smart cities," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 747–754, 2018.
- [8] J. Xu, H. Wang, M. Xu, F. Yang, Y. Zhou, and X. Yang, "Feature-enhanced occlusion perception object detection for smart cities," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5544194, 14 pages, 2021.
- [9] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *2008 IEEE conference on computer vision and pattern recognition*, pp. 1–8, Anchorage, AK, USA, 2008.
- [10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, USA, 2016.
- [11] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, USA, 2017.
- [12] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," <https://arxiv.org/abs/1804.02767>.
- [13] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, Chile, 2015.
- [14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection 404 with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 29, no. 6, 2015.
- [15] L. Wang, X. Fan, J. Chen, J. Cheng, J. Tan, and X. Ma, "3D object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities," *Sustainable Cities and Society*, vol. 54, article 102002, 2020.
- [16] S. Mettupally and V. Menon, "A smart eco-system for parking detection using deep learning and big data analytics," in *2019 SoutheastCon*, pp. 1–4, Huntsville, AL, USA, 2019.

Research Article

MidSiot: A Multistage Intrusion Detection System for Internet of Things

Nguyen Dat-Think ^{1,2}, Ho Xuan-Ninh ^{1,2} and Le Kim-Hung ^{1,2}

¹Faculty of Computer Networks and Communications, University of Information Technology, Ho Chi Minh City 70000, Vietnam

²Vietnam National University, Ho Chi Minh City 70000, Vietnam

Correspondence should be addressed to Le Kim-Hung; hunglk@uit.edu.vn

Received 20 October 2021; Revised 4 January 2022; Accepted 17 January 2022; Published 21 February 2022

Academic Editor: Hamed Nassar

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

Internet of Things (IoT) has been thriving in recent years, playing an important role in a multitude of various domains, including industry 4.0, smart transportation, home automation, and healthcare. As a result, a massive number of IoT devices are deployed to collect data from our surrounding environment and transfer these data to other systems over the Internet. This may lead to cybersecurity threats, such as denial of service attacks, brute-force attacks, and unauthorized accesses. Unfortunately, many IoT devices lack solid security mechanisms and hardware security supports because of their limitations in computational capability. In addition, the heterogeneity of devices in IoT networks causes nontrivial challenges in detecting security threats. In this article, we present a collaborative intrusion detection system (IDS), namely, MidSiot, deployed at both Internet gateways and IoT local gateways. Our proposed IDS consists of three stages: (1) classifying the type of each IoT device in the IoT network; (2) differentiating between benign and malicious network traffic; and (3) identifying the type of attacks targeting IoT devices. The last two stages are handled by the Internet gateways, whereas the first stage is on the local gateway to leverage the computational resources from edge devices. The evaluation results on three popular IDS datasets (IoTID20, CIC-IDS-2017, and BOT-IoT) indicate our proposal could detect seven common cyberattacks targeting IoT devices with an average accuracy of 99.68% and outperforms state-of-the-art IDSs. This demonstrates that MidSiot could be an effective and practical IDS to protect IoT networks.

1. Introduction

The number of devices connecting to the Internet has been growing at a breathtaking pace over the past decades. From two billion in 2006, it reached 200 billion in 2020 because of the proliferation of mobile computing and the Internet of Things (IoT) [1]. As a result, these devices play a critical role in primary industries (e.g., healthcare, manufacturing, retailing, security, and transportation) by providing intelligent services, such as tracking inventory, managing machines, monitoring patient health, and detecting abnormality. They are anticipated to boost the total global worth of IoT to 6.2 trillion dollars by 2025, most of which come from manufacturing (2.3 trillion dollars) and healthcare (2.5 trillion dollars) [2]. It is apparent that IoT is the driving force of evolution in every daily aspect.

However, ensuring security and privacy for the IoT devices is a nontrivial challenge due to their limitation in computational capability, which is insufficient for traditional security mechanisms. This makes them susceptible to wide-ranging cyberattacks, such as data leakage, spoofing, and DoS/DDoS. In a report published by Kaspersky, the first half of 2021 witnessed 1.5 billion attacks against smart devices aiming at stealing data, mining cryptocurrency, or building botnets [3]. In September 2016, an infamous attack performed by Mirai malware turned 380,000 devices into botnets that launched DDoS attacks against several services and organizations, including Dyn-a domain registration service provider [4]. Moreover, this malware is capable of mutating [5]. On 12 December 2017, its variant exploited a zero-day flaw in Huawei HG532 routers to speed up its infection. One year later, the number of variants was

increased significantly, such as Okiru, Masuta, OMG, Wicked, Hakai, and Yowai [6].

To eliminate such security threats, an intrusion detection system (IDS) is commonly deployed at network gateways. It constantly monitors network traffics coming from various sources to detect abnormalities, which may be security threats. Following [7], the attack detection approaches of the IDS are categorized into signature-based and anomaly-based. The former approach identifies cyberattacks by comparing a set of signatures (or rules) extracted from known attacks with incoming traffic. We note that the key difference between IoT network traffic and other network traffics is the diversity and volume. The diversity of IoT network traffic comes from the heterogeneity of IoT devices and their communication protocols, resulting in diverse network behaviors. Furthermore, IoT devices exponentially increase and generate massive data traversing the Internet. Due to these characteristics, the rule-based detection mechanism, SNORT is a typical example, is insufficient for IoT networks. In detail, SNORT is ineffective for complex attacks signs of which are various and implicit in network traffic. In addition, SNORT needs to maintain a large rule-set and security experts in the loop to update these rules frequently. Therefore, it is not efficient and scalable enough for IoT scenarios. The latter approach, which is the most popular, makes use of machine learning (ML) to construct a model of normal network traffic patterns. This model is then used to measure the similarity between the incoming traffic and known patterns to detect malicious traffic. Although the anomaly-based approach using machine learning considerably alleviates the weaknesses of the signature-based approach, it still has several limitations.

- (i) Neglecting a collaborative edge-cloud architecture: Training and inferring tasks in machine learning are resource-intensive, so they are usually handled by cloud platforms in existing work. This might decrease the detection performance because the network traffic at the cloud level, which is merged from several gateways consisting of various data sources, is intricate. In contrast, edge devices are resource-constraint IoT devices that are insufficient to handle complex machine learning tasks (e.g., detecting abnormal activities, training detection models). Offloading these tasks on edge devices severely affects other services running on these devices. However, we believe that these devices could handle specific tasks to increase the IDS's detection performance regardless of their limitation in computation capability. Therefore, a collaborative edge-cloud architecture for intrusion detection systems is necessary to overcome this limitation.
- (ii) Lacking IoT device-type identification: Because the IoT device types are various and heterogeneous, their network behaviors are highly diverse. For example, the high UDP packet rate coming from IoT cameras is normal, but the one from temperature sensors is a sign of a security threat. This may lead to false attack detection. Thus, identifying

device types and considering them as an input feature of the attack detection model is crucial to increase the detection accuracy of IDSs.

- (iii) Detecting a limited set of attack types: Existing works about IDS are extensive but primarily concerned with detecting a limited set of attacks in a general domain, such as DoS and spoofing. Given the rising prevalence of IoT, there is an essential need to address a larger set of attacks targeting IoT networks.

To solve the above limitation, we present MidSiot, a machine learning-based three-stage IDS designed for IoT networks supporting collaboration between local gateways and Internet gateways of Internet Service Provider (ISP) (solving the first limitation). In more detail, to leverage edge computing and enhance the attack detection accuracy, the first stage is operated at local gateways to identify IoT devices based on their behaviors in the network (solving the first limitation), whereas the next two stages powered by a machine learning model are handled by the Internet gateways to not only differentiate between normal and malicious network traffic, but also accurately identify the seven common attack types (solving the first limitation). The evaluation results on existing IDS datasets for the IoT domain (IoTID20, BOT-IoT, and CIC-IDS-2017) show that MidSiot could detect seven popular attacks targeting IoT devices with an average accuracy of 99.68%. Our main contributions presented in this study are as follows:

- (1) A collaborative architecture for IoT IDSs to leverage the computational resources of edge gateway to enhance IDS's detection performance.
- (2) A lightweight and robust machine learning-based IDS constituting of three stages to accurately detect various cyberattacks pointing at IoT devices.
- (3) We intensively evaluate our proposal on popular IDS datasets and examine the resampling techniques to address imbalanced datasets during our experiments.

The remainder of the article is organized as follows. In Section 2, we discuss related work. The MidSiot architecture and its detection method are presented in Section 3. Section 4 reports the evaluation of our method through IDS datasets, and we conclude our work in Section 5.

2. Related Works

In recent years, there has been an increased interest in exploring machine learning for enhancing the detection quality of IDSs [17, 18]. In [9], the authors proposed an anomaly detection mechanism using a single machine learning classifier. The authors of [10] presented a scalable k-NN-based online anomaly detection addressing the lazy-learning problem in wireless sensor networks [10]. The works in [16] also employed anomaly detection techniques for IDSs using binary classification. This means that they cannot identify the type of attack. In [11], the authors proposed an ensemble of autoencoders for online IDS whose

performance is comparable to offline anomaly detectors. Ref. [12] is a novel approach for IDSs in which the authors applied convolutional neural network to predict the attack types. 98% accuracy on the NSL-KDD dataset was achieved in this experiment. There are also hybrid-IDSs where anomaly-based and signature-based approaches are used to develop the IDS. Such a typical system is introduced in [19], in which packet header anomaly detection, network traffic anomaly detection, and SNORT are combined. In [13], the authors leveraged four machine learning algorithms to derive rule-sets used as signatures for their IDS. In [8], a hierarchical architecture including multiple neural networks was proposed to detect malicious packets and identify the attack types hidden inside these packets. The authors in [14] introduced a hierarchical structure for IDS that separates the detection process into different steps. The authors in [20] suggested a distributed architecture for smart home IDSs that offload complex tasks onto the Internet Service Provider (ISP) and deliver simple ones to the smart home gateway [20].

In terms of datasets used for IDSs, the authors in [21] proposed a new dataset called IoTID20, which was also evaluated in their work by implementing several machine learning algorithms (e.g., logistic regression, decision tree, random forest), which results in increasing F1-score for both binary classification and multiclass classification. The authors in [15] developed a new realistic botnet dataset for use in IoT networks, and as a result, it mainly consists of DDoS attacks. Another dataset is CICIDS-2017 including attack traffic generated by their testbed and realistic background traffic created by the B-profile system [22]. All of these 3 datasets were constructed using the CICFlowMeter tool (formerly known as ISCXFlowMeter), thereby having a similar set of features. Regarding resampling methods for network intrusion detection system (NIDS), the literature at [23] compared multiple undersampling techniques for NIDS on CICIDS-2017 and CICIDS-2018 datasets, including random, cluster centroids, and nearmiss algorithms. The authors concluded that these undersampling methods reduced models' training time, and K-nearest neighbor has the most significant improvement. There are also other works that implemented a combination of several resampling techniques, such as oversampling and undersampling [24]. Their experimental results showed that the oversampling method increases the training time, whereas the undersampling method decreases this time. In addition, if the dataset is highly imbalanced, these methods improve the recall score notably. The authors in [25] proposed an algorithm-level class balancing technique that addresses the underlying issue about attack class imbalance in IDS datasets, resulting in identifying various attack categories with better accuracy than the CNN models.

In recent years, many researchers have geared toward blockchain applications in intrusion detection systems thanks to its potential in protecting data integrity and privacy. The authors in [26] investigated the challenges and limitations of blockchain to intrusion detection in addition to their applications, such as the overhead traffic with limited handling capability of intrusion detection and extensive

energy and cost usage of blockchain. Despite these difficulties, blockchain still has the potential to mitigate the data sharing and trust management issues in collaborative intrusion detection. As far as collaborative intrusion detection systems are concerned, a series of research [27–29] provided blockchain challenge-based collaborative intrusion detections. In these systems, the authors leveraged the strength of blockchain to investigate the trust mechanism in a network of IDS nodes. Their goals are to enhance the robustness of trust management against attacks as well as to protect the alarm aggregation process from malicious inputs. The works in [30, 31] made some contributions in the same direction, but the authors specifically targeted intrusion detection systems in a software-defined network. The authors in [32] proposed a deep blockchain framework to offer security-based distributed intrusion detection and privacy-based blockchain with smart contracts in IoT networks. Although the experimental results of the intrusion detection system were optimistic, the classification algorithm in use was a bidirectional long short-term memory, which accompanying blockchain might aggregate more computational burden on operating IoT devices.

To summarize the related works, Table 1 presents the state-of-the-art intrusion detection systems and their characteristics, including targeting security threats, attack detection method, evaluation datasets, attack-type and device-type detection, and lightweight. We can see that none of them is lightweight enough to classify the type of attack and its target. In addition, several approaches are evaluated by non-IoT datasets or testbeds having a small number of IoT devices. Thus, previous IDS proposals are insufficient for deploying to practical IoT ecosystem.

3. The MidSiot IDS

3.1. System Overview. In this section, we explain how our proposal works. First of all, Figure 1 illustrates the architecture of the proposed IDS that comprises three stages distributed between local network infrastructures and ISPs. The first stage is operated at the local gateways to identify connected IoT devices through their network behaviors. The next stage, which is conducted at the Internet gateways of ISPs, classifies network traffic of such IoT devices as normality or abnormality. When abnormal traffic is detected, it is transferred to the third stage to identify the attack types. Since the last two stages are done on ISPs which aggregate a huge volume of network traffic, correctly identifying the IoT device types along with their network traffic at the first stage is essential in increasing attack detection performance at following stages, especially for large-scale attacks targeting at multiple networks.

Second, Figures 2 and 3 present the block diagram of main operational phases in MidSiot, including the training and prediction phases, respectively. They also illustrate the connection and interfaces of components of the proposed IDS. In more detail, as shown in Figure 2, the raw network packets are captured by the packet flow inspection component from network traffic and transformed into network flows. These flows are then fed into the feature extraction component, extracting network features and computing network flow

TABLE 1: Summary of current works on Intrusion Detection Systems for Internet of Things.

Work	Security threat	Detection method	Validation dataset	Attack-type detection	Device-type detection	Lightweight
Zhang et al. [8]	DoS, R2L, U2R, and PROBE	Deep learning	KDD Cup 1999 Data	Yes	No	—
Wang and Stolfo [9]	58 attack types with 1999 DARPA dataset CUCS dataset (Code Red II, Buffer overflow)	1-gram models	1999 DARPA IDS Dataset CUCS Dataset	Yes	No	—
Xie et al. [10]	—	Machine learning	Real WSN data sets	—	No	Yes
Mirsky et al. [11]	Recon., MITM, DoS, Botnet	Autoencoder	Real-testbed	Yes	No	Yes
Ince [12]	DoS, probe, R2L, U2R	Deep learning	NSL-KDD	Yes	No	—
Kumar et al. [13]	Dos, exploit, probe, generic	Hybrid	UNSW-NB15	Yes	No	—
Anthi et al. [14]	Attack reconnaissances, DoS attacks, man-in-the-middle attacks, replay attacks, DNS spoofing	Machine learning	Real-testbed	Yes	Yes	—
Koroniotis et al. [15]	DoS/DDoS attacks, keylogging, data theft	Deep learning	BOT-IoT Dataset	Yes	No	—
Liu et al. [16]	Vulnerability scanners, ARP spoofing, DoS attacks, Mirai Botnet	Machine learning	IOTID-20 Dataset	No	No	Yes
Proposed System	Scanning methods (Host Discovery, Port scanning, OS/Version Detection) ARP Spoofing, SYN Flooding, Host Discovery, Telnet Bruce-force, UDP/ACK/HTTP Flooding	Machine learning	IOTID-20, CICIDS-2017, BOT-IoT Dataset	Yes	Yes	Yes

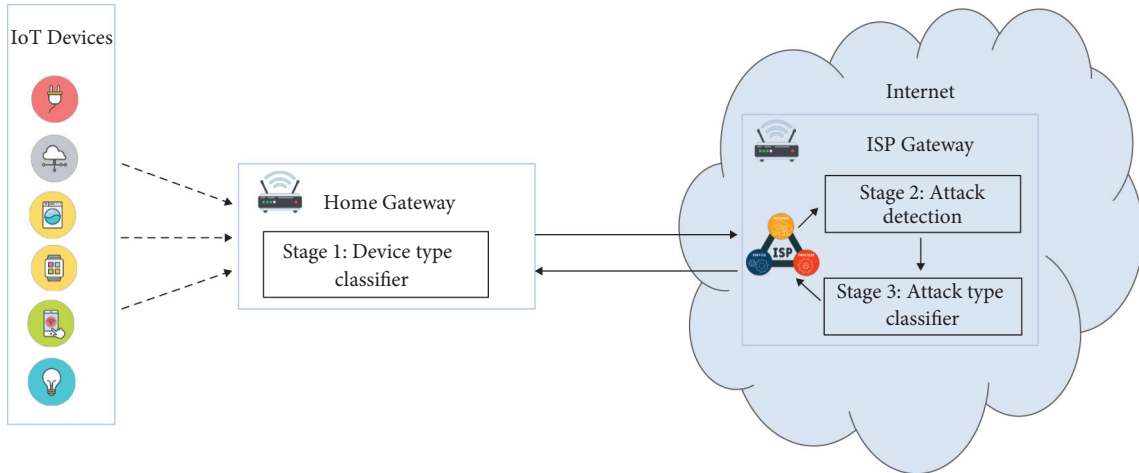


FIGURE 1: Overview of the architecture of the multistage intrusion detection system.

statistics. In addition, feature selection algorithm is applied to filter inappropriate features from the output features. In the training phase, these features are aggregated into a dataset used to train the models. Once models are trained successfully, the ISPs store these models used for the second and third stages in their local storages, while the models used for the first stage are sent to the local gateways. We note that the second stage employs several models, and each model is responsible for classifying network traffic for a specific device type.

In the prediction phase illustrated in Figure 3, the network features are constructed similarly with the training phase; however, they are then fed to models for detecting

malicious traffic. In more detail, the model of the first stage running on the local gateway is loaded to identify the device type of such features. All this information is transferred to ISP's Internet gateways, where a well-trained model corresponding to the device type is used to detect abnormality in these features. If malicious activities are detected, they are forwarded to the third stage to detect attack types by using a universal attack detection model. The detection results, including the IoT device under the attack and the type of attack, are sent to the action manager component to trigger necessary actions (logging attack behaviors, blocking the network traffic of victims, notifying administrators about the attack). Note that, because MidSiot's

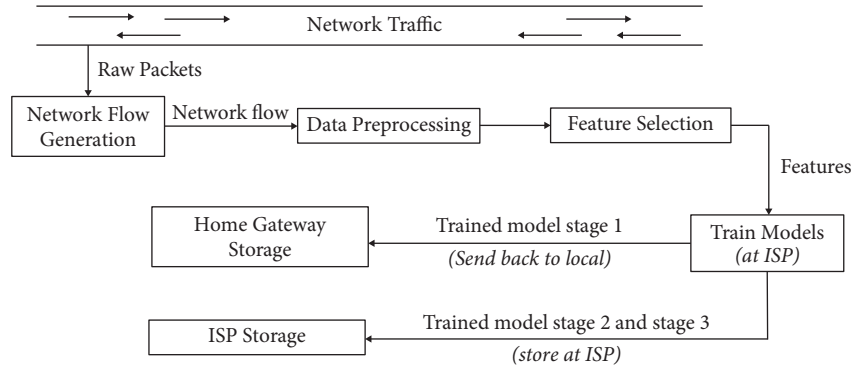


FIGURE 2: MidSiot’s training phase.

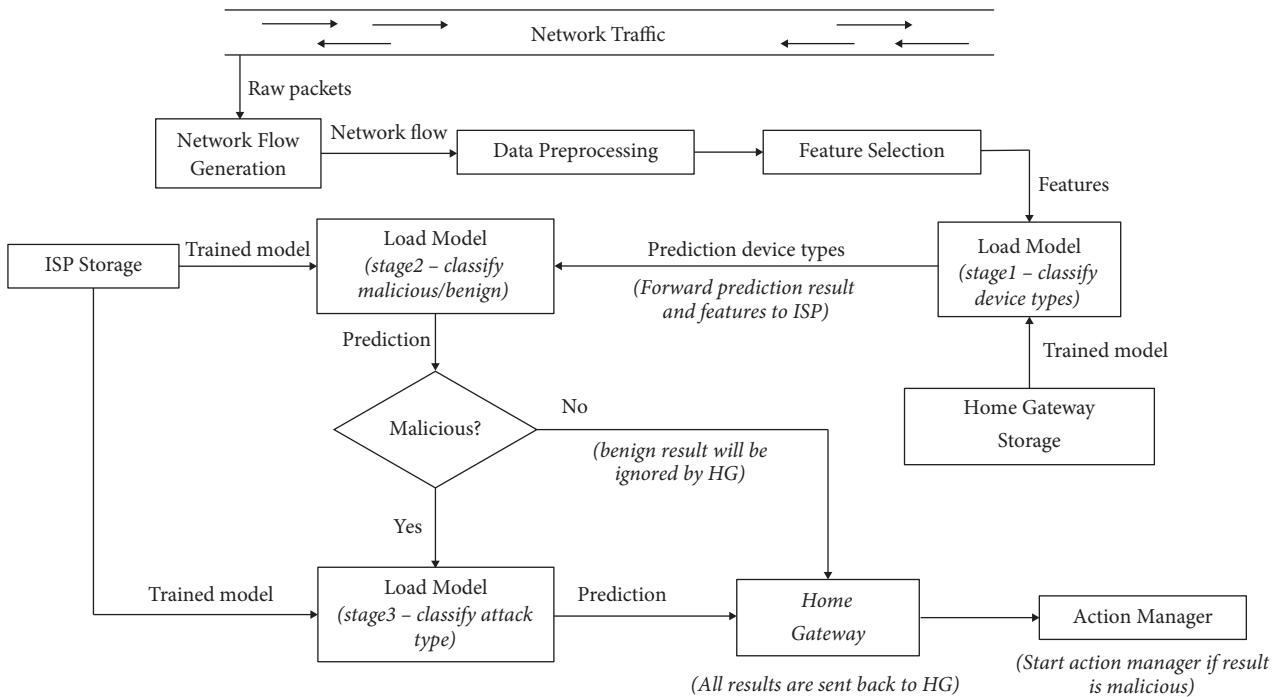


FIGURE 3: MidSiot’s prediction phase.

structure uses linked stages, the errors of one stage might affect not only the following stages but also the overall system’s performance. For example, if MidSiot misclassifies the device type, the second-stage results are potentially false. This is because the second-stage model is trained to learn the network patterns associated with a specific device type, and these patterns are different for each device type. Similarly, if the second-stage model misclassifies normal network traffic as abnormal, the final stage result is incorrect and triggers a false alert.

3.2. *Network Flow Generation.* Network flow generator is used to generate network flows from a batch of raw network packets. In MidSiot, it is powered by the deep packet inspection method that aggregates packets into flows sharing source/destination IP, source/destination port, and protocol and calculates flow features and statistics. In addition, this

method supports extracting MAC addresses, making it possible to label devices. Therefore, we could obtain 83 network features (e.g., FlowID, SourceIP, DestinationIP, SourcePort, DestinationPort, TimeStamp, and Protocols) listed in Appendix VI (Table 2).

3.3. *Data Preprocessing.* In IDS datasets, not all features are suitable for machine learning algorithms; some of them may degrade the model training performance, whereas others make models overfit. Therefore, employing a feature selection algorithm is necessary. First, all identity-based features (e.g., *ip_src*, *ip_dst*, *flow_id*, *timestamp*) are dropped to prevent the overfit issues, even features related to MAC addresses after labeling connected devices. We then adopt Pearson’s correlation coefficient to identify and remove unimportant features. In more detail, the importance index of each feature is its linear correlation coefficient value

TABLE 2: Extract network features.

Feature name	Description
fl_dur	Flow duration
tot_fw_pk	Total packets in the forward direction
tot_bw_pk	Total packets in the backward direction
tot_l_fw_pkt	Total size of the packet in the forward direction
fw_pkt_l_max	Maximum size of the packet in the forward direction
fw_pkt_l_min	Minimum size of the packet in the forward direction
fw_pkt_l_avg	Average size of the packet in the forward direction
fw_pkt_l_std	Standard deviation size of the packet in the forward direction
Bw_pkt_l_max	Maximum size of the packet in the backward direction
Bw_pkt_l_min	Minimum size of the packet in the backward direction
Bw_pkt_l_avg	Mean size of the packet in the backward direction
Bw_pkt_l_std	Standard deviation size of the packet in the backward direction
fl_byt_s	Flow byte rate that is the number of packets transferred per second
fl_pkt_s	Flow packets rate that is the number of packets transferred per second
fl_iat_avg	Average time between two flows
fl_iat_std	Standard deviation time two flows
fl_iat_max	Maximum time between two flows
fl_iat_min	Minimum time between two flows
fw_iat_tot	Total time between two packets sent in the forward direction
fw_iat_avg	Mean time between two packets sent in the forward direction
fw_iat_std	Standard deviation time between two packets sent in the forward direction
fw_iat_max	Maximum time between two packets sent in the forward direction
fw_iat_min	Minimum time between two packets sent in the forward direction
bw_iat_tot	Total time between two packets sent in the backward direction
bw_iat_avg	Mean time between two packets sent in the backward direction
bw_iat_std	Standard deviation time between two packets sent in the backward direction
bw_iat_max	Maximum time between two packets sent in the backward direction
bw_iat_min	Minimum time between two packets sent in the backward direction
fw_psh_flag	Number of times the PSH flag was set in packets travelling in the forward direction (0 for UDP)
bw_psh_flag	Number of times the PSH flag was set in packets travelling in the backward direction (0 for UDP)
fw_urg_flag	Number of times the URG flag was set in packets travelling in the forward direction (0 for UDP)
bw_urg_flag	Number of times the URG flag was set in packets travelling in the backward direction (0 for UDP)
fw_hdr_len	Total bytes used for headers in the forward direction
bw_hdr_len	Total bytes used for headers in the backward direction
fw_pkt_s	Number of forward packets per second
bw_pkt_s	Number of backward packets per second
pkt_len_min	Minimum length of a flow
pkt_len_max	Maximum length of a flow
pkt_len_avg	Mean length of a flow
pkt_len_std	Standard deviation length of a flow
pkt_len_va	Minimum interarrival time of the packet
fin_cnt	Number of packets with FIN
syn_cnt	Number of packets with SYN
rst_cnt	Number of packets with RST
pst_cnt	Number of packets with PUSH
ack_cnt	Number of packets with ACK
urg_cnt	Number of packets with URG
cwe_cnt	Number of packets with CWE
ece_cnt	Number of packets with ECE
down_up_ratio	Download and upload ratio
pkt_size_avg	Average size of packet
fw_seg_avg	Average size observed in the forward direction
bw_seg_avg	Average size observed in the backward direction
fw_byt_blk_avg	Average number of bytes bulk rate in the forward direction
fw_pkt_blk_avg	Average number of packets bulk rate in the forward direction
fw_blk_rate_avg	Average number of bulk rate in the forward direction
bw_byt_blk_avg	Average number of bytes bulk rate in the backward direction
bw_pkt_blk_avg	Average number of packets bulk rate in the backward direction
bw_blk_rate_avg	Average number of bulk rate in the backward direction
subfl_fw_pk	The average number of packets in a subflow in the forward direction

TABLE 2: Continued.

Feature name	Description
subfl_fw_byt	The average number of bytes in a subflow in the forward direction
subfl_bw_pkt	The average number of packets in a subflow in the backward direction
subfl_bw_byt	The average number of bytes in a subflow in the backward direction
fw_win_byt	Number of bytes sent in initial window in the forward direction
bw_win_byt	# of bytes sent in initial window in the backward direction
Fw_act_pkt	# of packets with at least 1 byte of TCP data payload in the forward direction
fw_seg_min	Minimum segment size observed in the forward direction
atv_avg	Mean time a flow was active before becoming idle
atv_std	Standard deviation time a flow was active before becoming idle
atv_max	Maximum time a flow was active before becoming idle
atv_min	Minimum time a flow was active before becoming idle
idl_avg	Mean time a flow was idle before becoming active
idl_std	Standard deviation time a flow was idle before becoming active
idl_max	Maximum time a flow was idle before becoming active
idl_min	Minimum time a flow was idle before becoming active

varying between -1 and 1 . Finally, we remove all features having an importance index lower than 0.8 . As a result, the final dataset only comprises 40 features, excluding all labels. We note that In MidSiot, Pearson correlation was only used during the training phase to construct a set of concise and suitable features for machine learning models. This feature set is then saved and loaded to the IDS in the detection phases. This means that Pearson correlation is inactive in the detection phase. Therefore, it has no impact on the detection procedure.

Algorithm 1 Overview. Let $X = [x_1, x_2, \dots, x_n]$ denote a list of raw packets, and D is the list of processed flows. The major steps of this algorithm are described as follows:

- (1) **Network Flow Generation**, in Line 2, receives a list of raw packets X and generates network flows by aggregating packets sharing Source/DestinationIP, Source/DestinationPort, Protocol.
- (2) **Label Device Types**, in Line 3, the device type of each flow is deduced via Source/DestinationMAC Address. In addition, this step is performed during the training phase only.
- (3) **Drop and Normalize Data**, in Lines 4 and 5, any flows in F having a null value at any field will be dropped. Afterward, the remaining flows are normalized to facilitate the machine learning processes.
- (4) **Pearson's Correlation Coefficient**, in Line 6, the Pearson's correlation coefficient is applied on the normalized flows F_2 to select only important features F_t . Finally, the flow list F_2 will have some features dropped and only features from F_t are retained, which results in D .

3.4. Multistage Attack Detection Algorithm

3.4.1. The Overview. The details of the multistage attack detection algorithm is described in Algorithm 2. In more detail, let $X = [x_1, x_2, \dots, x_n]$ denote a list of raw packets and R is the resulting attack type. The entire detection process consists of the following main steps:

- (i) **Processing data**, including generating network flows, dropping unnecessary features, and normalizing data are performed similarly to the Algorithm 1. However, as this is the detection process, device-type labeling and features selecting using Pearson's correlation coefficient are inactive.
- (ii) **Classifying device type**, the classification model from the storage of the local home gateway to perform prediction on the processed flow f_2 to deduce the device type.
- (iii) **Sending the result to the ISP**, the processed flow f_2 in addition to the prediction results dt is forwarded to the ISP to further perform abnormality and attack-type detection.
- (iv) **Detecting the attacks**, the model m is applied on the flow f_2 to deduce whether that flow is normal or abnormal. If it is abnormal, move on to the next step; otherwise, mark this flow as null (which represents benign).
- (v) **Classifying attack type**, a universal attack-type detection model is loaded from the storage of the ISP gateway. This model will be then applied on the malicious flow f_2 to deduce the kind of attack a that has happened. At the end of this process, we know the device type of the flow, whether the flow is malicious or benign, and the attack type of the flow if it is malicious.

3.4.2. The First Stage. The primary benefit of the first stage is to enhance the accuracy of the attack detection model in the next steps. In more detail, since the IoT device types are various and heterogeneous, their network behaviors are highly diverse. For example, the high UDP packet rate coming from IoT cameras is normal, but the one from temperature sensors is a sign of a security threat. This may lead to false attack detection. Thus, identifying device types and considering them as an input feature of the attack detection model is crucial to increase the accuracy. Furthermore, this stage should be done on the local gateway for two reasons: (1) local gateways have sufficient

Input: Raw network packets X
Output: Processed data D

- (1) **Initialize:** $D =$
- (2) **Generate network flows:** $F \leftarrow \text{NetworkFlowGenerator}(X)$
- (3) **Label device type:** $\text{LabelDeviceType}(F)$
- (4) **Drop invalid flows and identity columns:** $F_1 \leftarrow \text{DropInvalidData}(F)$
- (5) **Normalize data:** $F_2 \leftarrow \text{Normalize}(F_1)$
- (6) **Select features using Pearson's Correlation Coefficient:**
 - (a) $\text{Fts} \leftarrow \text{Corr}(F_2)$
 - (b) $D \leftarrow \text{SelectFeatures}(F_2, \text{Fts})$
- (7) **Return** D

ALGORITHM 1: Data preprocessing in training phases.

Input: Raw network packets X
Output: Attack type R

- (i) **Local home gateway:**
 - (1) **Generate network flows:**
 $f \leftarrow \text{GenerateFlows}(X)$
 - (2) **Drop unnecessary features:**
 $f_1 \leftarrow \text{DropFeatures}(f)$
 - (3) **Normalize data:**
 $f_2 \leftarrow \text{Normalize}(f_1)$
 - (4) **Load device type classification model:**
 $M_1 \leftarrow \text{LoadModel}$
 - (5) **Classify device type:** Device type
 $dt \leftarrow M_1.\text{predict}(f_2)$
 - (6) **Send to ISP:** $\text{Send}((dt, f_2))$
- (ii) **Internet Service Provider (ISP)**
 - (1) **Initialize:** $R =$
 - (2) **Load all abnormality detection models:**
 $M_2 \leftarrow \text{LoadModel2}$
 - (3) **Load abnormality detection model:**
 $m \leftarrow M_2[dt]$
 - (4) **Attack detection:**
 $y \leftarrow m.\text{predict}(f_2)$
 - (5) **If y is not normal then**
 - (a) **Load attack type detection model:**
 $M_3 \leftarrow \text{LoadModel3}$
 - (b) **Attack classification:**
 $a \leftarrow M_3.\text{predict}(f_2)$
 - (c) $R \leftarrow a$
 - (6) **Else**
 $R \leftarrow \text{null}$
 - (7) **Return** R

ALGORITHM 2: The overall detection process.

computation power to handle a part of the detection process, which reduces the burden for the cloud; and (2) if device-type classification is done with the other two steps on the cloud, merge a multitude of network packets coming from various IoT networks. This aggregation may make the network data exhibit more generic characteristics than device-specific ones, reducing the device-type classification performance. This directly affects the attack detection accuracy. Therefore, running the device-type classifiers in the local gateways closed to IoT devices could mitigate this issue since only a limited number of device types are considered.

3.4.3. *The Second and Third Stages.* Take a dataset

$$D = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N), \quad (1)$$

where $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})^T$, $i = 1, 2, \dots, N$, x_i is the input instance that represents a network flow. x_i has n features. N indicates the number of features of a network flow contained in the dataset D . $y_i \in [0, 1, 2, \dots, K-1]$ is the result of each detection record. A decision tree recursively partitions the feature space such that the samples with the same labels or similar target values are grouped together.

Let the data at node m be represented by Q_m . For each split $\theta = (j, t_m)$ consisting of a feature j and a threshold t_m , partition the data into $Q_m^{\text{left}}(\theta)$ and $Q_m^{\text{right}}(\theta)$ subsets:

$$\begin{aligned} Q_m^{\text{left}}(\theta) &= \{(x, y) | x_j < t_m\}, \\ Q_m^{\text{right}}(\theta) &= Q_m \setminus Q_m^{\text{left}}(\theta). \end{aligned} \quad (2)$$

The quality of a candidate split of node m is then computed using an impurity function or loss function H , the choice of which depends on the task being solved (classification or regression):

$$G(Q_m, \theta) = \frac{N_m^{\text{left}}}{N_m} H(Q_m^{\text{left}}(\theta)) + \frac{N_m^{\text{right}}}{N_m} H(Q_m^{\text{right}}(\theta)). \quad (3)$$

Select the parameters to minimize the impurity:

$$\theta^* = \operatorname{argmin}_{\theta} G(Q_m, \theta). \quad (4)$$

Repeat for subsets $Q_m^{\text{left}}(\theta)$ and $Q_m^{\text{right}}(\theta)$ until the maximum allowable depth is reached $N_m < \min_{\text{sample}}$ or $N_m = 1$.

For the classification of IDS, $y_i \in [0, 1, 2, \dots, K-1]$ for node m represents a region of R_m with instances of N_m . Assume that p_{mk} is the proportion of class k instance in m and can be obtained by the following formula:

$$p_{mk} = \frac{1}{N_m \sum_{x_j \in R_m} I(y_i = k)}. \quad (5)$$

The common measure of impurity is named Gini and can be obtained by the following formula:

$$H(X_m) = - \sum_k p_{mk} (1 - p_{mk}). \quad (6)$$

Cross-entropy can be obtained by the following formula:

$$H(X_m) = - \sum_k p_{mk} \log(p_{mk}). \quad (7)$$

Misclassification can be obtained by the following formula (not being used in the proposed system):

$$H(X_m) = - \sum_k 1 - \max(p_{mk}). \quad (8)$$

4. Results and Discussion

4.1. Evaluation Metrics. In our experiments, we adopted several evaluation metrics, such as precision (P), recall (R), F -measure (F), and accuracy. Let TP , FP , and FN denote true positives, false positives, and false negatives, respectively, and these evaluation metrics are defined as

$$\begin{aligned} P &= \frac{TP}{TP + FP}, \\ R &= \frac{TP}{TP + FN}, \\ F &= 2 \cdot \frac{P \cdot R}{P + R}. \end{aligned} \quad (9)$$

To evaluate the quality of attack detection models, we use the model accuracy as a primary evaluation metric that is directly computed from the confusion matrix based on the following formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (10)$$

We evaluate the performance of these resampling methods through macroaverage $F1$ -score performed independently on each class. Let MAP and MAR denote macroaverage precision and macroaverage recall, and the macroaverage $F1$ -score is defined as

$$\text{MacroAvgF1} = 2 * \frac{\text{MacroAvgPrec} * \text{MacroAvgRec}}{\text{MacroAvgPrec}^{-1} + \text{MacroAvgRec}^{-1}}, \quad (11)$$

where

$$\text{MacroAvgPrec} = \frac{\sum_{k=1}^K \text{Precision}_k}{K}, \quad (12)$$

$$\text{MacroAvgRec} = \frac{\sum_{k=1}^K \text{Recall}_k}{K},$$

where K is the total number of classes.

4.2. Dataset and Attack Class Balancing. We assessed Mid-Siot on three different datasets: IoTID20, CIC-IDS-2017, and BOT-IoT.

- (i) The IoTID20 dataset consists of two IoT devices (a smart home device SKT NGU and an EZVIZ Wi-Fi camera) and several non-IoT devices marked as external devices. The cyberattacks on these devices are classified into four attack categories and seven attack subcategories described in detail in Table 3.
- (ii) The CIC-IDS-2017 dataset contains the network traffic of six cyberattack types listed in Table 4 targeting 12 different IoT devices, which are labeled according to their operating systems and architectures. It also has several external devices to generate normal traffic.
- (iii) The BOT-IoT has five IoT devices and several external devices. The malicious network traffic of these devices is classified into three attack categories and detailed in Table 5.

Through rigorously analyzing the evaluation datasets, we figured out that the number of samples of each attack type is slightly imbalanced; thus, employing resampling methods is necessary. In our experiments, we experimented and evaluated the Random Undersampling algorithm (RU) and its conjunction with Synthetic Minority Oversampling Technique (RU-SMOTE).

- (i) Random Undersampling is a random selection process running on overwhelmed attack types to reduce their size. However, randomly selecting data points might accidentally ignore critical information, resulting in degraded classification performance.

TABLE 3: Devices, attack categories, and subcategories of the IoTID20 dataset.

Device	Category	Subcategory
EZVIZ, NUGU, External	Normal, DoS, Mirai, MITM, Scan	Normal, Syn Flooding, Brute Force, HTTP Flooding, UDP Flooding, ARP Spoofing, Host Port, OS

TABLE 4: Devices, attack categories, and subcategories of the CIC-IDS-2017 dataset.

Device	Category	Subcategory
Web server 16 Public, Ubuntu server 12 Public, Ubuntu 14.4 32bit, Ubuntu 14.4 64bit, Ubuntu 16.4 32bit, Ubuntu 16.4 64bit, Win 7 Pro, Win 8.1 64bit, Win Vista 64bit, Win 10 pro 32bit, win 10 64bit, MAC, External	Benign, Bot, Brute Force, Dos/Ddos, Infiltration, Portscan, Web Attack	Benign, Bot, FTP-Patator, SSH-Patator, DDoS, DoS, GoldenEye, DoS Hulk, DoS Slow, httptest, DoS slowloris, Infiltration, Portscan, Web Attack-Brute Force, Web Attack-Sql Injection, Web Attack-XSS, Heartbleed

TABLE 5: Devices, attack categories, and subcategories of the BOT-IoT dataset.

Device	Category	Subcategory
Ubuntu Server, Ubuntu Mobile, Metasploitable, Windows 7, Ubuntu Tap, External	Normal, DoS/DDoS, Reconnaissance, Theft	Normal, Service scanning, OS Fingerprinting, DoS/DDoS TCP, DoS/DDoS UDP, DoS/DDoS HTTP, Keylogging, data exfiltration

- (ii) Synthetic Minority Oversampling Technique (SMOTE) balances the dataset by synthesizing new samples for the minority class. In more detail, it selects a cluster of samples and draws a line between them; new samples are the points along this line.

To implement these resampling methods, we utilized *imblearn* library [33] supporting multiple resampling techniques along with several running strategies for both binary and multiclass classification. In binary classification, we need to configure the ratio between minority class and majority class after resampling, whereas this configuration is unnecessary in multiclass classification. From the results illustrated in Figures 4-6, it is obvious that these resampling techniques have no significant impact on the performance of models. Furthermore, if the number of samples in majority classes drastically outweighs ones of minority classes, they may decrease the attack detection quality of the smaller classes. Therefore, our models are trained without resampling techniques.

4.3. Results and Discussion. To select classification algorithms for our multistage IDS, we examined the detection time and accuracy of several supervised machine learning algorithms on the IoTID20 dataset. In detail, the ability to

classify benign and malicious network traffic (binary attack detection) and identify precisely the types of attacks (multiclass attack detection) are both considered. Since operating on network gateways requires a lightweight attack detection model, experimented algorithms are simple machine learning algorithms, including linear support vector machine, quadratic support vector machine, K-nearest-neighbor, linear discriminant analysis, quadratic discriminant analysis, multilayer perceptron, long short-term memory, autoencoder classifier, and decision tree classifier; their results are presented in Table 6. As shown in the table, the decision tree classifier outperforms other algorithms, and it is considered a lightweight machine learning algorithm [34]. This classifier is thus selected for the attack detection model of MidSiot. Note that empty values in the table (denoted by N/A) imply the long training time exceeding two hours. Moreover, training a decision tree classifier is trivial and possibly performed on IoT devices. We also applied *Classification and Regression Tree* (CART) to boost the detection performance. In more detail, CART splits training data into two subsets based on a specific feature k and a threshold t_k (e.g., “flow duration ≤ 100 ”). This split is repeated on each subset until it reaches the maximum depth or subset size equals to 0. As a result, the computational complexity of the classifier is reduced to $O(\log_2(m))$ with m

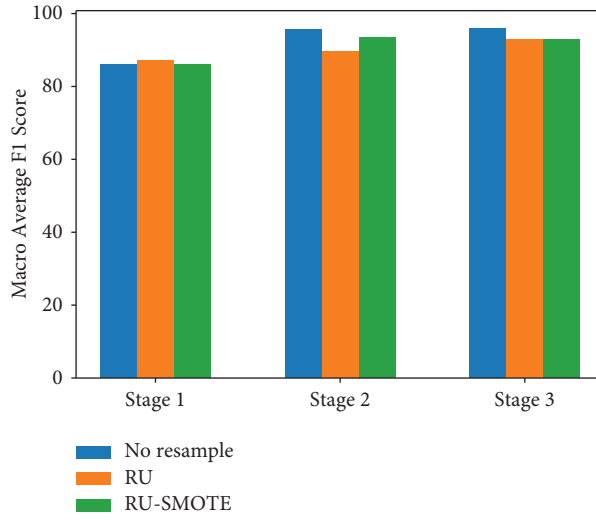


FIGURE 4: The macroaverage $F1$ -score of the IoTID20 dataset.

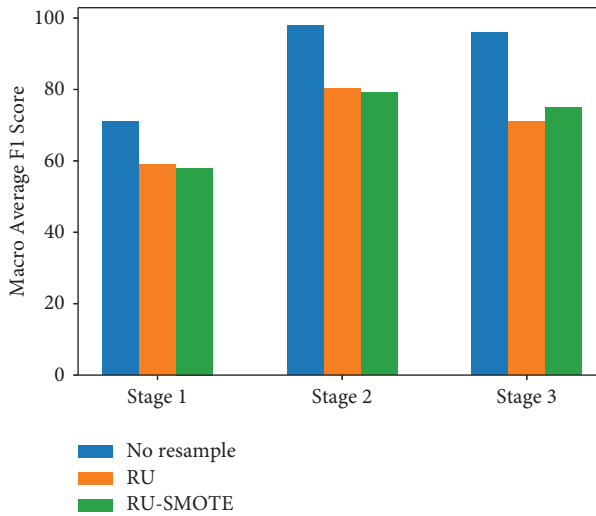


FIGURE 5: The macroaverage $F1$ -score of the CIC-IDS-2017 dataset.

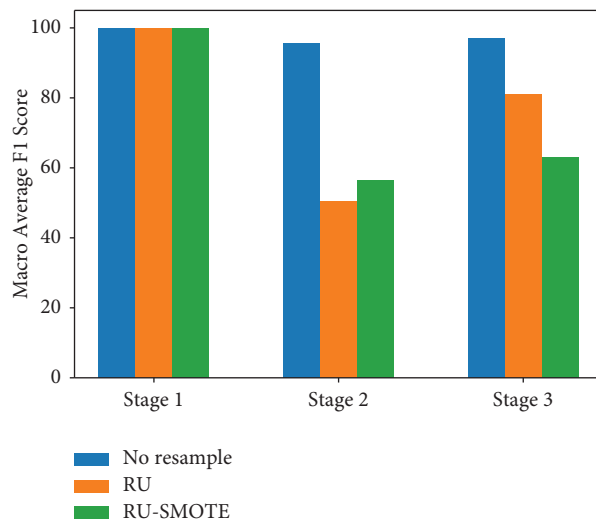


FIGURE 6: The macroaverage $F1$ -score of the BOT-IoT dataset.

TABLE 6: The results of examined classifiers.

Model	Classification	Accuracy (%)	Training time (μ s/flow)	Prediction time (μ s/flow)	Adjusted parameters
Linear support vector machine	Binary	98.16	1150.57	204.54	kernel = linear gamma = auto
	Multiclass	N/A	10 298.3	1453.1	
Quadratic support vector machine	Binary	98.25	792.61	311.5	kernel = poly gamma = auto
	Multiclass	N/A	N/A	N/A	
K-Nearest neighbor	Binary	99.79	0.17	2343.72	n_neighbors = 5
	Multiclass	98.61	0.19	2377.81	
Linear discriminant analysis	Binary	95.07	21.6	0.27	All default
	Multiclass	80.73	25.71	4.35	
Quadratic discriminant analysis	Binary	53.6	18.89	12.44	All default
	Multiclass	56.62	16.62	14.74	
Multilayer perceptron	Binary	99.6	2.57	0.65	Input layer and first layer with 50 neurons and activation = relu Output layer with activation = sigmoid input layer and LSTM layer with 50 neurons
	Multiclass	92.71	4.94	7.96	
Long short-term memory	Binary	96.53	572.44	62.64	Output layer with activation = sigmoid Encoding layer with 50 neurons and activation = relu
	Multiclass	N/A	N/A	N/A	
Autoencoder classifier	Binary	93.01	11.65	0.62	Decoding and output layer with activation = softmax
	Multiclass	87.74	13.35	0.81	
Decision tree classifier	Binary	99.94	12.49	0.43	criterion = entropy
	Multiclass	99.69	16.76	0.38	

TABLE 7: The evaluation results of the multistage IDS.

Dataset	Stage	Accuracy (%)	Training time (μ s/flow)	Prediction time (μ s/flow)
IoTID20	1	92.57	35.76	0.44
	2	99.95	11.33	0.26
	3	99.15	15.9	0.31
CIC-IDS-2017	1	94.17	110.31	0.89
	2	99.99	25.65	0.25
	3	99.97	78.68	0.27
BOT-IoT	1	99.92	19.27	0.15
	2	99.99	3.24	0.07
	3	99.93	15.05	0.17

being the number of samples in the training set. This significantly increases the training and prediction rate to deal with large datasets.

Attack detection results: Table 7 reports the overall performance of each MidSiot’s stage on evaluated datasets. We can see that MidSiot could not only accurately differentiate between normal and malicious traffic in the second stage, but also identify the type of attacks in the third stage. The average accuracy of such stages is about 99.98% and 99.68%, respectively. Regarding classifying IoT devices, our proposal achieved a high classification accuracy reported at 95.55% on average. In detail, device-type classification for BOT-IoT achieves the highest result, at 99.92%, whereas the results of IoTID20 and CIC-IDS-2017 are about 92.57% and 94.17%, respectively. To have a better understanding the attack detection performance of MidSiot, Tables 8–10 illustrate the confusion matrices, which present the comparison between predicted attacks and the actual ones.

Comparing with baseline methods: We compared attack detection quality between our proposal and state-of-the-art IDS and reported the results in Table 11. Overall, MidSiot

outperforms its competitors on CICIDS-2017 and BOT-IoT datasets and is comparable with them on IoTID2020 dataset. In more detail, regarding CIC-IDS-2017, the best of our competitors achieves 99.9% in both binary and multiclass classification, whereas our proposed IDS achieves better results recorded about 99.99% and 99.97%, respectively. Similar results are also found in the BOT-IoT dataset, in which our proposal achieves 99.99% accuracy in binary classification and 99.93% in multiclass classification problems. In the IoTID20 dataset, the best competitor detects the attack types with 100% accuracy, and MidSiot also has very competitive results reported at 99.15%. Compared with state-of-the-art IDSs, MidSiot employed more machine learning models to enhance detection accuracy. This demands high computation costs and training datasets to train these models. Indeed, each stage in MidSiot has a different training dataset, which requires a huge effort to label. For example, the first-stage model needs to label the device type of network traffic, whereas the second-stage model demands labeling abnormal traffic. Moreover, deploying the first-stage model from the cloud to IoT local gateway consumes network bandwidth and may trigger delays.

TABLE 8: Attack detection confusion matrix on the IoTID20 dataset.

		Predicted				
		Mirai	Scan	DoS	Normal	MITM ARP spoofing
Actual	Mirai	412446	2583	0	38	569
	Scan	618	74 473	0	9	155
	DoS	9	2	59356	5	3
	Normal	63	15	6	53	30
	MITM ARP Spoofing	624	200	2	16	34509

TABLE 9: Attack detection confusion matrix on the CIC-IDS-2017 dataset.

		Predicted				
		Benign	DoS/DDoS	Portscan	Bot	Web attack
Actual	Benign	158303	6	2	7	35
	DoS/DDoS	19	235881	8	0	2
	Portscan	6	6	158911	0	1
	Bot	10	0	0	1929	0
	Web Attack	40	1	2	0	1463

TABLE 10: Attack detection confusion matrix on the BOT-IoT dataset.

		Predicted			
		DDoS	DoS	Reconnaissance	Normal
Actual	DDoS	1926189	435	0	0
	DoS	1319	830786	0	0
	Reconnaissance	2	6	81818	2
	Normal	1	1	3	2

TABLE 11: Comparing the proposed system with other one-stage systems.

Dataset	Work	Attack detection model	Binary attack detection (%)	Multiclass attack detection (%)	Prediction time (μ s/netflow)
CICIDS-2017	MidSiot	Multistage	99.99	99.97	1.41
	Gamage and Samarabandu [35]	Random forest	N/A	99.86	60.48
	Vinayakumar et al. [36]	Deep neural network	93.10	95.60	N/A
	Elmrabit et al. [37]	Decision tree	99.90	99.90	N/A
	Manimurugan	Deep belief network	99.37	97.73	N/A
	MidSiot	Multistage	99.98	99.88	1.01
IOTID20	Ullah and Mahmoud [21]	Decision tree	99.94	99.69	N/A
	Alkahtani and Aldhyani [38]	Long short-term memory	98.20	N/A	N/A
	Song et al. [39]	Autoencoder	93.76	95.20	N/A
	Hussein et al. [40]	Random forest	99.90	99.90	N/A
	Ullah and Mahmoud [41]	Convolution neural network	99.98	97.76	N/A
	Islam et al. [42]	Decision tree	N/A	100.00	1139.37
BOT-IOT	MidSiot	Multistage	99.99	99.99	0.39
	Ferrag et al. [43]	Rules and decision tree	N/A	97.00	1.54
	Ferrag et al. [44]	Deep autoencoder	N/A	98.39	1916.55
	Dwibedi et al. [45]	Support vector machine	99.99	N/A	N/A
	Pokhrel et al. [46]	K-Nearest neighbor	92.10	N/A	N/A
	Ge et al. [47]	Support vector machine	99.74	99.03	693 040
	Ullah and Mahmoud [41]	Convolution neural network	99.90	99.97	N/A

In conclusion, by using a hierarchical architecture and chaining stages together, MidSiot effectively classifies device types, identifies abnormal network traffic, and differentiates cyberattack types.

5. Conclusion

In this article, we proposed a distributed intrusion detection system for IoT scenarios, in which connected devices are not only resource-constraint but also heterogeneous in hardware specification. To accurately detect various types of cyberattacks, the proposed IDS consists of three stages: (1) classifying device types; (2) detecting malicious network flows; and (3) identifying attack types. In the experiments on three popular IOT-IDS datasets (IoTID20, CIC-IDS-2017, and BOT-IoT), we demonstrated that our proposal could detect several attacks with an accuracy of 99.68% on average and outperforms state-of-the-art IDSs. In addition, we examined two resampling techniques to balance the datasets and discovered that these techniques slightly reduce the detection rate of minority attack types. In short, MidSiot is beneficial for both the industrial and research communities interested in further developing intrusion detection systems for IoT.

Data Availability

The training 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 research was funded by the fund supporting research activities from the University of Information Technology, Vietnam National University, Ho Chi Minh City.

References

- [1] W. a. Kassab and K. A. Darabkh, "A-Z survey of internet of things: architectures, protocols, applications, recent advances, future directions and recommendations," *Journal of Network and Computer Applications*, vol. 163, Article ID 102663, 2020.
- [2] A. Menard, "How can we recognize the real power of the internet of things," *Advanced Robotics*, vol. 1, pp. 4-5, 2017.
- [3] J. Sengupta, S. Ruj, and S. Das Bit, "A comprehensive survey on attacks, security issues and blockchain solutions for iot and iiot," *Journal of Network and Computer Applications*, vol. 149, Article ID 102481, 2020.
- [4] M. Eskandari, Z. H. Janjua, M. Vecchio, and F. Antonelli, "Passban ids: an intelligent anomaly-based intrusion detection system for iot edge devices," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6882-6897, 2020.
- [5] G. Kambourakis, C. Kolias, and A. Stavrou, "The mirai botnet and the iot zombie armies," in *Proceedings of the MILCOM 2017-2017 IEEE Military Communications Conference (MILCOM)*, pp. 267-272, IEEE, Baltimore, MD, USA, October 2017.
- [6] G. Gallopeni, B. Rodrigues, M. Franco, and B. Stiller, "A practical analysis on mirai botnet traffic," in *Proceedings of the 2020 IFIP Networking Conference (Networking)*, pp. 667-668, IEEE, Espoo, Finland, June 2020.
- [7] A. Gangwar and S. Sahu, "A survey on anomaly and signature based intrusion detection system (ids)," *International Journal of Engineering Research and Applications*, vol. 4, no. 4, 2014.
- [8] C. Zhang, J. Jiang, and M. Kamel, "Intrusion detection using hierarchical neural networks," *Pattern Recognition Letters*, vol. 26, no. 6, pp. 779-791, 2005.
- [9] K. Wang and S. J. Stolfo, "Anomalous payload-based network intrusion detection," *Lecture Notes in Computer Science*, Springer, in *Proceedings of the International Workshop on Recent Advances in Intrusion Detection*, pp. 203-222, September 2004.
- [10] M. Xie, J. Hu, S. Han, and H.-H. Chen, "Scalable hypergrid k-nn-based online anomaly detection in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1661-1670, 2012.
- [11] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: an ensemble of autoencoders for online network intrusion detection," 2018. arXiv preprint arXiv:1802.09089.
- [12] K. Ince, "A novel approach for intrusion detection systems: V-ids," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, no. 4, pp. 1929-1943, 2021.
- [13] V. Kumar, A. K. Das, and D. Sinha, "Uids: a unified intrusion detection system for iot environment," *Evolutionary Intelligence*, vol. 14, no. 1, pp. 47-59, 2021.
- [14] E. Anthi, L. Williams, M. Slowinska, G. Theodorakopoulos, and P. Burnap, "A supervised intrusion detection system for smart home iot devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9042-9053, 2019.
- [15] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: bot-iot dataset," *Future Generation Computer Systems*, vol. 100, pp. 779-796, 2019.
- [16] Z. Liu, N. Thapa, A. Shaver, K. Roy, X. Yuan, and S. Khorsandroo, "Anomaly detection on iot network intrusion using machine learning," in *Proceedings of the 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1-5, IEEE, KwaZulu Natal, South Africa, August 2020.
- [17] H. Kaur, G. Singh, and J. Minhas, "A review of machine learning based anomaly detection techniques," 2013. arXiv preprint arXiv:1307.7286.
- [18] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications surveys & tutorials*, vol. 18, no. 2, pp. 1153-1176, 2015.
- [19] M. A. Aydın, A. H. Zaim, and K. G. Ceylan, "A hybrid intrusion detection system design for computer network security," *Computers & Electrical Engineering*, vol. 35, no. 3, pp. 517-526, 2009.
- [20] M. Gajewski, J. M. Batalla, G. Mastorakis, and C. X. Mavromoustakis, "A distributed ids architecture model for smart home systems," *Cluster Computing*, vol. 22, no. 1, pp. 1739-1749, 2019.
- [21] I. Ullah and Q. H. Mahmoud, "A scheme for generating a dataset for anomalous activity detection in iot networks," in *Proceedings of the Canadian Conference on AI*, pp. 508-520, Ottawa, Ontario, May 2020.

- [22] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSp*, vol. 1, pp. 108–116, 2018.
- [23] B. Silva, R. Silveira, M. Silva Neto, P. Cortez, and D. Gomes, "A comparative analysis of undersampling techniques for network intrusion detection systems design," *Journal of Communication and Information Systems*, vol. 36, no. 1, pp. 31–43, 2021.
- [24] S. Bagui and K. Li, "Resampling imbalanced data for network intrusion detection datasets," *Journal of Big Data*, vol. 8, no. 1, pp. 1–41, 2021.
- [25] P. Bedi, N. Gupta, and V. Jindal, "I-siamids: an improved siam-ids for handling class imbalance in network-based intrusion detection systems," *Applied Intelligence*, vol. 51, no. 2, pp. 1133–1151, 2021.
- [26] W. Meng, E. W. Tischhauser, Q. Wang, Y. Wang, and J. Han, "When intrusion detection meets blockchain technology: a review," *IEEE Access*, vol. 6, pp. 10179–10188, 2018.
- [27] W. Li, Y. Wang, J. Li, and M. H. Au, "Towards blockchain challenge-based collaborative intrusion detection," in *Applied Cryptography and Network Security Workshops*, J. Zhou, R. Deng, Z. Li et al., Eds., Springer International Publishing, Cham, Switzerland, pp. 122–139, 2019.
- [28] W. Li, Y. Wang, J. Li, and M. H. Au, "Toward a blockchain-based framework for challenge-based collaborative intrusion detection," *International Journal of Information Security*, vol. 20, no. 2, pp. 127–139, 2021.
- [29] W. Li, S. Tug, W. Meng, and Y. Wang, "Designing collaborative blockchain signature-based intrusion detection in iot environments," *Future Generation Computer Systems*, vol. 96, pp. 481–489, 2019.
- [30] R. M. A. Ujjan, Z. Pervez, and K. Dahal, "Snort based collaborative intrusion detection system using blockchain in sdn," in *Proceedings of the 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, pp. 1–8, Ukulhas, Maldives, August 2019.
- [31] W. Fan, Y. Park, S. Kumar, P. Ganta, X. Zhou, and S.-Y. Chang, "Blockchain-enabled collaborative intrusion detection in software defined networks," in *Proceedings of the 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp. 967–974, Guangzhou, China, November 2020.
- [32] O. Alkadi, N. Moustafa, B. Turnbull, and K.-K. R. Choo, "A deep blockchain framework-enabled collaborative intrusion detection for protecting iot and cloud networks," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9463–9472, 2020.
- [33] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 559–563, 2017.
- [34] R. Bikmukhamedov and A. Nadeev, "Lightweight machine learning classifiers of iot traffic flows," in *Proceedings of the 2019 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*, pp. 1–5, IEEE, Minsk, Belarus, July 2019.
- [35] S. Gamage and J. Samarabandu, "Deep learning methods in network intrusion detection: a survey and an objective comparison," *Journal of Network and Computer Applications*, vol. 169, Article ID 102767, 2020.
- [36] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.
- [37] N. Elmrabit, F. Zhou, F. Li, and H. Zhou, "Evaluation of machine learning algorithms for anomaly detection," in *Proceedings of the 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pp. 1–8, IEEE, Dublin, Ireland, June 2020.
- [38] H. Alkahtani and T. H. Aldhyani, "Intrusion detection system to advance internet of things infrastructure-based deep learning algorithms," *Complexity*, vol. 2021, Article ID 5579851, 18 pages, 2021.
- [39] Y. Song, S. Hyun, and Y.-G. Cheong, "Analysis of autoencoders for network intrusion detection," *Sensors*, vol. 21, no. 13, p. 4294, 2021.
- [40] A. Y. Hussein, P. Falcarin, and A. T. Sadiq, "Enhancement performance of random forest algorithm via one hot encoding for iot ids," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 9, no. 3, pp. 579–591, 2021.
- [41] I. Ullah and Q. H. Mahmoud, "Design and development of a deep learning-based model for anomaly detection in iot networks," *IEEE Access*, vol. 9, pp. 103906–103926, 2021.
- [42] N. Islam, F. Farhin, I. Sultana et al., "Towards machine learning based intrusion detection in iot networks," *Computers, Materials & Continua*, vol. 69, no. 2, pp. 1801–1821, 2021.
- [43] M. A. Ferrag, L. Maglaras, A. Ahmim, M. Derdour, and H. Janicke, "Rdtids: rules and decision tree-based intrusion detection system for internet-of-things networks," *Future Internet*, vol. 12, no. 3, p. 44, 2020.
- [44] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study," *Journal of Information Security and Applications*, vol. 50, Article ID 102419, 2020.
- [45] S. Dwibedi, M. Pujari, and W. Sun, "A comparative study on contemporary intrusion detection datasets for machine learning research," in *Proceedings of the 2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 1–6, IEEE, Arlington, VA, USA, November 2020.
- [46] S. Pokhrel, R. Abbas, and B. Aryal, "Iot security: botnet detection in iot using machine learning," arXiv preprint arXiv: 2104.02231, 2021.
- [47] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo, and A. Robles-Kelly, "Deep learning-based intrusion detection for iot networks," in *Proceedings of the 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC)*, pp. 256–25609, IEEE, Kyoto, Japan, December 2019.

Research Article

Real-Time 3D Pedestrian Tracking with Monocular Camera

Peng Xiao ^{1,2}, Fei Yan ¹, Jiannan Chi ^{2,3} and Zhiliang Wang ¹

¹*School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China*

²*Engineering Research Center of Intelligence Perception and Autonomous Control, Beijing University of Technology, Beijing 100124, China*

³*School of Automation and Electronic Engineering, University of Science and Technology Beijing, Beijing 100083, China*

Correspondence should be addressed to Jiannan Chi; ustbjnc@ustb.edu.cn

Received 3 October 2021; Revised 5 November 2021; Accepted 25 November 2021; Published 17 February 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 Peng Xiao 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.

Target tracking has always been a popular research area in computer vision, and many important methods have been proposed. However, most methods can only solve partial and slight occlusion. If the target is lost, a common solution is to keep detecting, reidentify the target when it reappears, and then link the broken tracks together, but this makes tracking discontinuous. There are two key points in this problem: continuous tracking and occlusion judgment. In this paper, we propose a target tracking method with a short-time prediction function to solve this problem. For continuous tracking, we establish a 3D dynamic model to estimate the motion state of the target in each frame. For occlusion judgment, we use a depth prediction network to estimate the depth of the target and then determine whether the target is occluded by the depth. Without relying on depth sensors or multiple cameras, we achieve depth estimation using only a single monocular image, which greatly expands the application of our method. Benefit from the introduction of motion estimation and depth prediction, the tracking accuracy of our method has been significantly improved, especially for better robustness to occlusion. Even when the target is completely occluded, it can be tracked for a short time without reidentification. In addition, we improve the speed of depth prediction through knowledge distillation by 2.08 times, and the final tracking speed reaches 52.6 Hz on GPU, which meets the real-time tracking requirements.

1. Introduction

Modern society produces a large number of videos every day. As an important means of video analysis, video object tracking has a wide range of applications, such as autonomous driving [1], robotics [2], and augmented reality [3]. Although great progress has been made over these years, most methods are based on the assumption that the target is visible. Therefore, these methods can only solve partial and slight occlusion problems. However, in daily life, the target is often completely obscured. There are some ways to solve this problem, and the mostly used one is reidentification, but this breaks the continuity of the tracking. Even if the target is not visible, it still exists, and we should speculate its location based on our experience. And in some cases, we cannot pay the consequences of ignoring the completely missing target, such as online intelligent driving. To solve

this problem, we can start from two aspects: continuous tracking and occlusion judgment.

The key to continuous tracking is that the tracker should be able to use the historical information of the target's motion to estimate the current motion state of the target. Usually we assume that the motion of the target conforms to certain rules, so we can establish its motion model. Even if the target is lost, we can also infer the target's state according to the model. Regression model is a common prediction model in engineering, but regression model is not applicable to state prediction in tracking problem. Because the regression model requires that all points in state spaces roughly conform to a known relationship, such as a linear relationship or a polynomial relationship, which is almost impossible to satisfy in visual tracking. In addition, the regression model reflects the overall trend, and the prediction error of a single point may be very large, which may even directly

lead to tracking failure. A more practical prediction method is Kalman filtering (KF) [4], which has been widely used in tracking problems. It first predicts the ideal state value through the state transition equation and then corrects the predicted state value and the model based on the actual measured value. Compared with regression models, KF has three advantages. The first is low computation cost. Only five formulas are needed to perform in each a time step. The second is flexible. We can customize the target's state values that need to be estimated, such as coordinates, aspect ratio, speed, and acceleration, and the relationship between state values can also be defined by ourselves. The third is strong adjustment ability. Some parameters in KF model can be adjusted over time to ensure that the model conforms to the current motion state as much as possible. After comprehensive consideration, KF will be used for state estimation in this paper.

Occlusion judgment is a very challenging problem, especially for 2D tracking. In the visual tracking problem, occlusion occurs frame by frame, and the target and the block are gradually mixed together. There is no general and effective way to separate the target and the block or to measure the level of occlusion. In 3D tracking, the tracker usually calculates the distance between the target and the camera, that is, the depth. It is easy to determine whether and how much the target is occluded by simply checking the depth value of key points on the target. Since the block is nearer to the camera than the target, the depth value of the target region will suddenly increase once occlusion occurs. The only problem is how to obtain the depth value. The most direct way to obtain depth information is to use a depth sensor, such as Kinect and laser scanner. Although the hardware-based approaches can obtain accurate depths, the dependence on hardware also greatly limits the applicability of these approaches, and they cannot handle images that are already captured without depth information. For most tracking tasks, we only need to distinguish the relative position of each object in space, not the absolute position, so it is a more practical solution to estimate the relative position of the object through a model. This has been a hot topic in 3D tracking in recent years, and we will discuss this in detail in Section 2. In this paper, we will introduce a practical neural network [5] to predict the depth due to its excellent performance and published depth dataset. To speed up depth prediction, we carry out a knowledge distillation to the original net [5] and get great improvement.

Another point to focus on is target detection. Most of the state-of-the-art (SOTA) tracking methods are tracking-by-detection (TBD) framework, and the performance of the detector directly affects the accuracy of target localization. Before the rise of deep learning, landmark target detection methods include Viola Jones Detectors [6], HOG Detector [7], and Deformable Part-based Model (DPM) [8]. These well-designed artificial feature detectors often achieve good results on specific tasks but had poor generalization capabilities. Since 2014, various deep learning-based target detection methods having continuously refreshed the record of target detection and many classic methods have also been proposed, such as RCNN [9], Faster RCNN [10], YOLO

[11], SSD [12], RetinaNet [13], and Fast YOLO [14]. Here, we select Fast YOLO [14] as private target detector because of its high speed.

In this paper, we study the pedestrian tracking problem in video surveillance and aim to achieve real-time continuous 3D tracking using a single camera. Inspired by SORT [15], we propose a multipedestrian tracking method incorporating depth information. The basic idea is to use a Kalman filter to continuously estimate the motion state parameters for each target. To ensure the continuity of tracking, we do prediction when occlusion happens or the track does not match any target. The measurement of a KF consists of two parts: plane information and depth information, which are obtained by a target detection neural network and depth prediction network, respectively. To ensure real-time tracking, the neural networks used are trained in advance and are not updated online. Figure 1 shows the tracking effect of our method. When pedestrians are partially or completely occluded, our tracker still keeps tracking. In contrast, most trackers give up tracking once they cannot detect the targets.

To summarize, this paper presents a method for 3D tracking with a fixed monocular camera. The contributions of our work are summarized as follows:

- (1) We utilize KF's short-term prediction capabilities to achieve continuous tracking. The original SORT method cannot detect occlusion, so when the target is occluded, its trajectory has to be ended. While we introduce depth information to make occlusion detection very easy
- (2) We fuse the uncalibrated depth information to achieve 3D tracking on 2D images. We use a neural network to estimate the depth information of the target in real time, thereby turning the original 2D tracking into a 3D tracking
- (3) We take a series of measures to speed up tracking. The most important strategy is knowledge distillation of the deep prediction network. Beside this, we select a faster target detector and remove reidentification step that is usually used in MOT methods

The remainder of this paper is organized as follows. Section 2 introduces some related studies about multitarget detection, tracking, and depth prediction. Section 3 presents the overall architecture of our method and details its improved parts. In Section 4, the experimental results verify the effectiveness of the proposed method. In Section 5, we provide a comprehensive summary of this study, and a future research direction is presented.

2. Related Work

2.1. Multiple Pedestrian Tracking. Compared with single object tracking (SOT), multiple object tracking (MOT) is more complicated because there are extra issues to be considered, such as the matching of trajectory and target and target reidentification. Multiple pedestrian tracking (MPT)



FIGURE 1: Visualization of tracking when occlusion happens. Red bounding boxes represent occluded pedestrians that are tracked by our method. The figures below are partial enlargement of the above ones.

has become the main research direction of MOT. The related research work focuses on the following four areas for improvement: (a) design the association methods, (b) joint other vision tasks, (c) apply deep learning to MPT, and (d) multi-modality-based MPT. The core of TBD framework is data association. Some classical association methods are still used as basic algorithms. Hungarian method is a classic algorithm for solving the minimum weight matching problem of bipartite graph and is introduced into MPT by Singh et al. [16]. Although the algorithm is fast, the accuracy is not high due to local optimal nature. NF [17] extends local optimization to global optimization, and CRF [18] further considers association dependencies. GMMCP [19] are proposed to solve the problem of high computation complexity. Different from these graph-based methods, MCSM [20] formulates the association as a minimum cost subgraph multicut problem that links and performs clustering for the multiple plausible person detection jointly over time and space. Several researchers have leveraged other vision tasks to improve the tracking performance. One approach is to treat MOT as an extension of SOT. For example, Hu et al. [21] use Siamese-RPN to locate the target location. The other is to combine with image segmentation. Voigtlaender et al. [22] propose a new baseline method which jointly addresses detection, tracking, and segmentation with a single convolutional network. Since Kim et al. [23] first utilize CNN to extract 4096-dimensional features for each detection box, deep neural network models have been widely used in MPT, such as VGGNet [24, 25] and GoogleNet [21, 26]. When a single type of data is unreliable, people try to use multimodal data, such as Zhang et al. [27] using image and point cloud features and Gautam et al. [28] using image and radar features.

2.2. Depth Prediction. Scene depth estimation is an old-standing problem in vision. The most direct way to obtain depth information is to use a depth sensor, such as Kinect and laser scanner. Although sensors can capture accurate depth information, they are only applicable to specific scenes and cannot process existing RGB images. In recent years, using deep neural networks to estimate scene depth has

become a mainstream research direction. Various DCNN-based methods focus on designing structural features, especially in depth prediction. Fu et al. [29] propose an encoder-decoder network, which extracts multiscale features from the encoder and is trained in an end-to-end manner without iterative refinement. Jiao et al. [30] propose an attention-driven loss, which merges the semantic priors to improve the prediction precision on unbalanced distribution datasets. Chen et al. [31] apply the generative adversarial training to lead the network to learn a context-aware and patch-level loss automatically.

The basic idea of these methods is to train a deep network model with a RGBD dataset and then reconstruct the 3D structure of the target from the RGB image to predict the depth information. Deep network models are data-driven, and the models achieve good results when trained with large amounts of samples, but it is still not an easy task to obtain large amounts of RGBD data. Silberman et al. [32] construct the NYU dataset using Kinect, but it was limited to indoors. Although the datasets Make3D [33] and KITTI [34] built with laser scanners can be used outdoors, they are collected in specific scenarios (a university campus and atop a car, respectively). Another way to collect depth data for training is to ask people to manually annotate depth in images, but it is not only time-consuming and laborious but also can only give the relative position of the object. Estimating geometry from Internet photo collections has been an active research area for a decade. Li et al. [5] propose a method to generate an infinite dataset. First, a large number of images are collected from the Internet, and then, the structure-from-motion (SfM) and multiview stereo (MVS) method are used to generate depth maps, and these depth maps are further processed to form a large-scale depth data set MegaDepth (MD).

2.3. Pedestrian Detection. The deep feature-based detection approaches have become the main direction of pedestrian detection research because of their SOTA performance. In fact, many general detection methods are also suitable for

pedestrian detection, so we do not distinguish between them. We strongly recommend researchers to refer the surveys [35, 36], which give detailed summary about target detection. Here, we only review some classical algorithms. Ren et al. [37] propose a Recurrent Rolling Convolution (RRC) architecture, which can selectively integrate contextual information into the bounding box regressor. Liu et al. [12] propose SSD, the first deep network based object detector that does not resample pixels or features for bounding box hypotheses and is as accurate as approaches that do. Liu et al. [38] propose the ALFNet, which trains the SSD in multisteps and significantly improves the accuracy of pedestrian detection while maintaining the efficiency of the single-stage detector. Most detection methods are trained on datasets without occlusion or with reasonable occlusion. Once heavy occlusion occurs, the performance will decrease significantly. Therefore, the recent benchmark test pays special attention to pedestrian detection with heavy occlusion. Zhang et al. [39] design a new regression loss and introduce a part occlusion aware region of interest (PORoI) pooling unit to solve the problem of occluded pedestrian detection in crowded scenes. Tian et al. [40] design a set of component detectors; each component is designed to handle a specific occlusion mode. Zhou and Yuan [41] use a neural network to locate the full body and visible part of a pedestrian, respectively.

3. Method

The flowchart of the proposed multipedestrian tracking algorithm is shown in Figure 2. First camera motion is calculated to reduce the global error. Then, each KF predicts its trajectory at time k , denoted as $\tilde{\mathbf{S}}_k$, according to the former estimated trajectory \mathbf{S}_{k-1} . Next, a target detection network and a depth prediction network, respectively, extract the plane information and depth information of the target from the image at current time k , which constitute the measurement \mathbf{O}_k for KF. After that, the tracks and detected targets are matched with Hungarian method. There are three results for matching, and each result corresponds to a processing way. (1) For unmatched target, if a target does not match any track, the target will be treated as the starting point of a new track and a new KF will be initialized. (2) Unmatched track, if a track has not matched any target for several consecutive frames, the track is considered to have lost the target completely, and keep tracking is meaningless, so the track will be delete from tracking system. Otherwise, we treat the track as a temporary disappearance and update KF. (3) For matched track, if a track matches a target, the system will detect occlusion and then update KF. Finally, the added new tracks and the updated tracks compose the tracking results at time k . In the following subsections, we will elaborate on several key points in the algorithm.

3.1. State Estimation

3.1.1. KF Prediction. Kalman filter is a classic time series estimation method, which is very suitable for estimating the motion state of the target. The state can be any parameter related to the target, such as center position, height, aspect ratio, and their respective velocities in image coordinates.

According to the state of the target at the previous time, KF calculates a prediction value at time k by $\tilde{\mathbf{S}}_k = g(\mathbf{S}_{k-1}, \mathbf{Q})$, where \mathbf{Q} is a Gaussian noise that reflects the accuracy of the process model and g is a function that describes the state change law of the target in two adjacent frames. In multitarget tracking, g is usually assumed to be a constant noise model.

3.1.2. Observation Calculation. A target detection network and a depth prediction network, respectively, process the image at time k and obtain the plan and depth map of the targets, which constitute the measurement value \mathbf{O}_k . We adopt fast-YOLO [14] for target detection, which greatly reduces the number of deep inferences and speeds up the detection. For depth prediction, we follow the monocular depth estimation method open-sourced in paper [5]. To reduce the calculation, we use the trick of knowledge distilling to transfer the original depth prediction network to a 5-layer student CNN. The teacher network is trained on MD dataset [5]. Then, the student network is further trained to predict semantic segmentation maps from depth image. Eventually, the accuracy of the student network can be close to that of the teacher network, but the model size is much smaller, thus achieving accelerated computation. Both the two networks are trained offline and are not updated online when tracking.

3.1.3. KF Update. The final state value \mathbf{S}_k is composed of two parts: the predicted value and the corrected term,

$$\mathbf{S}_k = \tilde{\mathbf{S}}_k + K(\mathbf{O}_k - \mathbf{T}\tilde{\mathbf{S}}_k), \quad (1)$$

where K is the Kalman gain and \mathbf{T} is the state transition matrix. For multitarget tracking, we match tracks with targets and get three results: a target does not match any track, a track matches a target, and a track does not match any target. The first case indicates that the target emerges lately, so a new KF needs to be instantiated. Both the second and third cases can be calculated by Equation (1), which will be detailed in Section 3.3.

3.2. Constructing 3D Motion Model. Equipped with depth estimates, we construct a 3D linear motion model with a constant velocity assumption. The state of each target is modelled on the ten dimensional state space $(X, Y, Z, A, H, \dot{X}, \dot{Y}, \dot{Z}, \dot{A}, \dot{H})^T$ that contains the bounding box center position (X, Y, Z) , aspect ratio A , height H , and their respective velocities. The observations of the target are (X, Y, Z, A, H) .

The state change of the target at two adjacent moments can be denoted by the state equation:

$$\mathbf{S}_t = \mathbf{S}_{t-1} + \dot{\mathbf{S}}_{t-1} + \mathbf{q}_t, \quad (2)$$

where $\mathbf{q}_t \in N(0, \mathbf{Q})$ is process noise and \mathbf{Q} is the process noise covariance matrix. And we can derive the following formulas for X ,

$$X_t = X_{t-1} + \dot{X}_{t-1} + r_X. \quad (3)$$

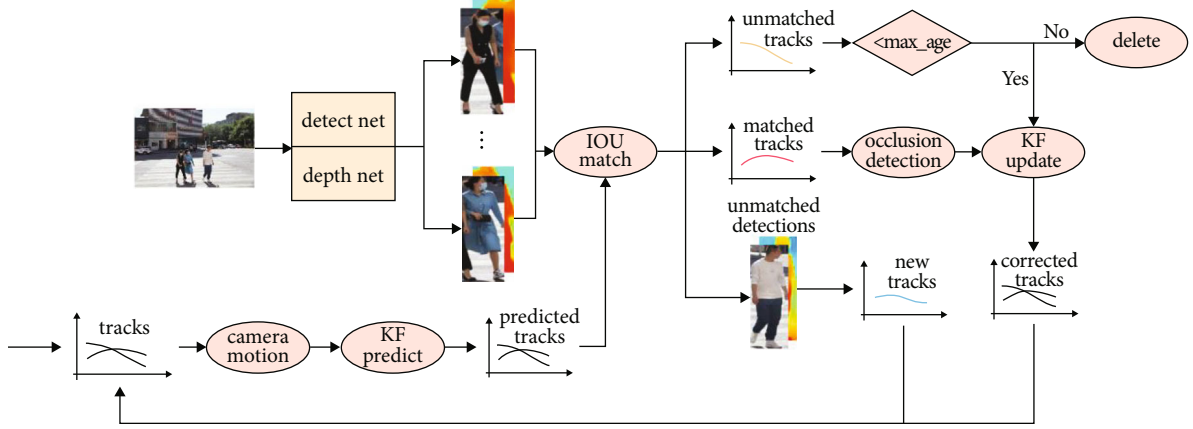


FIGURE 2: The flowchart of our tracking method.

Similarly, we can get the equations of Y, Z, A , and H . The intrinsic relationship between state \mathbf{S}_t and observation \mathbf{O}_t can be denoted by the measurement equation:

$$\mathbf{O}_{t-1} = \mathbf{S}_{t-1} + \mathbf{r}_{t-1}, \quad (4)$$

where $\mathbf{r}_t \in N(0, \mathbf{R})$ is observation noise and \mathbf{R} is the observation noise covariance matrix.

Inverse depth is a commonly used representation predicted due to the ability to represent points at infinity and to model uncertainty in pixel disparity space. So the depth used is $1/Z$. In order to simplify notation, we assume the camera focal length f is a constant. In fact, f can be folded into a motion noise parameter and can be easily tuned on a training set. Then, the adjusted parameters of the bounding boxes are as follows:

$$x = f \frac{X}{Z}, y = f \frac{Y}{Z}, z = \frac{1}{Z}, a = A, h = f \frac{H}{Z}. \quad (5)$$

This means that we dynamically scale the object with inverse depth. If depths are smooth over time, we can take Z_{t-1} as an approximation of Z_t , so we derive the following formulas from Equation (3),

$$f \frac{X_t}{Z_t} \approx f \frac{X_t}{Z_{t-1}} = f \frac{X_{t-1}}{Z_{t-1}} + f \frac{\dot{X}_{t-1}}{Z_{t-1}} + f \frac{r_X}{Z_{t-1}}, \quad (6)$$

which is

$$x_t = x_{t-1} + \dot{x}_{t-1} + f \frac{r_X}{Z_{t-1}}. \quad (7)$$

The equation suggests that one can approximately apply a Kalman filter on 2D image measurements augmented with a temporal noise model that is scaled by the estimated inverse-depth of the object.

3.3. Occlusion and Unmatched Tracks. Due to the use of depth information, we can easily determine whether the target is occluded by comparing the depth values in $\tilde{\mathbf{S}}_k$ and \mathbf{O}_k ,

because objects with smaller depth are always in front of objects with larger depth. To avoid accidental errors, we take the average depth of all points near the predicted location as the observed depth. The area size is $1/4$ of the predicted bounding box. When $z_k^{(\tilde{S})} < z_k^{(O)}$, we consider that the target is occluded, and \mathbf{O}_k is the information of the occluder rather than the target, so \mathbf{S}_k cannot be directly calculated by Equation (1). In this case, we take the KF prediction $\tilde{\mathbf{S}}_k$ as the approximate state value \mathbf{S}_k , which means the observation value \mathbf{O}_k is completely accurate without error.

If a track does not match any target, it may be because the target moves out of the image, not be detected, or not matched. To reduce the impact due to missed detections or matching errors, we introduce a counter for each track to count the number of frames since the last successful measurement association, denoted as c_k . When the track matches a target, the counter is reset to 0. When the track does not match any target, KF continues to predict the state of the target while c_k increases. If the track does not match any target for several consecutive frames c_{k_max} , there is a high probability that the target is lost, and the tracking of this track should be ended. It can be noticed that measurement association also works as identification. In most MOT methods, once a target is lost, its track will break off. So, when a new target appears, we have to judge whether the target is a new one or a lost one through an extra reidentification step. However, benefit from KF and occlusion detection, the track will not be interrupted in our method, so the reidentification is omitted.

In short, there are two ways to update \mathbf{S}_k according to different situations,

$$\mathbf{S}_k = \begin{cases} \tilde{\mathbf{S}}_k & z_k^{(\tilde{S})} < z_k^{(O)} \text{ or } c_k < c_{k_max} \\ \tilde{\mathbf{S}}_k + K(\mathbf{O}_k - \mathbf{T}\tilde{\mathbf{S}}_k) & \text{others} \end{cases}. \quad (8)$$

In our experiment, we set $c_{k_max} = 26$.


```

Input: The state vector  $\mathbf{S}_{k-1}$ , image  $\mathbf{img}_k$ , max unmatched consecutive frames  $c_{k-\max}$ .
Output: The state vector  $\mathbf{S}_k$ 
1: estimate camera motion;
2: calculate the predicted state vector  $\tilde{\mathbf{S}}_k$ ;
3: detect all pedestrian on  $\mathbf{img}_k$ ;
4: predict the depth of  $\mathbf{img}_k$ ;
5: match each track  $\mathbf{track}_i$  with targets  $\mathbf{target}_j$ ;
6: if  $\mathbf{target}_j$  does not match any  $\mathbf{track}_i$ 
    go step 7;
    elseif  $\mathbf{track}_i$  does not match any  $\mathbf{target}_j$ 
        count number of unmatched frames  $c_k$ , and go step 8
    else
        go step 9
7: initialize a new KF tracker with  $\mathbf{target}_j$ ;
8: if  $c_k < c_{k-\max}$ 
     $\mathbf{S}_k = \tilde{\mathbf{S}}_k$  else
    delete  $\mathbf{target}_j$ 
9: if  $z_k^{(\tilde{\mathbf{S}})} < z_k^{(O)}$ 
     $\mathbf{S}_k = \tilde{\mathbf{S}}_k$  else
     $\mathbf{S}_k = \tilde{\mathbf{S}}_k + K(\mathbf{O}_k - T\tilde{\mathbf{S}}_k)$ 
10: update KF

```

ALGORITHM 1: The overall algorithm for the proposed tracking method.

TABLE 1: Pedestrian detection results of different detectors on MOT17.

Detector	AP \uparrow	MODA \uparrow	MODP \uparrow	FAF \downarrow	TP \uparrow	FP \downarrow	FN \downarrow	Hz \uparrow
SDP [47]	0.81	76.9	78.0	1.3	95699	7599	18865	0.6
FRCNN [15]	0.72	68.5	78.0	1.7	88601	10081	25963	5.1
DPM [48]	0.61	31.2	75.8	7.1	78007	42308	36557	19.7
Fast YOLO [49]	0.77	73.1	77.8	1.5	92436	8694	20185	147.2

3.4. Camera Motion. Camera motion is an important factor to vision tracking, which not only changes the coordinates of the object but may also blur the image. The motion of dynamic objects is assumed to be small relative to the scene motion in most videos, so we use image alignment algorithm to approximate camera motion estimation. Philipp et al. proposed a practical work [42]. We first estimating a nonlinear pixel warp \mathbf{W} between neighbouring frames which maps pixel coordinates (x_{t-1}, y_{t-1}) in one frame to the next (x_t, y_t) and then use this wrap to align boxes forecasted using frames up to $t-1$ with frame t .

At last, we summarize the algorithm proposed in this article in pseudo-code as follows (Algorithm 1).

4. Results and Discussion

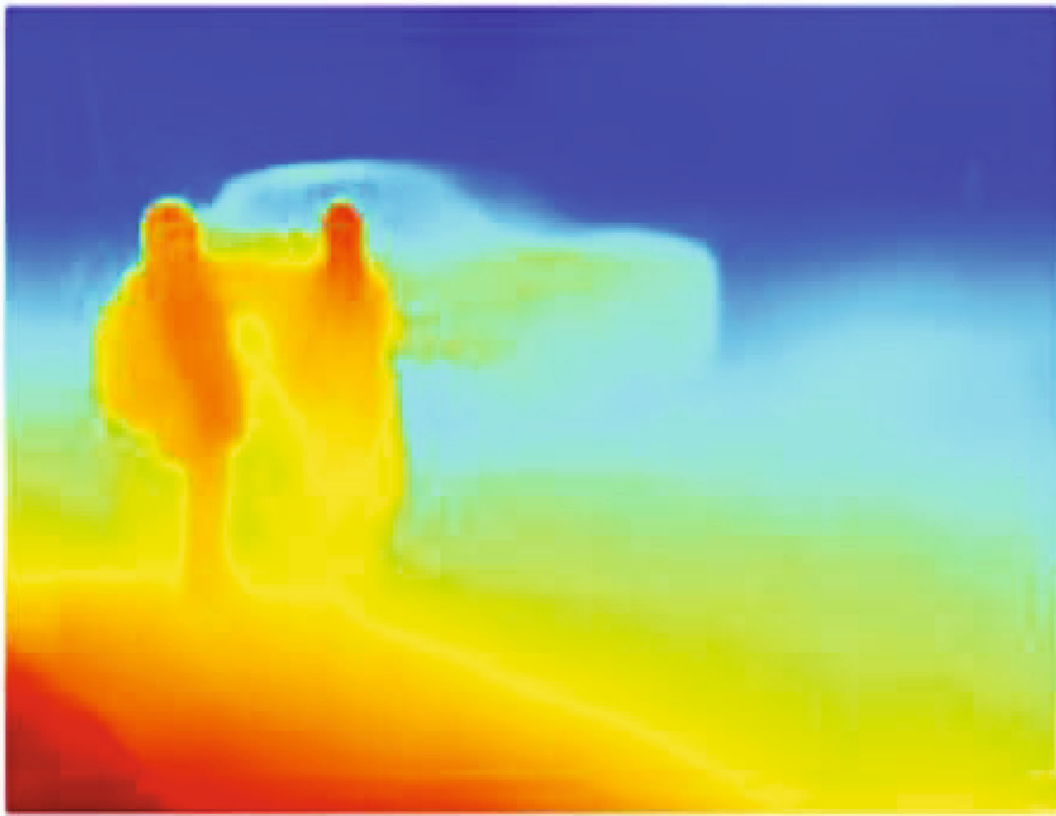
We conducted tracking experiments on the popular multi-target tracking datasets, and the results will be shown in Section 4.3. In addition, we also discuss the effect of depth prediction in Section 4.2.

4.1. Experiment Setting. We evaluated our method on two popular MOT datasets: MOT2017 [43] and MOT2020 [44]. MOT17 contains 14 videos, 7 for training and 7 for

testing. Faster R-CNN [10], SDP [45], and DPM [8] are provided as public target detector. MOT20 contains 8 videos, 4 for training and 4 for testing, and only Faster R-CNN [10] is provided. All of the datasets are very challenging including crowded scenes with heavy occlusions, camera motion, and both day and night sequences.

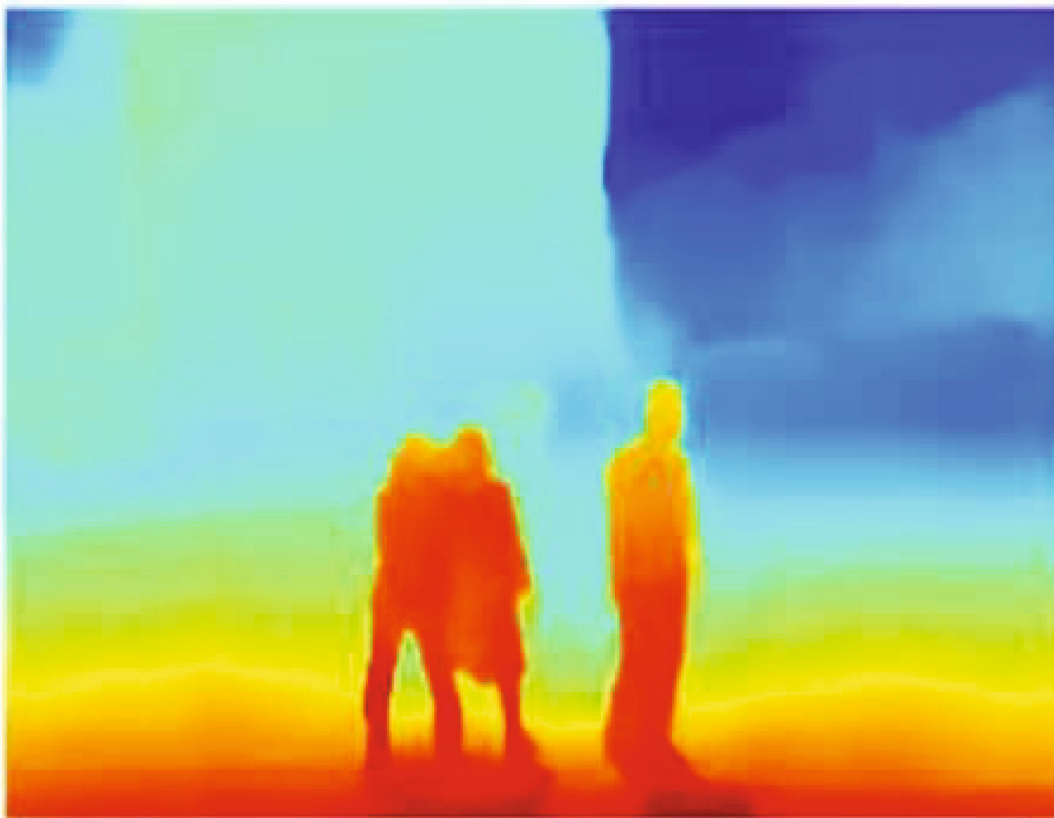
To evaluate the performance of the tracking methods, we adopted the widely used CLEAR MOT metrics [46]. Multiple object tracking accuracy (MOTA) evaluates accuracy in the presence of false positives (FP), false negatives (FN), and identity switches (IDS). IDS counts the total number of identity switches. At the same time, IDF1, MT, ML, and Hz are also considered. IDF1 is the ratio of correctly identified detections over the average number of ground-truth and computed detections, and it indicates the average maximum consistent tracking rate. MT evaluates the mostly tracked trajectories that are successfully tracked at least 80%. ML evaluates the mostly lost trajectories that are successfully tracked at most 20%. Hz indicates the processing speed (in frames per second). Among these metrics, MOTA and IDF1 are usually considered the most important.

For target detection, public detectors provided by the authors and private detector (Fast YOLO [14]) are all used for sufficient comparison. The private detector is trained



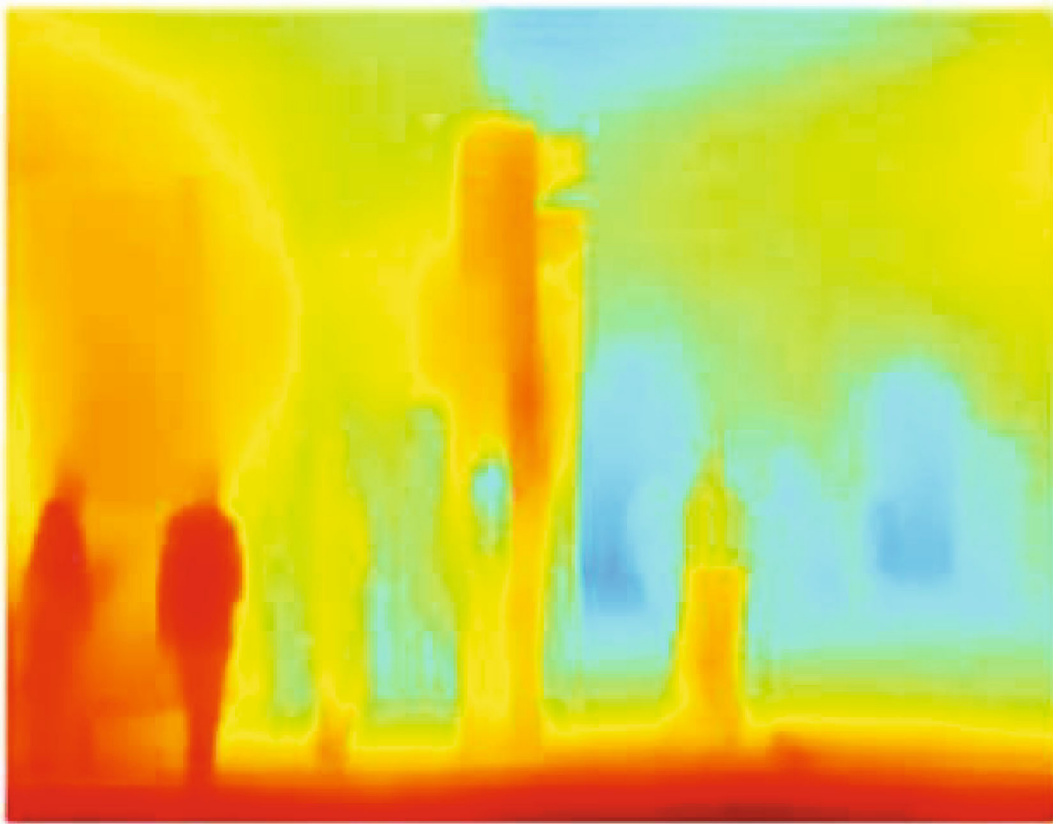
(a)

FIGURE 3: Continued.



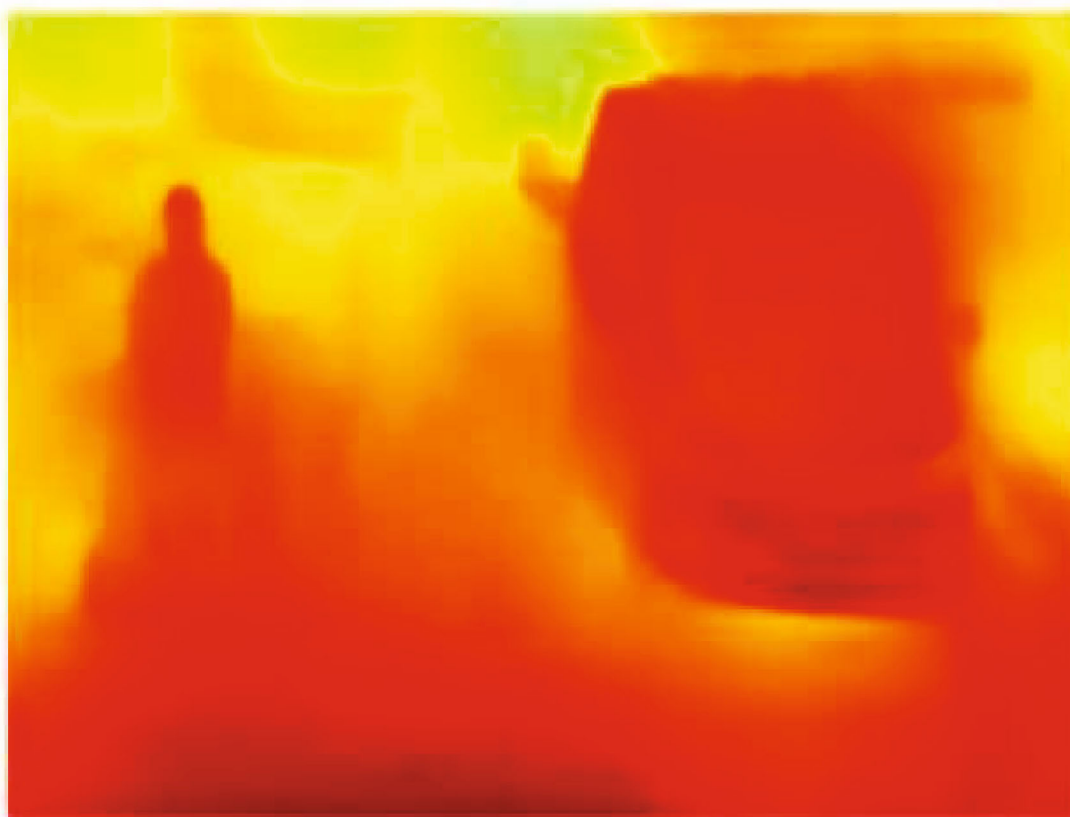
(b)

FIGURE 3: Continued.



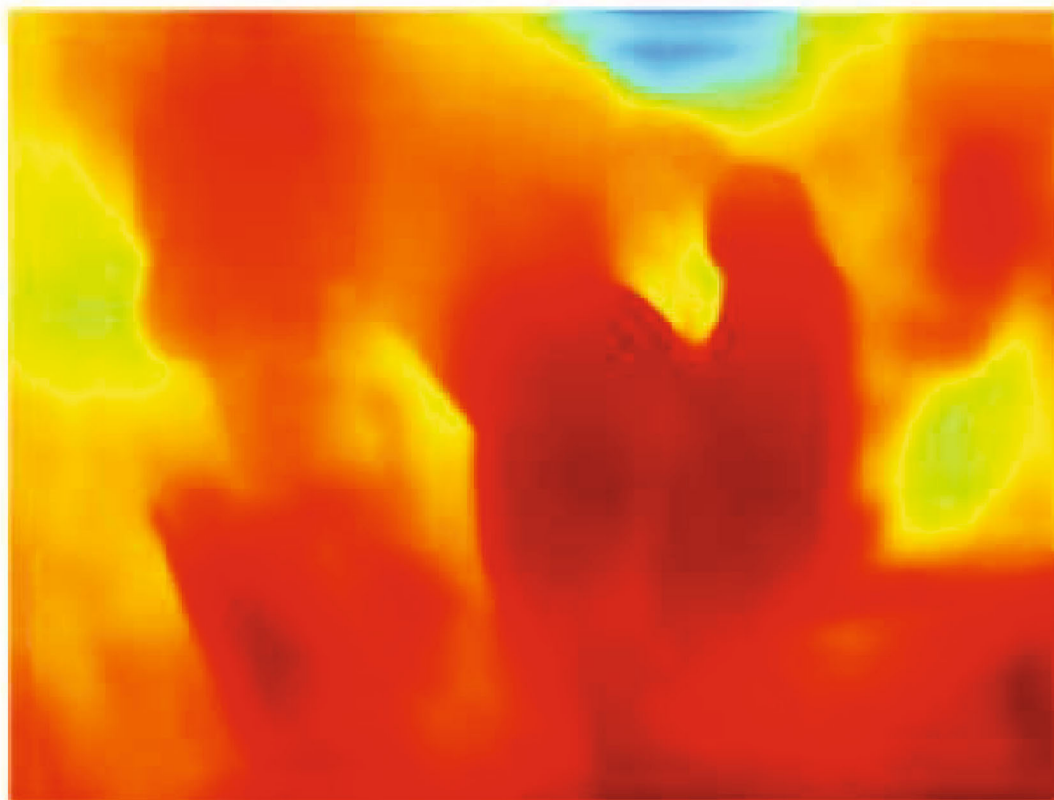
(c)

FIGURE 3: Continued.



(d)

FIGURE 3: Continued.



(e)

FIGURE 3: Result of pedestrian detection and depth prediction. The upper and under figures are the results of pedestrian detection and depth prediction, respectively.

TABLE 2: Tracking results using public detectors on MOT17.

Tracker	MOTA↑	IDF1↑	MT↑	ML↓	FP↓	FN↓	IDS↓	Hz↑
IOU [47]	45.5	39.4	15.7	40.5	19993	281643	5988	1523.0
SORT [15]	43.1	39.8	12.5	42.3	28398	287582	4852	143.3
GMPHD_Rd17 [48]	46.8	54.1	19.7	33.3	38452	257678	3865	30.8
FlowTracker [49]	40.4	38.0	14.0	36.6	60962	269136	5927	61.2
ours_pub	51.3	55.2	20.8	32.5	26236	247437	3734	40.4

TABLE 3: Tracking results using public detectors on MOT20.

Tracker	MOTA↑	IDF1↑	MT↑	ML↓	FP↓	FN↓	IDS↓	Hz↑
IOU_KMM [50]	46.5	49.4	29.9	19.6	57517	214777	4509	30.3
SORT [15]	42.4	45.1	16.7	26.2	27521	264694	4470	57.3
GMPHD_Rd20 [48]	44.7	43.5	23.6	22.1	42778	236116	7492	25.2
FlowTracker [49]	46.7	42.4	27.8	20.0	54732	217371	3532	19.2
ours_pub	52.6	53.8	29.3	19.5	26891	214278	3619	25.3

with the 7 training videos in MOT17. All models are retrained in advance and not updated when running. Our experiment was conducted in PyTorch and runs on a desktop with a CPU of Intel(R) Xeon(R) E5-2680@2.80GHz and a 1080Ti GPU.

4.2. Pedestrian Detection and Depth Prediction

4.2.1. Pedestrian Detection. To evaluate the effects of pedestrian detection, we tested Faster R-CNN [10], SDP [45], DPM [8], and Fast YOLO [14] on MOT17. Since there are many well-trained network models, we just fine-tuned them on the train set and then detect the pedestrians on the test set. Table 1 shows the result. The up arrow and down arrow, respectively, indicate that the larger the value, the better and the smaller the better, and the italic value indicates the best result. AP, MODA, MODP, and FAF mean average precision taken over a set of reference recall values, multiobject detection accuracy, multiobject detection precision, and the average number of false alarms per frame, respectively. As we can see, SDP has the best detection performance in most metrics; however, the speed is as slow as 0.6 FPS. Although Fast YOLO gets the second-best detection results, it achieves amazing speed, which is 245 times faster than SDP. There is no doubt that it is worthwhile to trade extremely small detection accuracy for extremely high detection speed for target tracking problem.

4.2.2. Depth Prediction. To evaluate the effects of depth prediction, we selected many pictures with different backgrounds for experiments, and five representative results are shown in Figure 3. It can be seen that pedestrians on all images are correctly detected. The effect of depth prediction is satisfactory, which can distinguish people from background. After careful observation, we found that depth prediction has the following two characteristics: One is that the estimation of the big target is more accurate. If the size of the target is small, the depth difference between the target and the background is smaller, and the target will not be well-marked on the depth map. As shown in Figure 3(d), the

depth of the car on the right is similar to that of the person on the left, but the outline of the car is obviously clearer. The other is that the estimation is more accurate under a simple background. It can be seen that the depth distinction between pedestrians and background is obvious in Figures 3(a) and 3(b), while in the depth maps in Figures 3(c)–3(e), some people and background are difficult to distinguish. In Figure 3(a), it is easy to see the difference in depth between the pedestrians and the car, which reflects the advantage of using depth to judge the occlusion. In Figure 3(b), the background can be divided into three parts, the ground, the building on the left rear, and the trees on the right rear. Because the building and trees are far away from people, they have little influence on the depth prediction of the pedestrians, so depth image is also outstanding. As a comparison, we can be seen that the depth of the people nearby in Figure 3(c) is obvious, while the people faraway are difficult to distinguish from the background. Interestingly, in the depth map of Figure 3(d), the pedestrian in the middle seems to have blended into the background, while the rear car is vaguely discernible, which seems to be related to the contrast of pixels. In addition, the words on the image will also have a significant impact. In the depth map in Figure 3(e), the silhouette of human is not clear, probably because the interference from the ground is very serious. Although sometimes the result of depth estimation is not very accurate, it can help us judge occlusion as long as it can be distinguished from the background depth.

4.3. Pedestrian Tracking

4.3.1. Results on MOT17 and MOT20. Our goal is to build a real-time tracking system, so the algorithms that we select from MOT17 and MOT20 for comparison all reach a speed of 25 frames per second. Meanwhile, all these algorithms have been published in papers.

Table 2 shows the results on MOT17. IOU [47], SORT [15], GMPHD_Rd [48], and FlowTracker [49] are selected as the baseline, because most of them also appear in MOT20,

TABLE 4: Tracking results of all real-time trackers on MOT17.

Tracker	MOTA \uparrow	IDF1 \uparrow	MT \uparrow	ML \downarrow	FP \downarrow	FN \downarrow	IDS \downarrow	Hz \uparrow	Det.
TrTrack [51]	75.2	63.5	55.3	10.2	50157	86442	3603	59.2	1
Fair [52]	73.7	72.3	43.2	17.3	27507	117477	3303	25.9	1
RekTCL [53]	73.3	73.2	41.3	18.7	22944	124980	2790	88.8	1
TraDeS [54]	69.1	63.9	36.4	21.5	20892	150060	3555	66.9	1
ours_pri	65.1	64.8	34.2	25.8	22065	171873	2976	52.6	1
ours_pub	51.3	55.2	20.8	32.5	26236	247437	3734	40.4	0
GMPHDOGGM [55]	49.9	47.1	19.7	38.0	24024	255277	3125	30.7	0
GMPHD_Rd17 [48]	46.8	54.1	19.7	33.3	38452	257678	3865	30.8	0
PHD_LMP [56]	45.9	42.5	15.5	37.9	27946	272196	4977	29.4	0
IOU17 [47]	45.5	39.4	15.7	40.5	19993	281643	5988	1522.9	0
SORT [15]	43.1	39.8	12.5	42.3	28398	287582	4852	143.3	0
FlowTracker [49]	40.4	38.0	14.0	36.6	60962	269136	5927	61.2	0

and this is helpful for comparison. All trackers use the same public detectors (Faster R-CNN [10], SDP [45], and DPM [8]). We can see that our method achieves the best performance on MOTA, IDF1, MT, ML, FN, and IDS. IOU [47] gets the smallest FP score and amazing tracking speed, and our method gets the second smallest FP score and the third fastest speed. IDS scores largely reflect the ability of continuous tracking. Our method uses different strategies for different matching results, so we obtain better IDS score.

The results on MOT20 are shown in Table 3. Compared with MOT17, MOT20 contains less frames and less trajectories. Due to more crowdedness and more pedestrians, the detection task is much more challenging. On the whole, tracking is significantly slower, and MT and ML are better, but there is no clear trend in the other metrics. Our method gets the best scores in most metrics except for MT, IDS, and Hz.

Next, we compared our method and more published real-time trackers on MOT17. Ten trackers are selected, including 4 trackers using private detectors (TrTrack [51], Fair [52], RekTCL [53], TraDeS [54]) and 6 trackers using public detectors (GMPHDOGGM [55], GMPHD_Rd [48], PHD_LMP [56], IOU17 [47], SORT [15], FlowTracker [49]). Table 4 shows the detail results. Benefit from better detectors and better training strategies, trackers using private detectors achieve high MOTA scores than trackers using public detectors. While using public detectors, ours_pub performs best. Although ours_pri ranks last in the trackers using private detectors, it has a relatively high speed. Because an additional network is introduced to calculate the depth, the speed of our method has to be encumbered. In the future, we plan to use light architectures to optimize the processing of depth prediction. Better detection and association methods are also considered to improve accuracy.

We demonstrate the relationship between tracker accuracy and speed in Figure 4 (excluding IOU17). The farther to the right, the higher the MOTA score, and the higher the upward, the faster the tracking speed. It can be clearly seen in the figure that the methods using private detectors have great advantages in tracking accuracy and not slow in

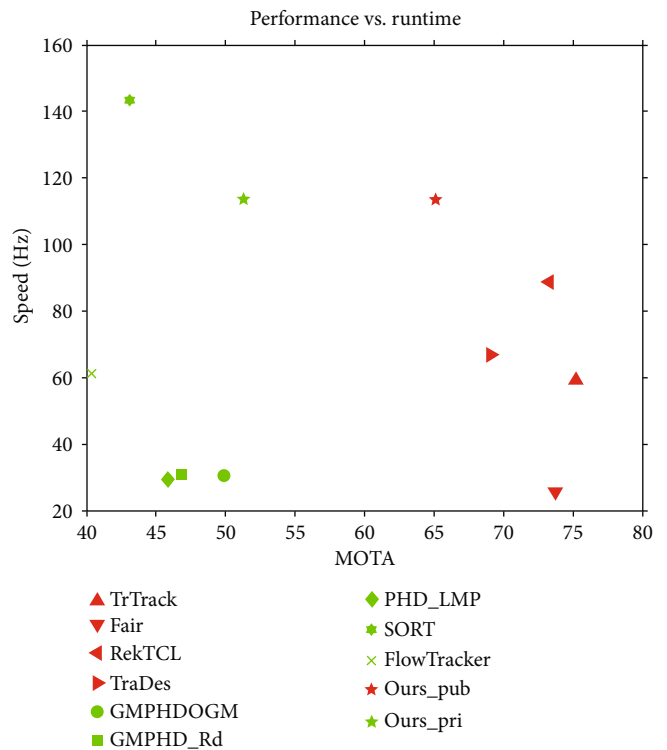


FIGURE 4: The performance compare between different trackers. Each marker denotes a tracker accuracy and speed. The red and green markers represent private and public detector, respectively.

speed. Our method is in the middle in speed and accuracy, and there is still much room for improvement.

4.3.2. Impact of Different Detectors. To explore the influence of different detectors on multiobject tracking, we ran our tracker with public and private detectors on MOT17. The results are shown in Table 5. Obviously, using better detector can improve the performance of the tracking algorithm, which is mainly reflected in three aspects: the total number of detected targets, the number of correctly recognized targets, and the detection precision. The MOTA scores from high to low are ours_pri, SDP, FrRCNN, and DPM, and it

TABLE 5: Tracking results of different detectors on MOT17.

Detector	MOTA \uparrow	IDF1 \uparrow	MT \uparrow	ML \downarrow	FP \downarrow	FN \downarrow	IDS \downarrow	Hz \uparrow
SDP	51.8	56.1	9.9	9.6	2938	68995	1096	43.2
FrRCNN	43.6	47.8	6.7	10.4	5725	84076	956	9.1
DPM	35.8	39.4	4.2	12.5	18573	94736	1682	48.4
ours_pri	65.4	64.8	11.4	8.6	7355	57291	992	52.6

TABLE 6: Tracking results of different depth predictors on MOT17.

Detector	MOTA \uparrow	IDF1 \uparrow	MT \uparrow	ML \downarrow	FP \downarrow	FN \downarrow	IDS \downarrow	Hz \uparrow
HG	62.8	61.1	12.2	9.5	8573	64736	1082	17.1
Distilled-HG	65.4	64.8	11.4	8.6	7355	57291	992	52.6

directly reflects the performance of the detector. The private detector performs very well, especially for FN, which is 11704 lower than the best public detector. The tracking speed is not only related to the detection speed but also related to the number of targets detected, because the more the number of targets to be tracked, the greater the amount of calculation required to match the trajectory to the target. The private detector achieves 52.6 Hz, which is much higher than the public detectors.

4.3.3. Impact of Distillation. We compare the tracking performance with and without distilling the HG network, and the result is shown in Table 6. Accuracy of tracking improves as that of depth prediction improves after distillation. The most attractive change is speed. The tracking speed before distillation was only 17.1 Hz, and after distillation, it reached 52.6 Hz, increasing by 2.08 times.

4.3.4. Impact of Different Components. To get a deeper insight into our method, we run several experiments to test the effect of each component. The complete method with private detector is set as the baseline. We remove one component each time and then evaluate the method on MOT17. The result is shown in Table 7.

The function of KF is mainly reflected in two aspects: one is to predict the target search area, and the other is to smooth the trajectory. Since the detector is not updated online, its performance remains stable. But KF can affect the detection result by change the target search area. Therefore, the indicator related to accuracy (IDF1) has not changed much. When the occlusion or unmatched trajectory occurs, KF can continue to estimate the trajectory, effectively avoiding the interruption of the trajectory, so the impact on IDS is very obvious. Overall, MOTA drops by 6.2, mainly due to the increase in IDS. Since the computation of KF is very small, it has little influence on the tracking speed.

Occlusion detection involves depth prediction and occlusion judgment. Occlusion judgment will affect the processing method after IOU matching, so it has a greater impact on IDS but has almost no impact on IDF1. Since the depth prediction network is time-consuming, the tracking speed increases by 26.8 frame per second without occlu-

TABLE 7: Result of ablation study on MOT17.

Components	MOTA \uparrow	IDF1 \uparrow	IDS \downarrow	Hz \uparrow
Complete	65.1	64.8	2976	52.6
No KF	-6.2	-1.4	+127	+4.4
No occlusion detection	-4.5	-0.1	+54	+26.8
No motion estimation	-5.9	-1.1	+46	+11.3

sion detection, which is the largest increase amplitude in the three components.

The position of the target in the image is determined by the motion of the target and the camera. When using Kalman filtering, we assume that the camera is fixed. In fact, in the MOT17 dataset, three camera sequences are stationary, and four are captured from a moving camera. If we ignore the motion of the camera, it will inevitably affect the Kalman filter. It can be seen from Table 7 that motion estimation has an obvious impact on the overall performance. MOTA dropped by 5.9, IDF1 dropped by 1.1, IDS increased by 46, and tracking speed increased by 11.3.

Referring the results in Table 7 by column, we can get the impact of each component on a single indicator. Figure 5 shows this impact intuitively. KF, occlusion detection, and motion estimation all contribute to MOTA, IDF1, and IDS. For MOTA and IDF1, the effect of KF and motion estimation is slightly greater than that of occlusion detection, because KF and motion estimation have a greater impact on the target detection. For IDS, KF is particularly useful. This is due to the processing of missing values by KF, which results in a significant decrease in the number of reidentified targets. More components mean more computation, so the tracking speed inevitably decreases. Among them, occlusion detection has the greatest impact on the speed, because it calculates the depth through a deep neural network. Accuracy and speed are often difficult to compromise. Fortunately, our method improves accuracy while keeps high speed.

5. Discussion

In Section 4.2, we evaluate the performance of depth prediction. Experiment shows that depth prediction is more

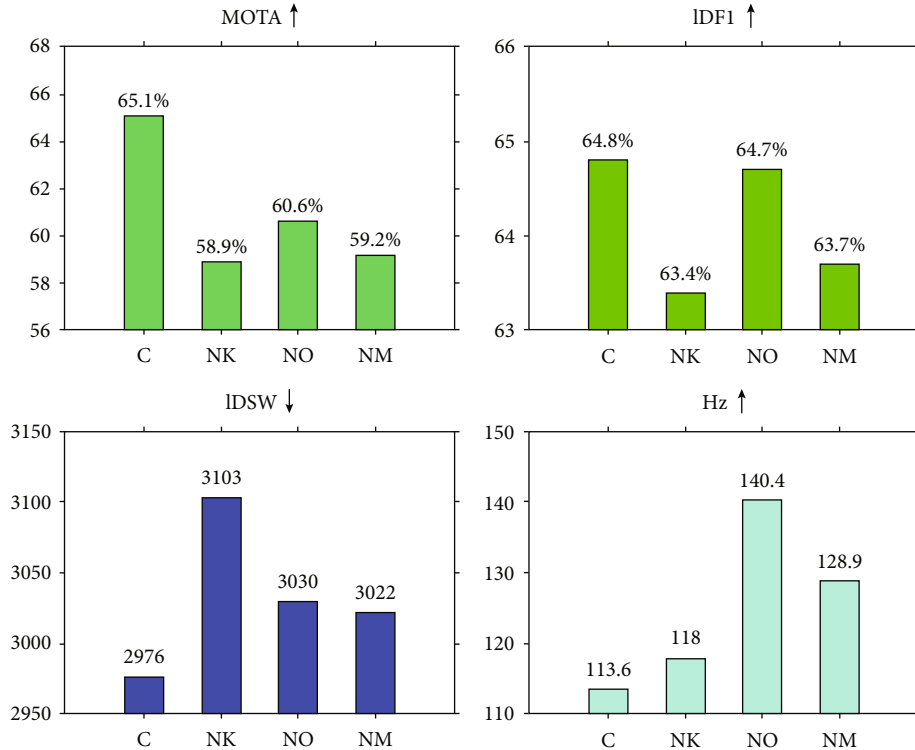


FIGURE 5: Impact of each component on indicators. C, NK, NO, and NM denote complete, no KF, no occlusion detection, and no motion estimation, respectively.

effective when the target is relatively large, or when it differs greatly from the background. Except for very few cases, the depth prediction is reliable. Although the estimated depth values cannot replace the true depth values, they completely reflect the sequential order of the objects, which is enough for us to judge the spatial location of the targets.

In Section 4.3, we evaluate our method with both public and private detectors on MOT17 and MOT20. Our method has achieved high accuracy while maintaining a high speed. It should be pointed out that the methods we selected are all real-time methods. In fact, more methods trade speed for accuracy. Through the ablation study, we show the impact of each component on the tracking results. KF framework has the greatest impact on IDS, as it is the key to maintaining continuous tracking. Occlusion detection plays the most important role in IDF1, because it determines the update method of KF, which in turn affects the final accuracy. For tracking speed, depth estimation involves neural network calculations and has the greatest impact on speed, while KF has relatively small impact. On the whole, each component has similar contributions to MOTA.

6. Conclusions

In this paper, we study pedestrian detection and tracking with a fixed monocular camera. To improve the robustness of tracking, we focus on continuous tracking and

occlusion detection to deal with heavily occlusion. To keep continuous tracking, we introduce Kalman filter to estimate the motion state of the target in each frame. With the help of KF's predictive ability, we can continue to estimate the state of the target when the target is occluded. Once the target reappears, the tracker can quickly locate the target, thus avoiding reidentification and ensuring the continuity of the target's trace during the occlusion period. To detect occlusion, we introduce depth information in our tracking system. Once the depth of the target suddenly decreases a lot, we think the target is obscured by something, because the closer the distance to the camera the smaller the depth. Depth information is usually obtained through depth sensors or multiview images, which is not practical in daily life, because most cameras are monocular and cannot provide depth value. Here, we introduce a light depth prediction network, which is distilled from a large but well-performed network. Distillation can not only slightly improve the accuracy of depth prediction (2.6 MOTA score) but also greatly accelerate the speed of model inference (35.5 Hz). We evaluated our method on public object tracking dataset. The results show that our method can not only achieve high accuracy (65.1 MOTA score) on MOT 17 but also high tracking speed (52.6 Hz). In the future, we intend to introduce better association rules to improve the tracking accuracy, continuously optimize the target detection network and depth estimation network, and strive to achieve real-time tracking in a more universal environment.

Data Availability

The data used to support the findings of this study are included within the article. All datasets used in our research are publicly available and are cited in our article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Beijing Municipality (4212023), the Fundamental Research Funds for the Central Universities, USTB (FRF-GF-20-04A), the Major Program of National Social Science Foundation of China (No. 17ZDA331), the National Key Research and Development Program of China (2018YFC2001700), the Scientific and Technological Innovation Foundation of Shunde Graduate School, and the Engineering Research Center of Intelligence Perception and Autonomous Control, Ministry of Education, Beijing (100124).

References

- [1] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, 2020.
- [2] S. James, P. Wohlhart, M. Kalakrishnan et al., "Sim-to-real via sim-to-sim: data-efficient robotic grasping via randomized-to-canonical adaptation networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12627–12637, Long Beach, CA, USA, 2019.
- [3] M. Gattullo, G. W. Scurati, M. Fiorentino, A. E. Uva, F. Ferrise, and M. Bordegoni, "Towards augmented reality manuals for industry 4.0: a methodology," *Robotics and Computer-Integrated Manufacturing*, vol. 56, pp. 276–286, 2019.
- [4] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 1995.
- [5] Z. Li and N. Snavely, "Megadepth: learning single-view depth prediction from internet photos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2041–2050, Salt Lake City, UT, USA, 2018.
- [6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*, Kauai, HI, USA, 2001.
- [7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pp. 886–893, San Diego, CA, USA, 2005.
- [8] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *2008 IEEE conference on computer vision and pattern recognition*, pp. 1–8, Anchorage, AK, USA, 2008.
- [9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 580–587, Columbus, OH, USA, 2014.
- [10] 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.
- [11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, 2016.
- [12] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 21–37, Cham, Springer, 2016.
- [13] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.
- [14] M. J. Shafiee, B. Chywl, F. Li, and A. Wong, "Fast YOLO: a fast you only look once system for real-time embedded object detection in video," 2017, <https://arxiv.org/abs/1709.05943>.
- [15] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*, pp. 3464–3468, Phoenix, AZ, USA, 2016.
- [16] V. K. Singh, B. Wu, and R. Nevatia, "Pedestrian tracking by associating tracklets using detection residuals," in *2008 IEEE workshop on motion and video computing*, pp. 1–8, Copper Mountain, CO, USA, 2008.
- [17] H. Shen, L. Huang, C. Huang, and W. Xu, "Tracklet association tracker: an end-to-end learning-based association approach for multi-object tracking," 2018, <https://arxiv.org/abs/1808.01562>.
- [18] H. Zhou, W. Ouyang, J. Cheng, X. Wang, and H. Li, "Deep continuous conditional random fields with asymmetric inter-object constraints for online multi-object tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1011–1022, 2019.
- [19] A. Dehghan, S. Modiri Assari, and M. Shah, "Gmmcp tracker: globally optimal generalized maximum multi clique problem for multiple object tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 4091–4099, Boston, MA, USA, 2015.
- [20] S. Tang, B. Andres, M. Andriluka, and B. Schiele, "Subgraph decomposition for multi-target tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5033–5041, Boston, MA, USA, 2015.
- [21] W. Feng, Z. Hu, W. Wu, J. Yan, and W. Ouyang, "Multi-object tracking with multiple cues and switcher-aware classification," 2019, <https://arxiv.org/abs/1901.06129>.
- [22] P. Voigtlaender, M. Krause, A. Osep et al., "MOTS: multi-object tracking and segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7942–7951, Long Beach, CA, USA, 2019.
- [23] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *Proceedings of the IEEE international conference on computer vision (ICCV)*, pp. 4696–4704, Santiago, Chile, 2015.
- [24] S. Tang, M. Andriluka, B. Andres, and B. Schiele, "Multiple people tracking by lifted multicut and person re-identification," in *Proceedings of the IEEE Conference on Computer*

- Vision and Pattern Recognition (CVPR)*, pp. 3539–3548, Honolulu, HI, USA, 2017.
- [25] Y. Xu, L. Qin, X. Liu, J. Xie, and S. C. Zhu, “A causal and-or graph model for visibility fluent reasoning in tracking interacting objects,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2178–2187, Salt Lake City, UT, USA, 2018.
- [26] L. Chen, H. Ai, Z. Zhuang, and C. Shang, “Real-time multiple people tracking with deeply learned candidate selection and person re-identification,” in *2018 IEEE international conference on multimedia and expo (ICME)*, pp. 1–6, San Diego, CA, USA, 2018.
- [27] W. Zhang, H. Zhou, S. Sun, Z. Wang, J. Shi, and C. C. Loy, “Robust multi-modality multi-object tracking,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2365–2374, Seoul, Korea (South), 2019.
- [28] S. Gautam, G. P. Meyer, C. Vallespi-Gonzalez, and B. C. Becker, “SDVTracker: real-time multi-sensor association and tracking for self-driving vehicles,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 3012–3021, Montreal, Canada, 2021.
- [29] F. Huan, M. Gong, C. Wang, K. Batmanghelich, and D. Tao, “Deep ordinal regression network for monocular depth estimation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2002–2011, Salt Lake City, UT, USA, 2018.
- [30] J. Jiao, Y. Cao, Y. Song, and R. Lau, “Look deeper into depth: monocular depth estimation with semantic booster and attention-driven loss,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 53–69, Munich, Germany, 2018.
- [31] R. Chen, F. Mahmood, A. Yuille, and N. J. Durr, “Rethinking monocular depth estimation with adversarial training,” 2018, <https://arxiv.org/abs/1808.07528>.
- [32] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from RGBD images,” in *Computer Vision – ECCV 2012. ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds., vol. 7576 of Lecture Notes in Computer Science, pp. 746–760, Springer, Berlin, Heidelberg, 2012.
- [33] A. Saxena, M. Sun, and A. Y. Ng, “Make3d: learning 3D scene structure from a single still image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 824–840, 2009.
- [34] M. Menze and A. Geiger, “Object scene flow for autonomous vehicles,” in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 3061–3070, Boston, MA, USA, 2015.
- [35] J. Cao, Y. Pang, J. Xie, F. S. Khan, and L. Shao, “From hand-crafted to deep features for pedestrian detection: a survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, p. 1, 2021.
- [36] Z. Zou, Z. Shi, Y. Guo, and J. Ye, “Object detection in 20 years: a survey,” 2019, <https://arxiv.org/abs/1905.05055>.
- [37] J. Ren, X. Chen, J. Liu et al., “Accurate single stage detector using recurrent rolling convolution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5420–5428, Honolulu, HI, USA, 2017.
- [38] W. Liu, S. Liao, W. Hu, X. Liang, and X. Chen, “Learning efficient single-stage pedestrian detectors by asymptotic localization fitting,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 618–634, Munich, Germany, 2018.
- [39] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, “Occlusion-aware R-CNN: detecting pedestrians in a crowd,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 637–653, Munich, Germany, 2018.
- [40] Y. Tian, P. Luo, X. Wang, and X. Tang, “Deep learning strong parts for pedestrian detection,” in *Proceedings of the IEEE international conference on computer vision (ICCV)*, pp. 1904–1912, Santiago, Chile, 2015.
- [41] C. Zhou and J. Yuan, “Bi-box regression for pedestrian detection and occlusion estimation,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 135–151, Munich, Germany, 2018.
- [42] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, “Tracking without bells and whistles,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 941–951, Seoul, Korea (South), 2019.
- [43] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, “MOT16: a benchmark for multi-object tracking,” 2016, <https://arxiv.org/abs/1603.00831>.
- [44] P. Dendorfer, H. Rezatofighi, A. Milan et al., “Mot20: a benchmark for multi object tracking in crowded scenes,” 2020, <https://arxiv.org/abs/2003.09003>.
- [45] F. Yang, W. Choi, and Y. Lin, “Exploit all the layers: fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers,” in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 2129–2137, Las Vegas, NV, USA, 2016.
- [46] K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: the clear mot metrics,” *EURASIP Journal on Image and Video Processing*, vol. 2008, Article ID 246309, 10 pages, 2008.
- [47] E. Bochinski, V. Eiselein, and T. Sikora, “High-speed tracking-by-detection without using image information,” in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, Lecce, Italy, 2017.
- [48] N. L. Baisa, “Occlusion-robust online multi-object visual tracking using a GM-PHD filter with CNN-based re-identification,” *Journal of Visual Communication and Image Representation*, vol. 80, article 103279, 2021.
- [49] H. Nishimura, S. Komorita, Y. Kawanishi, and H. Murase, “SDOF-Tracker: fast and accurate multiple human tracking by skipped-detection and optical-flow,” 2021, <https://arxiv.org/abs/2106.14259>.
- [50] O. Urbann, O. Bredtmann, M. Otten, J.-P. Richter, T. Bauer, and D. Zibriczky, “Online and real-time tracking in a surveillance scenario,” 2021, <https://arxiv.org/abs/2106.01153>.
- [51] P. Sun, Y. Jiang, R. Zhang et al., “Transtrack: multiple-object tracking with transformer,” 2020, <https://arxiv.org/abs/2012.15460>.
- [52] Y. Zhang, C. Wang, X. Wang, W. Zeng, and W. Liu, “Fairmot: on the fairness of detection and re-identification in multiple object tracking,” *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3069–3087, 2021.
- [53] W. Li, Y. Xiong, S. Yang, M. Xu, Y. Wang, and W. Xia, “Semi-TCL: semi-supervised track contrastive representation learning,” 2021, <https://arxiv.org/abs/2107.02396>.
- [54] J. Wu, J. Cao, L. Song, Y. Wang, M. Yang, and J. Yuan, “Track to detect and segment: an online multi-object tracker,” in

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12352–12361, 2021.

- [55] Y. Song, K. Yoon, Y. C. Yoon, K. C. Yow, and M. Jeon, “Online multi-object tracking with GMPHD filter and occlusion group management,” *IEEE Access*, vol. 7, pp. 165103–165121, 2019.
- [56] R. Sanchez-Matilla and A. Cavallaro, “Motion prediction for first-person vision multi-object tracking,” in *Computer Vision – ECCV 2020 Workshops. ECCV 2020*, A. Bartoli and A. Fusiello, Eds., vol. 12538 of Lecture Notes in Computer Science, pp. 485–499, Springer, Cham, 2020.

Research Article

Gene Selection and Classification of scRNA-seq Data Combining Information Gain Ratio and Genetic Algorithm with Dynamic Crossover

Junhong Feng ¹, Xishuan Niu ¹, Jie Zhang ¹, and Jian-Hong Wang ²

¹School of Computer Science and Engineering, Yulin Normal University, Yulin 537000, Guangxi, China

²Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung 411030, Taiwan

Correspondence should be addressed to Jie Zhang; jgxyzjz@126.com and Jian-Hong Wang; wwwccucomtw@gmail.com

Received 2 December 2021; Accepted 7 January 2022; Published 31 January 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 Junhong 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.

Single-cell RNA sequencing (scRNA-seq) is emerging as a promising technology. There exist a huge number of genes in a scRNA-seq data. However, some genes are high quality genes, and some are noises and irrelevant genes because of unspecific technology reasons. These noises and irrelevant genes may have a strong influence on downstream data analyses, such as a cell classification, gene function analysis, and cancer biomarker detection. Therefore, it is very significant to obviate these irrelevant genes and choose high quality genes by gene selection methods. In this study, a novel gene selection and classification method is presented by combining the information gain ratio and the genetic algorithm with dynamic crossover (abbreviated as IGRDCGA). The information gain ratio (IGR) is employed to eliminate irrelevant genes roughly and obtain a preliminary gene subset, and then the genetic algorithm with a dynamic crossover (DCGA) is utilized to choose high quality genes finely from the preliminary gene subset. The main difference between the IGRDCGA and the existing methods is that the DCGA and IGR are integrated first and used to select genes from scRNA-seq data. We conduct the IGRDCGA and several competing methods on some real-world scRNA-seq datasets. The obtained results demonstrate that the IGRDCGA can choose high quality genes effectively and efficiently and outperforms the other several competing methods in terms of both the dimensionality reduction and the classification accuracy.

1. Introduction

In scRNA-seq data, there often are amounts of genes and may reach tens of thousands. Some genes are irrelevant or unsuitable for classification tasks, and they may seriously affect the efficiency of downstream data analysis. If all genes are utilized in data classifications, the classification accuracy and classification efficiency may be low. In order to obviate these irrelevant genes and select high quality genes, an effective and efficient gene selection algorithm is vital.

Feature selection (FS) problems can be taken as large-scale global optimization problems [1]; therefore, we can use bioinspired intelligence optimization algorithms to address feature selection problems. Wang et al. [1] took FS problems regard as large-scale global optimization problems. Nakisa

et al. [2] utilized the evolutionary computation (EC) to search the optimal feature subset. Eroglu and Kilic [3] integrated a genetic local search algorithm and a k -nearest neighbor classifier to select feature subset. Maleki and Zeinali [4] used a hybrid genetic algorithm (GA) to address dimension reductions and applied it to the classification of lung cancer. Tahir et al. [5] presented a binary chaotic GA to select feature for healthcare datasets. However, how to correctly use the GA to address the gene selection and classification of scRNA-seq data is a significant issue to consider first. To the best of our knowledge, there are only a few literatures so far.

The study integrates the IGR and DCGA to address the gene selection and classification of scRNA-seq data and proposes a novel gene selection and classification algorithm

IGRDCGA. The IGRDCGA utilizes the IGR to eliminate irrelevant genes roughly and obtain a preliminary gene subset and then employs the DCGA to choose high quality genes finely from the preliminary gene subset.

The rest of this study is organized as follows. Section 2 briefly describes the information gain ratio and genetic algorithm. Section 3 states three evaluation metrics. The dataset and preprocessing method to use in the study are described in Sections 4. The coding and the other details of the IGRDCGA are described in Section 5. The numerical results of the IGRDCGA and several competing algorithms are given in Section 6. The conclusion of the study is made in Section 7.

2. Related Work

In this section, the information gain ratio and genetic algorithm are described as follows.

2.1. Information Gain Ratio. The information gain is a metric derived from information entropy, often used to evaluate the mutual dependence level between two random variables. Namely, it is a symmetrical metric of dependency [6].

For two discrete random variables X and Y , their information entropy can be calculated, respectively, in terms of the following formulas:

$$H(X) = - \sum_{x \in X} p(x) \log_2(p(x)), \quad (1)$$

$$H(Y) = - \sum_{y \in Y} p(y) \log_2(p(y)), \quad (2)$$

where $x \in X$, $y \in Y$, and $p(x)$, $p(y)$ represent the marginal probability of x and y , respectively.

The conditional entropy and information gain [6–8] of X versus Y can be calculated in terms of the first two following formulas, respectively. The information gain ratio of X versus Y is the ratio of the information gain to the information entropy, which is formulated in the last following formula.

$$H(X|Y) = - \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2(p(x|y)), \quad (3)$$

$$IG(X; Y) = H(X) - H(X|Y), \quad (4)$$

$$IGR(X; Y) = \frac{IG(X; Y)}{H(X)}. \quad (5)$$

The $IGR(X; Y)$ ranges from 0 to 1 while 1 represents that X completely leads to Y and 0 represents that X and Y are completely independent.

2.2. Genetic Algorithm. The genetic algorithm (GA) is a bioinspired intelligence optimization algorithm. It is inspired by the process of a natural selection and belongs to one of evolutionary algorithms (EAs) [9–13]. It is commonly

utilized to generate feasible solutions for optimization problems by performing the operators such as selection, crossover, and mutation [14–16]. The selection operator is designed to choose a part of chromosomes for crossover operator from previous population. The frequently used selection operator is random selection strategy, such as tournament selection strategy. The crossover operator is designed to exchange one or many genes in two parents that are selected by selection operator. It simulates reproduction or recombination in biological evolution process. GA determines whether to perform mutation operator or not according to crossover probability. By crossover operator, two parents may generate two or many offsprings in terms of different crossover strategy. The frequently used crossover operator includes single-point crossover, two-point crossover, and multipoint crossover. Mutation operator is designed to modify one or many genes in certain chromosome. It simulates gene mutation in biological evolution process. Similarly, GA determines whether to perform mutation operator or not according to mutation probability. The frequently used mutation operator includes locus mutation, exchange mutation, and insertion mutation. Mutation operator only acts on one chromosome while crossover operator acts on two chromosomes. Generally, crossover operator probability is much larger than mutation probability. This accords with a biological evolution process.

3. Evaluation Metrics

To evaluate the performance of the IGRDCGA, in the study, we utilize the following evaluation metrics: NMI (normalized mutual information) [17–19], ARI (adjusted random index) [20, 21], and *purity* [22].

3.1. NMI (Normalized Mutual Information). The NMI is a frequently used evaluation metric. It can be often used to evaluate the accuracy and the difference between the obtained clustering results and the ground truth results.

For two discrete random variables X and Y , their MI (mutual information) can be calculated in terms of the following formula:

$$MI(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}, \quad (6)$$

where $p(x)$ and $p(y)$, respectively, denote marginal probability of x and y and $p(x, y)$ denotes a joint distribution probability.

The normalized MI is taken as NMI, which can be calculated in terms of the following formula.

$$NMI(X; Y) = \frac{2 * MI(X; Y)}{H(X) + H(Y)}, \quad (7)$$

where $H(X)$ and $H(Y)$ denote the entropy of X and Y , respectively. The NMI ranges from 0 to 1 while 1 is the optimal score, which represents that X and Y have identical mutual information. A larger NMI signifies higher concordance between X and Y .

Example 1. Suppose that the ground truth class labels of a single-cell dataset are as follows: $y_1 = [1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 3]$ and the classification result of an algorithm is as follows: $y_2 = [1\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 3\ 1\ 1\ 3\ 3\ 3]$. It follows from y_1 and y_2 that their unique value is both $[1\ 2\ 3]$. Then, the probability values in formula (6) are calculated as follows:

$$p(y_1 = 1) = p(y_1 = 2) = \frac{6}{16} = 0.3529,$$

$$p(y_1 = 3) = \frac{5}{17} = 0.2941,$$

$$p(y_2 = 1) = \frac{8}{17} = 0.4706,$$

$$p(y_2 = 2) = \frac{5}{17} = 0.2941,$$

$$p(y_2 = 3) = \frac{4}{17} = 0.2353,$$

$$p(1, 1) = p(y_1 = 1, y_2 = 1) = \frac{5}{17} = 0.2941,$$

$$p(1, 2) = p(y_1 = 1, y_2 = 2) = \frac{1}{17} = 0.0588,$$

$$p(1, 3) = p(y_1 = 1, y_2 = 3) = \frac{0}{17} = 0,$$

$$p(2, 1) = 0.0588, p(2, 2) = 0.2353, p(2, 3) = 0.0588,$$

$$p(3, 1) = 0.1176, p(3, 2) = 0, p(3, 3) = 0.1765.$$

(8)

According to formula (6), it follows that

$$MI(y_1; y_2) = \sum_{x \in y_1} \sum_{y \in y_2} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} = 0.5654. \quad (9)$$

According to formula (1) or (2), it follows that

$$\begin{aligned} H(y_1) &= - \sum_{x \in y_1} p(x) \log_2(p(x)) = - \left(\frac{6}{17} \times \log_2 \left(\frac{6}{17} \right) \right) \\ &\quad - \left(\frac{6}{17} \times \log_2 \left(\frac{6}{17} \right) \right) - \left(\frac{5}{17} \times \log_2 \left(\frac{5}{17} \right) \right) = 1.5799 \\ H(y_2) &= 1.5222. \end{aligned} \quad (10)$$

According to formula (7), it follows that

$$NMI(y_1; y_2) = \frac{2 * MI(y_1; y_2)}{H(y_1) + H(y_2)} = 0.3645. \quad (11)$$

3.2. ARI (Adjusted Rand Index). The ARI is another widely used evaluation metric for measuring the concordance between two clustering results. The RI (Rand Index) measures the similarity between two results. It calculates all pairs of

samples, including the pairs in identical or different clusters, and pairs in the predicted and ground truth clusters.

For two clustering results X and Y containing n elements, the RI of them is calculated in terms of the following formula.

$$RI = \frac{a + b}{(a + b + c + d)}, \quad (12)$$

where a is the number of pairs in identical class in X and identical cluster in Y ; b is the number of pairs in identical class in X but not identical cluster in Y ; c is the number of pairs that are not in identical class in X but in identical cluster in Y ; d is the number of pairs that are neither in identical class in X nor in identical cluster in Y .

The overlap between X and Y can be summarized in a contingency table, in which each element represents the number of objects in common between X and Y . The ARI is adjusted RI, which can be calculated in terms of the following formula.

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{1/2 \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}, \quad (13)$$

where n_{ij} are values from the contingency table, a_i is the sum of the i -th row of the contingency table, b_j is the sum of the j -th column of the contingency table, and $\binom{n}{2}$ denotes a binomial coefficient.

The ARI is the corrected-for-chance version of the RI. The RI may only yield a value between 0 and 1; the ARI can yield negative values if the index is less than the expected index. The optimal score of the ARI is 1, which represents that two clustering results are identical. A larger ARI signifies higher concordance between X and Y .

3.3. Purity. The *purity* [22] is another simple and transparent evaluation metric for evaluating the clustering performance. For *purity*, each identified cluster is assigned to the label which is most frequent in the cluster, and then the accuracy of this label assignment is computed by counting the number of correctly assigned cells and dividing by the number of cells N . This mapping is not one-to-one and may be biased to the class which has the largest size. Nonetheless, it provides us a simple metric.

For two clustering results X and Y containing n elements, the *purity* of them is calculated in terms of the following formula.

$$\text{purity}(X, Y) = \frac{\sum_p \max_q |X_p \cap Y_q|}{n}, \quad (14)$$

where p and q are the elements of X and Y , respectively.

The *purity* ranges from 0 to 1 while 1 is the optimal score, which denotes that X and Y have identical clustering accuracy. A larger *purity* signifies higher concordance between X and Y .

4. Dataset and Preprocessing

4.1. Dataset. In the study, 33 publicly available scRNA-seq datasets are utilized to testify the performance of the IGRDCGA, which are shown in Table 1. The datasets contain various single-cell gene expression data that are published by many different publications. Each row in the datasets denotes an observation or cell while each column denotes a feature or gene. Table 1 shows the features of the datasets, such as GSE, name, the number of cells (#cell), the number of genes (#gene), the number of the ground truth classes (#class), and references.

4.2. Data Preprocessing. For a scRNA-seq dataset containing N cells and M genes, if x_{ij} represents gene expression level of the i -th cell versus the j -th gene, then its adjacent matrix can be expressed as $D = (x_{ij})_{N \times M}$.

To lower the impacts of large gene expression levels on little gene expression levels, all the data in D are normalized according to the following formula.

$$x'_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min} + \delta}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, M, \quad (15)$$

where x_j^{\min} and x_j^{\max} , respectively, represent the minimum and maximum values of the j -th gene in D . The parameter δ is a very small value, which is to escape the denominator of 0, because x_j^{\min} and x_j^{\max} may be both equal to 0 in scRNA-seq dataset.

It is obvious that all the elements are normalized according to each column (each column denotes a gene). Therefore, after preprocessing, the gene expression level of each gene ranges from 0 to 1, and all the elements in D range from 0 to 1 as well.

5. The Proposed Algorithm

In this study, we present a novel algorithm to address the gene selection and classification for scRNA-seq data by combining information gain ratio and genetic algorithm with dynamic crossover (IGRDCGA for short). The coding and the other details of the IGRDCGA are as follows.

5.1. Coding and Initialization. To choose high quality genes from a huge number of genes, we need to design a good coding for each chromosome. For the scRNA-seq dataset containing N cells of M genes, we design a coding of variable length, whose length can be changed. We number M genes into 1~ M ; the coding of a chromosome can be expressed as follows:

$$C = c_1, c_2, \dots, c_l; 1 \leq c_1, c_2, \dots, c_l \leq M, \quad (16)$$

where c_1, c_2, \dots, c_l , respectively, denote the first locus, the second locus, and the l -th locus of the coding, and they are any integer values larger than 1 and less than M ; l denotes the length of the coding of the chromosome, which is variable for different chromosomes.

Obviously, the coding of a chromosome signifies a combination of genes, namely, a result of gene selection. Therefore, different chromosomes can signify different gene selection, and coding of variable length can signify gene selection of variable length.

According to the above coding rules of a chromosome, Algorithm 1 presents the initialization algorithm to generate the initial population whose population size is N_{pop} .

5.2. Crossover Operator. As the length of the coding of a chromosome is variable, we need to design a new crossover operator to fit the coding of variable length. Given the coding of two parents C_1 and C_2 , their lengths are, respectively, l_1 and l_2 , and the minimum value of l_1 and l_2 is marked as l_3 . The number of the genes to exchange in crossover operator should be less than or equal to l_3 . The study presents a new crossover operator as shown in Algorithm 2.

5.3. Mutation Operator. Similarly, we also need to design a new mutation operator to fit the coding of variable length. Given the coding of one parents C_3 , its length is l_5 . The location of the genes to mutate in mutation operator should be less than or equal to l_5 . The study presents a new mutation operator as shown in Algorithm 3.

5.4. Detailed Steps of IGRDCGA. To illustrate the main process of the proposed algorithm, Figure 1 shows the flow chart of the IGRDCGA.

The detailed steps of the IGRDCGA are summarized as follows.

Step (1). Compute information gain ratio IGR of each gene in terms of formula (5) first. Then, eliminate those genes whose IGR is 0.

Step (2). To eliminate irrelevant genes and choose high quality genes, the threshold method is used to choose those genes whose IGR is the top best δ . In later numerical experiment, $\delta = 1/3$.

Step (3). Implement Algorithm 1 described in Section 5.1 to obtain the initial population P_1 .

Step (4). Compute the fitness and obtain the best chromosome.

- (1) For each chromosome p_i in P_1 , select those genes determined by p_i to obtain a new data D' ; then conduct k means clustering on D' to compute NMI metric as the fitness $f(p_i)$ of p_i .
- (2) Obtain the best chromosome x^* and its fitness $f(x^*)$ by comparing the fitness.

Step (5). Let the number of iterations $t \leftarrow t + 1$; then judge whether $t > t_{\max}$ or not. If yes, then output x^* and $f(x^*)$, and the IGRDCGA terminates; otherwise, turn to the next step.

TABLE 1: ScRNA-seq datasets and their features.

GSE/ID	Name	# Cell	# Gene	#Class	References
GSE74543	Allodiploid	16	2406	2	Li et al., 2016 [23]
GSE42268	Sasagawa	23	32700	3	Sasagawa et al., 2013 [24]
GSE38495	Ramskold	33	21042	7	Ramskold et al., 2012 [25]
GSE57249	Biase	56	25734	4	Biase et al., 2014 [26]
GSE51372	Ting	187	21583	7	Ting et al., 2014 [27]
GSE75688	Chung	518	41821	4	Chung et al., 2017 [28]
GSE85908	Yeo	214	27473	4	Yeo et al., 2017 [29]
GSE64016	Ning	247	19080	3	Ling et al., 2015 [30]
GSE60783	Ginhoux	251	11834	3	Schlitzer et al., 2015 [31]
GSE87795	Su	367	15333	6	Su et al., 2017 [32]
GSE59739	Usoskin	622	25334	4	Usoskin et al., 2015 [33]
GSE45719	Deng	268	22431	6	Deng et al., 2014 [34]
GSE53386	Fan	66	26357	6	Fan et al., 2015 [35]
E-MTAB-3321	Goolam	124	40315	5	Goolam et al., 2016 [36]
E-MTAB-2600	Kolodz	704	13473	3	Kolodziejczyk et al., 2015 [37]
GSE81682	Nestorowa	1656	4773	9	Nestorowa et al., 2016 [38]
GSE52583	Treutlein	80	23271	5	Treutlein et al., 2014 [39]
GSE75140	Camp15	734	18927	6	Camp et al., 2015 [40]
GSE81252	Camp17	777	19020	7	Camp et al., 2017 [41]
GSE36552	Yan	90	20214	6	Yan et al., 2013 [42]
GSE83139	Wang	635	19950	8	Wang et al., 2016 [43]
GSE81861	Li1	561	55186	9	Li et al., 2017 [44]
GSE73727	Li2	60	180253	6	Li et al., 2016 [45]
GSE57872	Patel	430	5948	5	Patel et al., 2014l [46]
SRP041736	Pollen	301	23730	11	Alex et al., 2014 [47]
GSE76381	Manno_m	2150	24378	32	La Manno et al.2016 [48]
GSE76381	Manno_h	4029	20560	56	La Manno et al.2016 [48]
GSE71585	Tasic	1679	24057	18	Tasic et al., 2016 [49]
GSE60361	Zeisel	3005	19486	9	Amit 2015 [50]
GSE76983	Grun	1502	23536	2	Grun et al., 2016 [51]
GSE84133	Baron	1886	14878	13	Baron et al., 2016 [52]
GSE85241	Muraro	2126	19127	10	Muraro et al., 2016 [53]
GSE81608	Xin	1600	39851	8	Xin et al., 2016 [54]

Step (6). Tournament selection strategy [55, 56] is used to choose $Npop$ chromosomes from P_1 to obtain new population P_2 .

Step (7). Generate a random r . If r is less than or equal to the crossover probability p_c , then randomly select two chromosomes from P_2 to perform the **crossover operator** described in Section 5.2. The final population is marked as P_3 .

Step (8). Generate a random r . If r is less than or equal to the mutation probability p_m , then randomly select one chromosome from P_3 to perform the **mutation operator** described in Section 5.3. The final population is marked as P_4 .

Step (9). Compute the fitness value of each chromosome in P_4 and obtain the best chromosome x_2^* and its fitness $f(x^*)$. If $f(x_2^*) > f(x^*)$, then $f(x^*) \leftarrow f(x_2^*)$. $x^* \leftarrow x_2^*$.

Step (10). Turn to **Step 5**.

5.5. Time Complexity Analysis of IGRDCGA. Suppose that a scRNA-seq dataset contains N cells and M genes, and the

population size in GA is $Npop$. The time complexity of the IGRDCGA is analyzed as follows.

The time complexity of Step 1 is $O(M)$, the time complexity of Step 2 is $O(1)$, and the time complexity of Algorithm 1 in Step 3 is $O(Npop)$. In Step 4, the time complexity of the k means clustering is $O(N)$, and that of obtaining the best chromosome is also $O(Npop)$. Step 5 to Step 10 are the main iteration process of the IGRDCGA; their time complexities determine the time complexity of the IGRDCGA. The execution count of Steps 5 and 6 is both t_{max} (the maximal number of iterations), and the time complexities are both $O(t_{max})$. In Step 7, the time complexity of selecting two chromosomes is $O(t_{max})$, while the counts of executions of the crossover operator are $l3$ (the minimum value of the lengths of two chromosome codes), and its time complexity is still $O(l3 * t_{max})$. Therefore, the time complexity of Step 7 is $O(l3 * t_{max})$. Similarly, the time complexity of Step 8 is still $O(l6 * t_{max})$. In Step 9, the time complexity of computing the fitness is $O(t_{max} * Npop * N)$, and the time complexity of obtaining the best chromosome is also $O(t_{max} * Npop)$. Namely, the time complexity of Step 9 is still $O(t_{max} * Npop)$. Obviously, the time complexity of Step 10 is $O(t_{max})$.

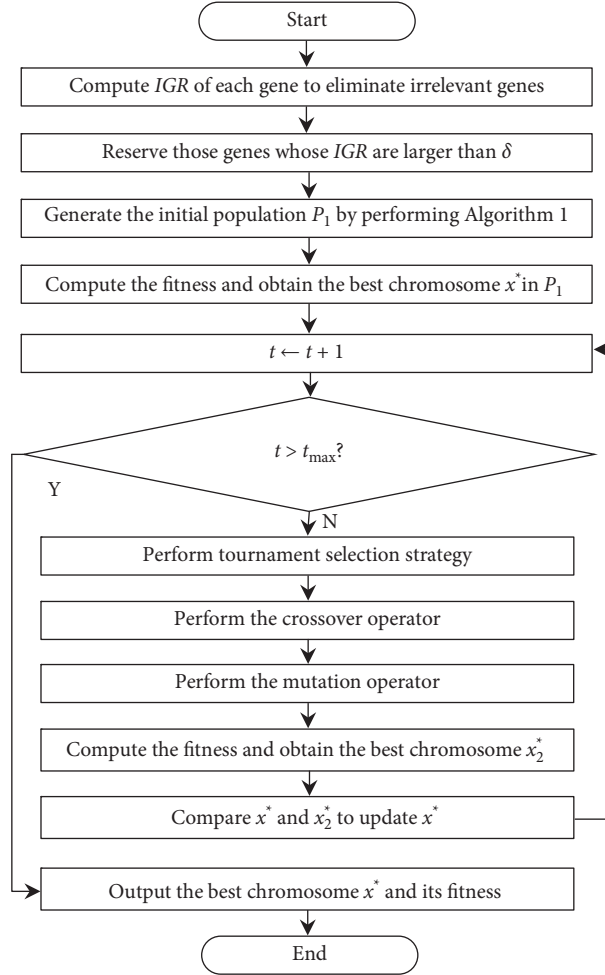


FIGURE 1: Flow chart of the IGRDCGA.

To sum up, the time complexity of the IGRDCGA involves the following time complexities: $O(1)$, $O(M)$, $O(Npop)$, $O(t_{max})$, $O(l3 * t_{max})$, $O(l6 * t_{max})$, $O(t_{max} * Npop)$, and $O(t_{max} * Npop * N)$. Obviously, $O(1) < O(M)$, $O(Npop) < O(t_{max} * Npop) < O(t_{max} * Npop * N)$, $O(t_{max}) < O(t_{max} * Npop) < O(t_{max} * Npop * N)$. Therefore, the decisive time complexities of the IGRDCGA are $O(M)$, $O(l3 * t_{max})$, $O(l6 * t_{max})$, $O(t_{max} * Npop * N)$. As $l3 < M$ and $l6 < M$, it follows that $O(l3 * t_{max}) < O(M * t_{max})$, $O(l6 * t_{max}) < O(M * t_{max})$, and $O(M) < O(M * t_{max})$. Thus, the time complexity of the IGRDCGA is determined by $O(t_{max} * Npop * N)$ and $O(M * t_{max})$. From Table 1, it can be clearly shown that, for most single-cell datasets, $M > N$. Commonly, $M > Npop$. Consequently, for most single-cell datasets, the time complexity of the IGRDCGA can be considered as $O(M * t_{max})$.

6. Numerical Results

In order to evaluate the performances of the IGRDCGA, two frequently used clustering algorithms, k means and spectral clustering [57], a state-of-the-art single-cell classification algorithm SIMLR [58], are employed to compare it. To

compare the performance of gene selection of the IGRDCGA, two frequently used dimensionality reduction algorithms, the PCA and tSNE, are also utilized to compare the IGRDCGA. For the SIMLR, we use its MATLAB program, which can be accessed by the address: <https://github.com/BatzoglouLabSU/SIMLR>. For the k means, PCA, and tSNE, the study utilizes the built-in functions in MATLAB.

6.1. Parameter Values. The related parameters of the IGRDCGA are described as follows.

- (i) *Parameter for IGR.* The threshold $\delta = 1/3$.
- (ii) *Terminal Condition for the IGRDCGA.* The population size $Npop = 100$; the crossover probability $p_c = 0.6$; the mutation probability $p_m = 0.3$; the maximal number of iterations $t_{max} = 100$.

The other competing algorithms use their default parameter values.

6.2. Results

6.2.1. Comparisons of the IGR. The IGR of each gene shows its relevance. The IGR ranges from 0 to 1 while 0 represents

- (1) Generate one random integer l in the interval $[1, M]$.
- (2) Generate l random integers in the interval $[1, M]$.
- (3) Loop the above (1) and (2) $Npop$ times to obtain the initial population.

ALGORITHM 1: Initialization of population.

- (1) Generate one random integer l_4 that is less than or equal to l_3 .
- (2) For C_1 , generate the location to exchange loc_1 , which contains l_4 different random integers that are larger than or equal to 1 and less than or equal to l_1 .
- (3) For C_2 , generate the location to exchange loc_2 , which contains l_4 different random integers that are larger than or equal to 1 and less than or equal to l_2 .
- (4) The coding bits of the location loc_1 in C_1 and the coding bits the location loc_2 in C_2 are exchanged among each other.

ALGORITHM 2: Crossover operator.

- (1) Generate one random integer l_6 that is less than or equal to l_5 .
- (2) For C_3 , generate the location to mutate loc_3 , which contains l_6 different random integers that are larger than or equal to 1 and less than or equal to l_5 .
- (3) The coding bits of the location loc_3 in C_3 are replaced into the other integers that are not equal to any of the coding bits of C_3 .

ALGORITHM 3: Mutation operator.

that the gene has no relevance for the classification. Therefore, this can be employed to obviate irrelevant genes roughly. We compute the *IGR* of each gene for all scRNA-seq datasets and obtain the number of the irrelevant genes (#irrelevant genes) and their percentages in respect of the total genes. The obtained results are shown in Table 2.

From Table 2, we can obviously see that there exist irrelevant genes in most datasets. The percentage of irrelevant genes versus the total genes can indicate gene redundancy rates. There are 24 datasets whose gene redundancy rate is larger than 0 in Table 2, which account for 72.72% (total 33 datasets). This demonstrates that the $IGR = 0$ is a good way to obviate the irrelevant genes.

However, Table 2 also shows that there are 9 datasets whose redundancy rate is 0. Namely, it cannot find the irrelevant genes from the 9 datasets by means of the $IGR = 0$. This also demonstrates that the $IGR = 0$ can only determine irrelevant genes and cannot determine irrelevant genes. Therefore, our proposed algorithm IGRDCGA utilizes a threshold method to eliminate irrelevant genes.

Comparatively speaking, the datasets containing more genes possess higher redundancy rates of genes and vice versa. Nevertheless, this is not absolute, as can be shown in Table 2.

6.2.2. Comparisons of Evaluation Metrics. We perform the IGRDCGA and several competing algorithms on the above scRNA-seq datasets described in Section 4.1. Three evaluation metrics NMI, ARI, and *purity* are employed to evaluate the performances of the IGRDCGA and the other four competing algorithms. All the algorithms are independently

performed for 20 runs to obtain their average values. For all 33 datasets, the average values of NMI, ARI, and *purity* metrics are shown in Tables 3–5, respectively.

From Table 3, we can obviously see that, for 22 of 33 datasets, the IGRDCGA gains the largest NMI in six algorithms, which account for 66.67% (22 of 33 datasets). Meanwhile, we can also observe that, for 6 of the rest 12 datasets, the differences of the largest NMI and the NMI obtained by the IGRDCGA are very little. For the k means, spectral clustering, SIMLR, PCA, and tSNE, the number of the datasets that they obtain the largest NMI is, respectively, 0, 1, 6, 2, and 6. By comparison, the IGRDCGA outperforms the other five competing algorithms in terms of NMI.

Table 4 shows that, for 24 of 33 datasets, the IGRDCGA acquires the maximal ARI in six algorithms, which account for 72.72%. For the k means, spectral clustering, SIMLR, PCA, and tSNE, the number of the datasets that they obtain the maximal ARI is, respectively, 0, 0, 3, 1, and 7. By contrast, the IGRDCGA is superior to the other five competing algorithms in terms of ARI.

From Table 5, it can be clearly seen that, for 21 of 33 datasets, the *purity* metrics obtained by the IGRDCGA are the largest in six algorithms, which account for 63.63%. For the k means, spectral clustering, SIMLR, PCA, and tSNE, the number of the datasets that they obtain the largest *purity* metrics is, respectively, 0, 1, 6, 1, and 6. By comparison, the IGRDCGA outperforms the other five competing algorithms in terms of *purity* metric.

By comparing Tables 3–5, we summarize the best evaluation metrics obtained by six algorithms in Table 6, where the NMI, ARI, and *purity* metrics are, respectively, denoted by 1, 2, and 3.

TABLE 2: Irrelevant genes and their percentage.

Dataset	#Of genes	#Irrelevant genes	Percentage (%)
Allodiploid	2406	0	0
Sasagawa	32700	15412	47.13
Ramskold	21042	1005	4.78
Biase	25734	51	0.2
Ting	21583	4349	20.15
Chung	41821	14217	33.99
Yeo	27473	4042	14.71
Ning	19080	1962	10.28
Ginhoux	11834	0	0
Su	15333	0	0
Usoskin	25334	5800	22.89
Deng	22431	1257	5.6
Fan	26357	941	3.57
Goolam	40315	12592	31.23
Kolodz	13473	0	0
Nestorowa	4773	0	0
Treutlein	23271	10497	45.11
Camp15	18927	2135	11.28
Camp17	19020	1285	6.76
Yan	20214	619	3.06
Wang	19950	0	0
Li1	55186	12131	21.98
Li2	180253	40335	22.38
Patel	5948	0	0
Pollen	23730	1920	8.09
Manno_m	24378	2390	9.8
Manno_h	20560	0	0
Tasic	24057	2443	10.16
Zeisel	19486	0	0
Grun	23536	7928	33.68
Baron	14878	17	0.11
Muraro	19127	225	1.18
Xin	39851	5962	14.96

TABLE 3: NMI metric obtained by the IGRDCGA and several competing algorithms.

Dataset	K means	Spectral	SIMLR	IGRDCGA	PCA	tSNE
Allodiploid	0.3793	0.0778	1.0000	1.0000	1.0000	0.0490
Sasagawa	0.3048	0.1815	0.4665	0.9128	0.6349	0.0903
Ramskold	0.6868	0.4946	0.7594	0.9709	0.9394	0.3444
Biase	0.6690	0.5114	0.8321	1.0000	0.9501	0.0544
Ting	0.2161	0.1211	0.5082	0.5778	0.5746	0.0585
Chung	0.1280	0.0295	0.3170	0.4196	0.2278	0.3336
Yeo	0.2358	0.0480	0.7744	0.8168	0.7650	0.6831
Ning	0.1100	0.0388	0.3465	0.7432	0.3579	0.0073
Ginhoux	0.4094	0.0326	0.4189	0.5580	0.4777	0.0074
Su	0.3145	0.0676	0.4963	0.5951	0.3910	0.0201
Usoskin	0.1823	0.0363	0.4552	0.8334	0.6235	0.6579
Deng	0.7451	0.3860	0.6979	0.8977	0.7671	0.0334
Fan	0.2939	0.1910	0.4925	0.5874	0.5249	0.1417
Goolam	0.2826	0.1917	0.4464	0.7428	0.3282	0.0527
Kolodz	0.7024	0.0235	0.9634	1.0000	0.9790	0.0035
Nestorowa	0.3499	0.1983	0.3514	0.3774	0.3562	0.3795
Treutlein	0.2520	0.2585	0.4574	0.8074	0.5445	0.0754
Camp15	0.4025	0.0596	0.5098	0.4945	0.4257	0.4979
Camp17	0.6479	0.0735	0.7420	0.7344	0.7366	0.0117
Yan	0.7034	0.4988	0.7937	0.9480	0.8830	0.0872
Wang	0.0609	0.0613	0.2588	0.3038	0.2226	0.3274

TABLE 3: Continued.

Dataset	K means	Spectral	SIMLR	IGRDCGA	PCA	tSNE
Li1	0.4852	0.0634	0.7377	0.7623	0.7168	0.0275
Li2	0.2117	0.1807	0.3387	0.5915	0.3452	0.1425
Patel	0.1669	0.0429	0.6446	0.7436	0.7349	0.8248
Pollen	0.5846	0.1790	0.9272	0.9008	0.9055	0.8886
Manno_m	0.3084	0.0887	0.4026	0.3979	0.4065	0.4710
Manno_h	0.3796	0.6052	0.4867	0.4849	0.5287	0.5927
Tasic	0.6758	0.7707	0.7456	0.8458	0.8039	0.8180
Zeisel	0.4681	0.6586	0.7097	0.6076	0.5229	0.7391
Grun	0.0059	0.1145	0.4133	0.3729	0.0055	0.0005
Baron	0.2079	0.5731	0.5950	0.5480	0.4281	0.6618
Muraro	0.2930	0.6647	0.7001	0.6897	0.4189	0.0095
Xin	0.2587	0.4237	0.2422	0.7201	0.5445	0.0100

TABLE 4: ARI metric obtained by the IGRDCGA and several competing algorithms.

Dataset	K means	Spectral	SIMLR	IGRDCGA	PCA	tSNE
Allodiploid	0.2767	-0.0500	1.0000	1.0000	1.0000	-0.0125
Sasagawa	0.1322	0.0262	0.4306	0.9023	0.5238	-0.0037
Ramskold	0.3999	0.1379	0.5358	0.9481	0.8590	0.0024
Biase	0.5343	0.2827	0.7346	1.0000	0.9556	-0.0118
Ting	0.0985	0.0196	0.3019	0.3905	0.3044	0.0002
Chung	0.0420	-0.0021	0.2011	0.2059	0.1222	0.2216
Yeo	0.1408	-0.0068	0.7760	0.7964	0.7675	0.5966
Ning	0.0608	0.0007	0.3269	0.7456	0.3137	-0.0004
Ginhoux	0.3753	-0.0016	0.3379	0.6062	0.4790	-0.0002
Su	0.3035	0.0084	0.3329	0.5590	0.3180	0.0000
Usoskin	0.1033	0.0020	0.3007	0.8343	0.5319	0.5538
Deng	0.7033	0.1435	0.5059	0.8786	0.5460	0.0011
Fan	0.1156	0.0386	0.3323	0.4049	0.3529	0.0060
Goolam	0.1436	0.0656	0.2284	0.7564	0.1433	0.0005
Kolodz	0.6468	0.0010	0.9762	1.0000	0.9880	0.0011
Nestorowa	0.2236	0.0735	0.2166	0.2572	0.2284	0.2584
Treutlein	0.1204	0.0741	0.2615	0.8138	0.4066	-0.0038
Camp15	0.3614	0.0141	0.4402	0.4835	0.3537	0.4121
Camp17	0.4667	0.0031	0.4859	0.6025	0.5641	-0.0005
Yan	0.5801	0.2894	0.6304	0.9425	0.8453	-0.0045
Wang	0.0084	0.0025	0.1445	0.1239	0.1111	0.1926
Li1	0.2265	0.0002	0.6032	0.6023	0.5354	-0.0002
Li2	0.0280	0.0111	0.1496	0.3423	0.1793	0.0014
Patel	0.0581	-0.0012	0.6065	0.7383	0.7424	0.8437
Pollen	0.2423	0.0177	0.8436	0.8258	0.8326	0.7889
Manno_m	0.0721	0.0024	0.1317	0.1351	0.1328	0.1822
Manno_h	0.0774	0.3050	0.1489	0.1593	0.2006	0.2767
Tasic	0.4326	0.6113	0.6176	0.8090	0.6889	0.6716
Zeisel	0.3023	0.4970	0.5730	0.4676	0.3418	0.5866
Grun	-0.0300	0.0730	0.4375	0.5053	-0.0355	-0.0001
Baron	0.0516	0.2526	0.2887	0.3202	0.1450	0.3251
Muraro	0.1306	0.4838	0.5716	0.6391	0.2261	-0.0002
Xin	0.2802	0.2389	0.0771	0.8145	0.3374	0.0000

From Table 6, we can clearly see that, for 15 of 33 datasets, the NMI, ARI, and *purity* metrics obtained by the IGRDCGA are all the best in six algorithms, which account for 45.45%. For the k means, spectral clustering, SIMLR, PCA, and tSNE, the number of the datasets that they obtain the best NMI, ARI and *purity* metrics is, respectively, 0, 0, 2, 1, and 6. Thus, the IGRDCGA outperforms the other five competing algorithms in terms of NMI, ARI, and *purity* metrics.

In the meantime, Table 6 shows that only in partial datasets does the IGRDCGA gain the best NMI, ARI, and *purity* metrics. For the Allodiploid, the IGRDCGA, SIMLR, and PCA obtain the best NMI, ARI, and *purity* metrics. We can observe from Table 1 that the Allodiploid possesses much less number of cells and genes compared with the other datasets. Namely, for the datasets with smaller dimensions, an algorithm is easy to obtain the best NMI, ARI,

TABLE 5: *Purity* metric obtained by the IGRDCGA and several competing algorithms.

Dataset	K means	spectral	SIMLR	IGRDCGA	PCA	tSNE
Allodiploid	0.2767	-0.0500	1.0000	1.0000	1.0000	-0.0125
Sasagawa	0.0387	-0.2253	0.5020	0.9170	0.5731	0.1462
Ramskold	0.5953	0.1801	0.7386	0.9756	0.9331	0.5491
Biase	0.5124	0.1796	0.7896	1.0000	0.9636	0.2149
Ting	-0.1852	-0.4401	0.5643	0.4239	0.5204	0.3951
Chung	-0.1686	-0.4338	0.3331	0.1288	0.2860	0.3774
Yeo	-0.0742	-0.4393	0.8247	0.8352	0.8196	0.6856
Ning	-0.1955	-0.3228	0.3667	0.7629	0.3466	0.1128
Ginhoux	0.3983	-0.3019	0.3912	0.6425	0.5308	0.1045
Su	0.4949	-0.4755	0.5004	0.7007	0.5594	0.3739
Usoskin	-0.1365	-0.4286	0.3846	0.8584	0.6343	0.6506
Deng	0.7271	-0.0730	0.6137	0.8920	0.6378	0.2573
Fan	0.1441	-0.2779	0.5980	0.6421	0.6073	0.4231
Goolam	0.1340	-0.1726	0.3632	0.7725	0.2284	0.1841
Kolodz	0.5854	-0.2917	0.9783	1.0000	0.9891	0.1003
Nestorowa	0.5335	-0.0128	0.6124	0.6035	0.6294	0.6444
Treutlein	-0.0020	-0.1692	0.4146	0.8372	0.5156	0.2137
Camp15	0.4989	-0.4715	0.6333	0.6594	0.5855	0.6197
Camp17	0.6779	-0.6129	0.6976	0.7836	0.7586	0.4854
Yan	0.6861	0.3208	0.7748	0.9620	0.9001	0.3921
Wang	-0.3533	-0.5255	0.4835	0.2183	0.4099	0.5145
Li1	0.2909	-0.7039	0.8271	0.8050	0.7802	0.5820
Li2	-0.2315	0.4114	0.4866	0.4227	0.3781	0.4231
Patel	-0.3625	-0.5657	0.7379	0.8244	0.8327	0.8982
Pollen	0.2559	-0.5278	0.9414	0.9271	0.9369	0.9210
Manno_m	0.4832	-0.8098	0.8236	0.7025	0.8149	0.8482
Manno_h	0.6085	0.9307	0.8803	0.8370	0.9099	0.9268
Tasic	0.6178	0.8780	0.8747	0.9231	0.8948	0.8939
Zeisel	0.5849	0.7180	0.7558	0.6948	0.6250	0.7646
Grun	0.2741	0.0728	0.4554	0.6157	0.2331	-0.0003
Baron	0.1026	0.5293	0.5466	0.5121	0.4274	0.5723
Muraro	0.2705	0.7018	0.7430	0.7552	0.4868	0.4388
Xin	0.1638	0.3518	0.2060	0.8166	0.4312	0.1518

and *purity* metrics. In this case, the PCA is the fittest method as it is the simplest and easiest to implement in the above three algorithms. For the Ting, Chung, and Li2, the IGRDCGA obtains the best NMI and ARI, while the SIMLR obtains the best *purity* metrics. For the Camp15, Camp17, Grun, and Muraro, the IGRDCGA obtains the best ARI and *purity* metrics, while the SIMLR obtains the best NMI. By the comparisons of the IGRDCGA with the SIMLR, we can clearly see that the number of the best metrics obtained by the IGRDCGA is far larger than that obtained by the SIMLR. This fully demonstrates that the IGRDCGA is superior to the SIMLR in terms of NMI, ARI, and *purity* metrics.

For the Nestorowa, Wang, Patel, Manno_m, Zeisel, and Baron datasets, the tSNE gains the best NMI, ARI, and *purity* metrics. However, for the other datasets, three evaluation metrics obtained by the tSNE are far worse than those obtained by the IGRDCGA. As shown in Table 3, there are 17 datasets whose NMI of the tSNE is less than 0.1, while there are none datasets whose NMI of the IGRDCGA is less than 0.1. In Table 4, there are 20 datasets whose ARI of the tSNE is less than 0.1, while there are none datasets whose ARI of the IGRDCGA is less than 0.1. In Table 5, there are 19 datasets whose *purity* metrics of the tSNE are less than 0.5, while there are only 4 datasets whose *purity* metrics of the

IGRDCGA are less than 0.5. Namely, for most datasets except the above six datasets, the NMI, ARI, and *purity* metrics obtained by the IGRDCGA are superior to those obtained by the tSNE. This attests that the IGRDCGA outperforms the tSNE for most datasets in terms of NMI, ARI, and *purity* metrics.

By elaborative analyses for the above six datasets and the other datasets, we can clearly observe the following principal features and differences. To begin with, the six datasets all possess very low redundancy rates of genes. Table 2 shows that the redundancy rates of the Zeisel and Baron are, respectively, 9.8 and 0.11, and the other four datasets are all 0. For the datasets with low redundancy rates, the IGRDCGA may lose some genes to cause the decreasing of the classification performances. In addition, the six datasets have relatively low NMI and ARI for six algorithms. From Table 3, we can easily see that all the NMI metrics of the Nestorowa, Wang, and Manno_m are, respectively, less than 0.38, 0.33, and 0.48. From Table 4, it can be obviously seen that all the ARI metrics of the Nestorowa, Wang, Manno_m, and Baron are, respectively, less than 0.26, 0.20, 0.19, and 0.33. Thirdly, the six datasets are approximately and completely sparse. From our provided appendix file, it can obviously observe that the data

TABLE 6: The best evaluation metrics obtained by six algorithms.

Dataset	K means	spectral	SIMLR	IGRDCGA	PCA	tSNE
Allodiploid			1, 2, 3	1, 2, 3	1, 2, 3	
Sasagawa				1, 2, 3		
Ramskold				1, 2, 3		
Biase				1, 2, 3		
Ting			3	1, 2		
Chung			3	1, 2		
Yeo				1, 2, 3		
Ning				1, 2, 3		
Ginhoux				1, 2, 3		
Su				1, 2, 3		
Usoskin				1, 2, 3		
Deng				1, 2, 3		
Fan				1, 2, 3		
Goolam				1, 2, 3		
Kolodz				1, 2, 3		
Nestorowa						1, 2, 3
Treutlein				1, 2, 3		
Camp15			1	2, 3		
Camp17			1	2, 3		
Yan				1, 2, 3		
Wang						1, 2, 3
Li1			2, 3	1		
Li2			3	1, 2		
Patel						1, 2, 3
Pollen			1, 2, 3			
Manno_m						1, 2, 3
Manno_h		1, 3				2
Tasic				1, 2, 3		
Zeisel						1, 2, 3
Grun			1	2, 3		
Baron						1, 2, 3
Muraro			1	2, 3		
Xin				1, 2, 3		

values within $[0, 0.1]$ in all datasets are more than those of the other data intervals. Nevertheless, the situation of the six datasets is a great deal more highlighted, and they are approximately and completely sparse if we let the data values within $[0, 0.1]$ to be 0. For the Nestorowa, Wang, Patel, Manno_m, Zeisel, and Baron, the data values within $[0, 0.1]$, respectively, account for more than 88%, 95%, 73%, 91%, 80%, and 91%, which are so high that they are nearly cover up the data values within the other ranges.

6.2.3. Iteration Plots and Heat Maps. In order to clearly illustrate the performance of the IGRDCGA, three scRNA-seq datasets, Allodiploid, Chung, and Kolodz, are selected as representative datasets according to the different values of the NMI metrics in Table 3. For the Allodiploid, Yeo, and Grun, their iteration plots are, respectively, illustrated in Figures 2–4. For the Allodiploid, the NMI metric obtained by the IGRDCGA is 1, which represents that the NMI metrics are consistent with the ground truth class labels. It can be clearly seen from Figure 2 that the iteration plot of the Allodiploid is parallel to the X -axis and its value is the maximal NMI of 1. Namely, the IGRDCGA obtains the maximal NMI at the first iteration. Therefore, the later iterations are useless. In the above 33 scRNA-seq datasets, only the Allodiploid displays the plot illustrated in Figure 2.

In Figure 3, the iteration plot of Yeo is ascending as the number of iterations increases. The ascending trend is very obvious at the early period of the iteration, and the ascending trend stops as the number of iterations increases. Namely, the NMI metric turns into a fixed number at the later period of the iteration. In the above 33 scRNA-seq datasets, most datasets display a similar plot shown in Figure 3.

From Figure 4, we can obviously see that the iteration plot of Grun is always ascending as the number of iterations increases. The ascending trend is always very obvious at the whole period of the iteration. In the above 33 scRNA-seq datasets, a few datasets display a similar plot shown in Figure 4.

In order to illustrate the classification performance of the IGRDCGA more visually, the heat maps of the Allodiploid, Sasagawa, and Yan are illustrated in Figures 5–7, respectively.

The Allodiploid contains 16 cells, 2406 genes, and 2 classes, whose ground truth class labels are illustrated in the first column of Figure 5. From Figure 5, it can be obviously seen that the results obtained by the IGRDCGA, SIMLR, and PCA are fully concordant with the ground truth class labels; those obtained by the k means and spectral clustering are nearly one class; and those obtained by the tSNE are oscillatory. Obviously, the results obtained by the k means, spectral clustering, and the tSNE are a great deal worse than

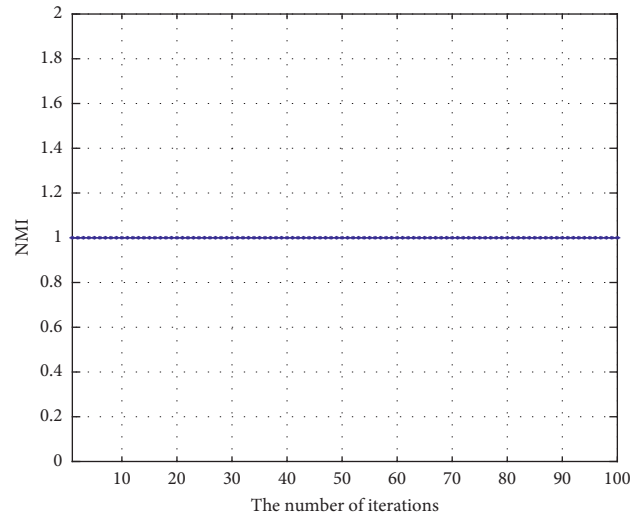


FIGURE 2: Iteration plot of Allodiploid.

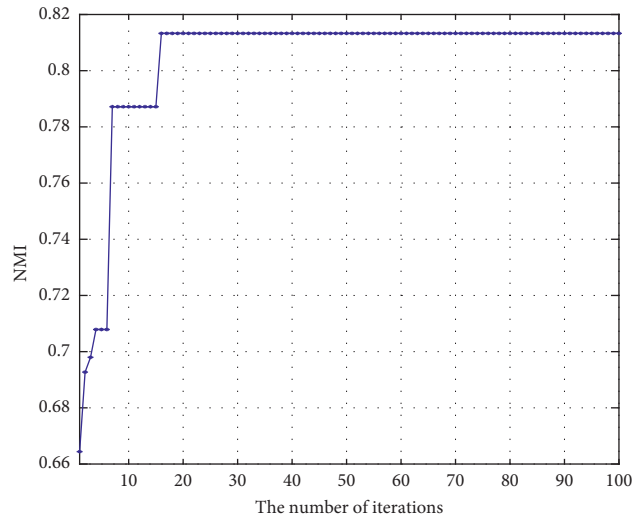


FIGURE 3: Iteration plot of Yeo.

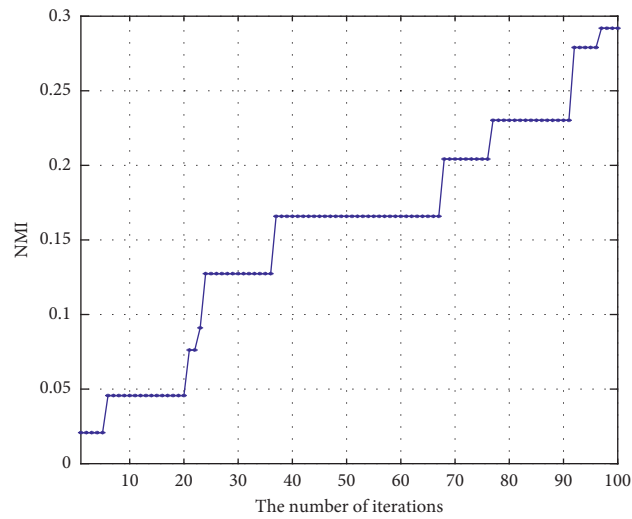


FIGURE 4: Iteration plot of Grun.

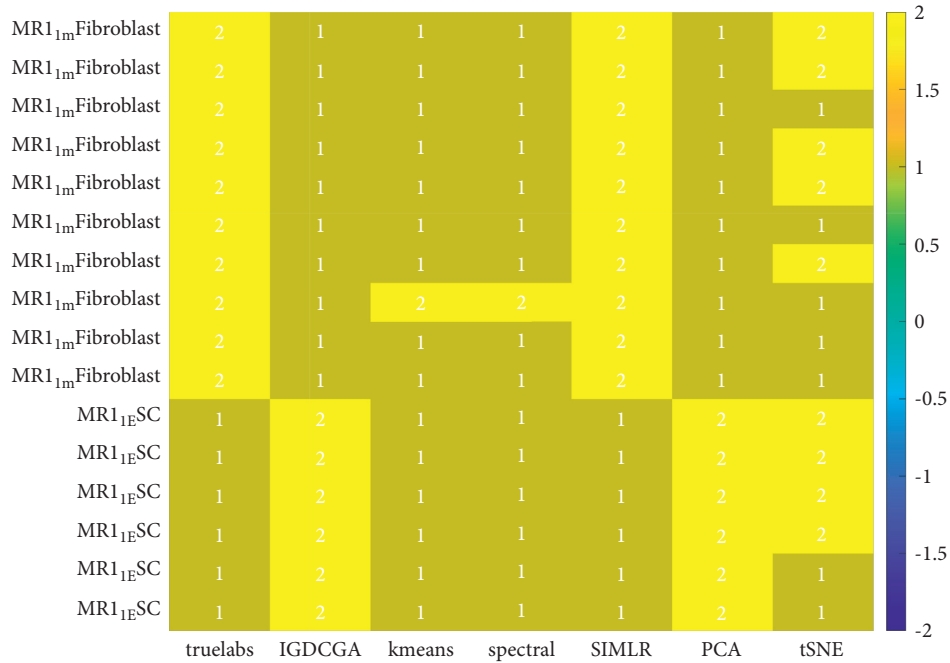


FIGURE 5: Heat map of the Allodiploid.

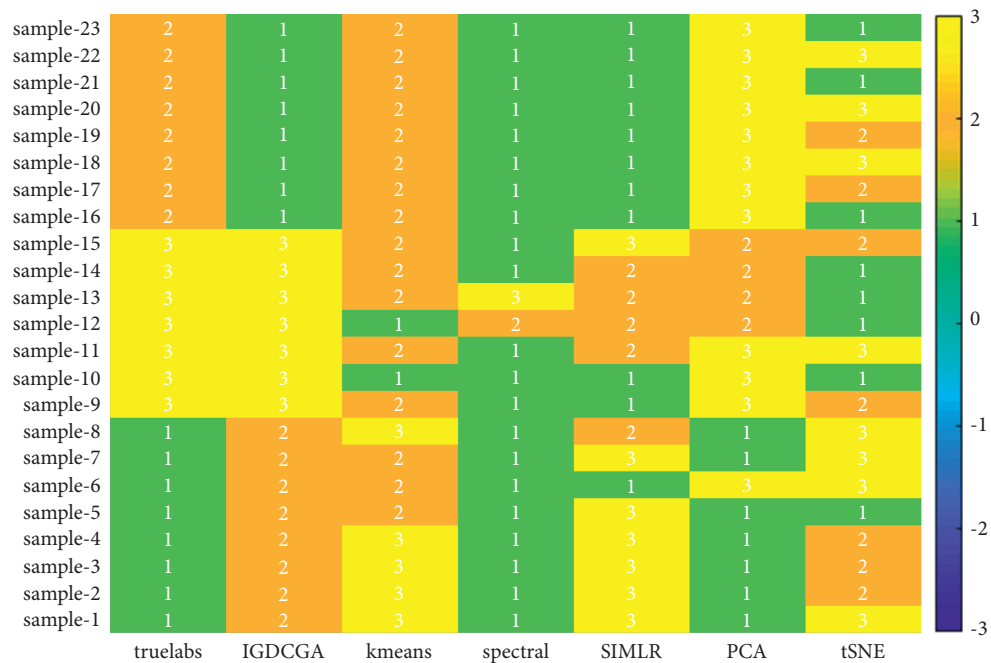


FIGURE 6: Heat map of the Sasagawa.

those obtained by the IGRDCGA, SIMLR, and PCA; the results of the tSNE are the worst in the six algorithms.

The dataset Sasagawa contains 23 cells, 32700 genes, and 3 classes, whose ground truth labels are illustrated in the first column of Figure 6. From Figure 6, it can be obviously observed that the results obtained by the IGRDCGA are fully concordant with the ground truth class labels; those obtained by spectral clustering are nearly one class; those obtained by the tSNE are oscillatory and the worst. In the

results obtained by the k means, SIMLR, and PCA, the results of the subclass 2 (the ground truth class labels are 2) are fully concordant with the ground truth class labels while the other subclasses are confused. Therefore, for the Sasagawa, the IGRDCGA gains the best results while the tSNE gains the worst results.

The first column of Figure 7 illustrates the ground truth labels of the Yan, which contains 90 cells, 20214 genes, and 6 classes. From Figure 7, we can clearly see that

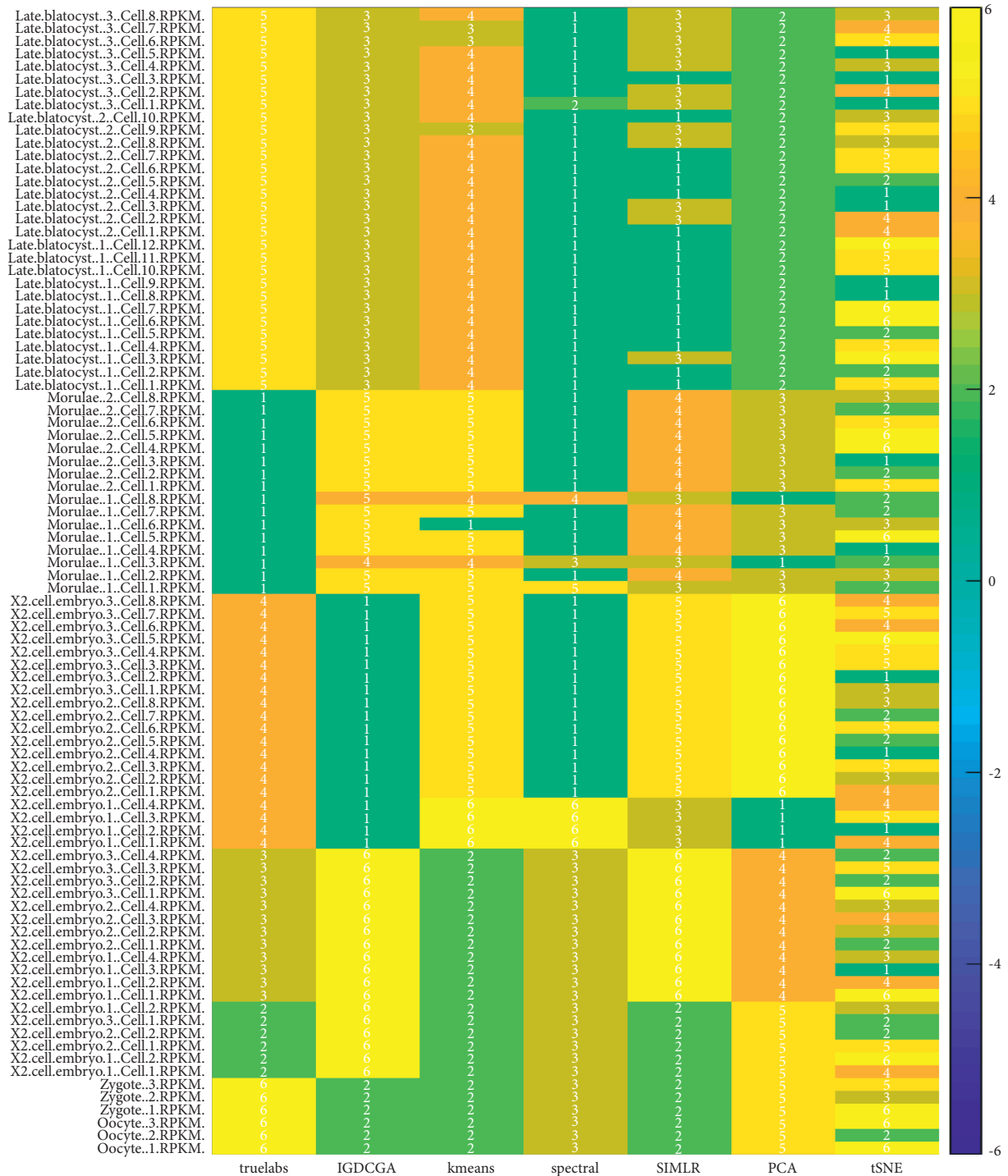


FIGURE 7: Heat map of the Yan.

none of the results obtained by six algorithms are concordant with the ground truth class labels. However, the results obtained by the IGRDCGA are closer to the ground truth class labels than those obtained by the other algorithms, and the results of the subclasses 2, 3, 4, 5, and 6 are fully concordant with the ground truth class labels. The results obtained by the tSNE are oscillatory and the worst in the six algorithms. The results obtained by the other algorithms are all confused. Therefore, for the Yan, the

IGRDCGA obtains the best results whereas the tSNE obtains the worst results.

7. Conclusion and Future Work

In this study, we present a novel algorithm to address the gene selection and classification for scRNA-seq data. It combines information gain ratio (IGR) and genetic algorithm with dynamic crossover (DCGA) and are abbreviated

as IGRDCGA. It utilizes information gain ratio to eliminate irrelevant genes roughly and utilizes DCGA to choose high quality genes finely. We have conducted the IGRDCGA and several competing algorithms on 33 publicly available scRNA-seq datasets. The obtained results demonstrate that the IGRDCGA can eliminate irrelevant genes and choose high quality genes effectively, and it is superior to the other several competing algorithms in terms of classification accuracy.

This algorithm is going on for further enhancement and improvement. One attempt is to utilize a more efficient coding to speed up its converging rate and stability. Another attempt is to extend the IGRDCGA to classification algorithms of the other high dimensional problems.

Data Availability

The datasets supporting this study are publicly available and they can be downloaded from EMBL-EBI (<https://www.ebi.ac.uk/>) or the NCBI Gene Expression Omnibus (GEO) repository (<https://www.ncbi.nlm.nih.gov/geo/>).

Conflicts of Interest

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

Acknowledgments

This research was supported by Science Research Foundation for High-level Talents of Yulin Normal University (no. G2021ZK17).

References

- [1] Y. Wang, H. Liu, F. Wei, T. Zong, and X. Li, "Cooperative coevolution with formula-based variable grouping for large-scale global optimization," *Evolutionary Computation*, vol. 26, no. 4, pp. 569–596, 2018.
- [2] B. Nakisa, M. N. Rastgoo, D. Tjondronegoro, and V. Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors," *Expert Systems with Applications*, vol. 93, pp. 143–155, 2018.
- [3] D. Y. Eroglu and K. Kilic, "A novel hybrid genetic local search algorithm for feature selection and weighting with an application in strategic decision making in innovation management," *Information Sciences*, vol. 405, pp. 18–32, 2017.
- [4] N. Maleki and Y. Zeinali, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Systems with Applications*, vol. 164, p. 113981, 2020.
- [5] M. Tahir, A. Tubaishat, F. Al-Obeidat et al., "A novel binary chaotic genetic algorithm for feature selection and its utility in affective computing and healthcare," *Neural Computing and Applications*, 2020.
- [6] A. H. Mohammad, "Comparing two feature selections methods (information gain and gain ratio) on three different classification algorithms using Arabic dataset," *Journal of Theoretical & Applied Information Technology*, vol. 96, no. 6, pp. 1561–1569, 2018.
- [7] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems*, vol. 24, no. 7, pp. 1024–1032, 2011.
- [8] A. Chinnaswamy and R. Srinivasan, "Hybrid information gain based fuzzy roughset feature selection in cancer microarray data," in *Proceedings of the International Conference on Innovations in Power and Advanced Computing Technologies (I-PACT)*, Vellore, India, April 2017.
- [9] C. Dai, Y. Wang, M. Ye, X. Xue, and H. Liu, "An orthogonal evolutionary algorithm with learning automata for multi-objective optimization," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3306–3319, 2016.
- [10] X. Xue and Y. Wang, "Using memetic algorithm for instance coreference resolution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 580–591, 2016.
- [11] J. Liu, Y. Wang, N. Fan, S. Wei, and W. Tong, "A convergence-diversity balanced fitness evaluation mechanism for decomposition-based many-objective optimization algorithm," *Integrated Computer-Aided Engineering*, vol. 26, no. 2, pp. 159–184, 2019.
- [12] H. Liu, Y. Wang, L. Liu, and X. Li, "A two phase hybrid algorithm with a new decomposition method for large scale optimization," *Integrated Computer-Aided Engineering*, vol. 25, no. 4, pp. 349–367, 2018.
- [13] M. Ye, "A hybrid genetic algorithm for the minimum exposure path problem of wireless sensor networks based on a numerical functional extreme model," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8644–8657, 2015.
- [14] X. Xue and Y. Wang, "Optimizing ontology alignments through a memetic algorithm using both MatchFmeasure and unanimous improvement ratio," *Artificial Intelligence*, vol. 223, pp. 65–81, 2015.
- [15] D. Cai and Y. Wang, "A new decomposition based evolutionary algorithm with uniform designs for many-objective optimization," *Applied Soft Computing*, vol. 30, no. 1, pp. 238–248, 2015.
- [16] Y.-M. Cheung, F. Gu, and H.-L. Liu, "Objective extraction for many-objective optimization problems: algorithm and test problems," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 755–772, 2016.
- [17] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.
- [18] A. F. McDaid, D. Greene, and N. Hurley, "Normalized mutual information to evaluate overlapping community finding algorithms," 2011, <https://arxiv.org/abs/1110.2515>.
- [19] O. Abedinia, N. Amjadi, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62–74, 2017.
- [20] C. Xu and Z. Su, "Identification of cell types from single-cell transcriptomes using a novel clustering method," *Bioinformatics*, vol. 31, no. 12, pp. 1974–1980, 2015.
- [21] V. Y. Kiselev, K. Kirschner, M. T. Schaub et al., "SC3: consensus clustering of single-cell RNA-seq data," *Nature Methods*, vol. 14, no. 5, pp. 483–486, 2017.
- [22] S. Wagner and D. Wagner, "Comparing clusterings: an overview," Universität Karlsruhe, Fakultät für Informatik Karlsruhe, Karlsruhe, Germany, 2007.
- [23] X. Li, X. L. Cui, J. Q. Wang, Y. K. Wang, Y. F. Li, and L. Y. Wang, "Generation and application of mouse-rat

- allodiploid embryonic stem cells,” *Cell*, vol. 164, no. 1-2, pp. 279–292, 2016.
- [24] Y. Sasagawa, I. Nikaido, T. Hayashi et al., “Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity,” *Genome Biology*, vol. 14, no. 4, p. 3097, 2013.
- [25] Ramsköld, S. Luo, Y. C. Wang et al., “Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells,” *Nature Biotechnology*, vol. 30, no. 8, pp. 777–782, 2012.
- [26] F. H. Biase, X. Cao, and S. Zhong, “Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing,” *Genome Research*, vol. 24, no. 11, pp. 1787–1796, 2014.
- [27] D. T. Ting, B. S. Wittner, M. Ligorio et al., “Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells,” *Cell Reports*, vol. 8, no. 6, pp. 1905–1918, 2014.
- [28] W. Chung, H. H. Eum, H.-O. Lee et al., “Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer,” *Nature Communications*, vol. 8, no. 1, p. 15081, 2017.
- [29] T. Yeo, S. J. Tan, C. L. Lim et al., “Microfluidic enrichment for the single cell analysis of circulating tumor cells,” *Scientific Reports*, vol. 6, no. 1, pp. 22076–22087, 2016.
- [30] N. Leng, L. F. Chu, C. Barry et al., “Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments,” *Nature Methods*, vol. 12, no. 10, pp. 947–950, 2015.
- [31] A. Schlitzer, V. Sivakamasundari, J. Chen et al., “Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow,” *Nature Immunology*, vol. 16, no. 7, pp. 718–728, 2015.
- [32] X. Su, Y. Shi, X. Zou et al., “Single-cell RNA-Seq analysis reveals dynamic trajectories during mouse liver development,” *BMC Genomics*, vol. 18, no. 1, pp. 1–14, 2017.
- [33] D. Usoskin, A. Furlan, S. Islam et al., “Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing,” *Nature Neuroscience*, vol. 18, no. 1, pp. 145–153, 2015.
- [34] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, “Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells,” *Science*, vol. 343, no. 6167, pp. 193–196, 2014.
- [35] X. Fan, X. Zhang, X. Wu et al., “Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos,” *Genome Biology*, vol. 16, no. 1, pp. 148–217, 2015.
- [36] M. Goolam, A. Scialdone, S. J. L. Graham et al., “Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos,” *Cell*, vol. 165, no. 1, pp. 61–74, 2016.
- [37] A. A. Kolodziejczyk, J. K. Kim, J. C. H. Tsang et al., “Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation,” *Cell Stem Cell*, vol. 17, no. 4, pp. 471–485, 2015.
- [38] S. Nestorowa, F. K. Hamey, B. Pijuan Sala et al., “A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation,” *Blood*, vol. 128, no. 8, pp. e20–e31, 2016.
- [39] B. Treutlein, D. G. Brownfield, A. R. Wu et al., “Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq,” *Nature*, vol. 509, no. 7500, pp. 371–375, 2014.
- [40] J. G. Camp, F. Badsha, M. Florio et al., “Human cerebral organoids recapitulate gene expression programs of fetal neocortex development,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 51, pp. 15672–15677, 2015.
- [41] J. G. Camp, K. Sekine, T. Gerber et al., “Multilineage communication regulates human liver bud development from pluripotency,” *Nature*, vol. 546, no. 7659, pp. 533–538, 2017.
- [42] L. Yan, M. Yang, H. Guo et al., “Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells,” *Nature Structural & Molecular Biology*, vol. 20, no. 9, pp. 1131–1139, 2013.
- [43] Y. J. Wang, J. Schug, K.-J. Won et al., “Single-cell transcriptomics of the human endocrine pancreas,” *Diabetes*, vol. 65, no. 10, pp. 3028–3038, 2016.
- [44] H. Li, E. T. Courtois, D. Sengupta et al., “Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors,” *Nature Genetics*, vol. 49, no. 5, pp. 708–718, 2017.
- [45] L. Jin, “Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types,” *EMBO Reports*, vol. 17, no. 2, pp. 178–187, 2016.
- [46] A. P. Patel, I. Tirosh, J. J. Trombetta et al., “Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma,” *Science*, vol. 344, no. 6190, pp. 1396–1401, 2014.
- [47] A. Alex, “Pollen, Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex,” *Nature Biotechnology*, vol. 32, no. 10, pp. 1053–1058, 2014.
- [48] G. La Manno, D. Gyllborg, S. Codeluppi et al., “Molecular diversity of midbrain development in mouse, human, and stem cells,” *Cell*, vol. 167, no. 2, pp. 566–580, 2016.
- [49] B. Tasic, V. Menon, T. N. Nguyen et al., “Adult mouse cortical cell taxonomy revealed by single cell transcriptomics,” *Nature Neuroscience*, vol. 19, no. 2, pp. 335–346, 2016.
- [50] Z. Amit, “Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq,” *Science*, vol. 347, no. 6226, pp. 1138–1142, 2015.
- [51] D. Grün, M. J. Muraro, J. C. Boisset et al., “De novo prediction of stem cell identity using Single-Cell transcriptome data,” *Cell Stem Cell*, vol. 19, no. 2, pp. 266–277, 2016.
- [52] M. Baron, A. Veres, S. L. Wolock et al., “A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure,” *Cell Systems*, vol. 3, no. 4, pp. 346–360, 2016.
- [53] M. J. Muraro, G. Dharmadhikari, D. Grün et al., “A single-cell transcriptome atlas of the human pancreas,” *Cell Systems*, vol. 3, no. 4, pp. 385–394, 2016.
- [54] Y. Xin, J. Kim, H. Okamoto et al., “RNA sequencing of single human islet cells reveals type 2 diabetes genes,” *Cell Metabolism*, vol. 24, no. 4, pp. 608–615, 2016.
- [55] L. Adam and D. Lipowska, “Roulette-wheel selection via stochastic acceptance,” *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 6, pp. 2193–2196, 2012.
- [56] V. Ho-Huu, T. Nguyen-Thoi, T. Truong-Khac, L. Le-Anh, and T. Vo-Duy, “An improved differential evolution based on roulette wheel selection for shape and size optimization of truss structures with frequency constraints,” *Neural Computing and Applications*, vol. 29, no. 1, pp. 167–185, 2018.
- [57] W. Chen and G. Feng, “Spectral clustering: a semi-supervised approach,” *Neurocomputing*, vol. 77, no. 1, pp. 229–242, 2012.
- [58] B. Wang, J. Zhu, E. Pierson, D. Ramazzotti, and S. Batzoglou, “Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning,” *Nature Methods*, vol. 14, no. 4, pp. 414–416, 2017.

Research Article

Enhanced Intelligent Smart Home Control and Security System Based on Deep Learning Model

Olutosin Taiwo ¹, **Absalom E. Ezugwu** ², **Olaide N. Oyelade**,²
and Mubarak S. Almutairi ³

¹*School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Westville Campus, Private Bag X54001, Durban 4000, South Africa*

²*School of Computer Science, University of KwaZulu-Natal, King Edward Road, Pietermaritzburg Campus, Pietermaritzburg, KwaZulu-Natal 3201, South Africa*

³*College of Computer Science, University of Hafr Al Batin, Saudi Arabia*

Correspondence should be addressed to Absalom E. Ezugwu; ezugwua@ukzn.ac.za

Received 1 October 2021; Revised 15 November 2021; Accepted 27 December 2021; Published 15 January 2022

Academic Editor: Tarek Gaber

Copyright © 2022 Olutosin Taiwo 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.

Security of lives and properties is highly important for enhanced quality living. Smart home automation and its application have received much progress towards convenience, comfort, safety, and home security. With the advances in technology and the Internet of Things (IoT), the home environment has witnessed an improved remote control of appliances, monitoring, and home security over the internet. Several home automation systems have been developed to monitor movements in the home and report to the user. Existing home automation systems detect motion and have surveillance for home security. However, the logical aspect of averting unnecessary or fake notifications is still a major area of challenge. Intelligent response and monitoring make smart home automation efficient. This work presents an intelligent home automation system for controlling home appliances, monitoring environmental factors, and detecting movement in the home and its surroundings. A deep learning model is proposed for motion recognition and classification based on the detected movement patterns. Using a deep learning model, an algorithm is developed to enhance the smart home automation system for intruder detection and forestall the occurrence of false alarms. A human detected by the surveillance camera is classified as an intruder or home occupant based on his walking pattern. The proposed method's prototype was implemented using an ESP32 camera for surveillance, a PIR motion sensor, an ESP8266 development board, a 5 V four-channel relay module, and a DHT11 temperature and humidity sensor. The environmental conditions measured were evaluated using a mathematical model for the response time to effectively show the accuracy of the DHT sensor for weather monitoring and future prediction. An experimental analysis of human motion patterns was performed using the CNN model to evaluate the classification for the detection of humans. The CNN classification model gave an accuracy of 99.8%.

1. Introduction

Home-based crime attacks, theft, and burglary are on the increase annually. Despite the lockdown and stay-at-home order, several homes were still attacked and burgled in South Africa [1]. According to the report in [1], attacks on homes and properties have aggravated into sexual offenses and assaults. This has shown a need for additional and improved

levels of security for homes, properties, and individuals. The growth in the industrial development of smart home automation systems and the rate of research works carried out in the field give assurance of using the field to solve security issues arising in smart home environments. A secured home will make occupants live in peace and without fear, irrespective of their location. Having remote control over the home, a clear view of the situation within the home, and a system that

notifies the home owner of impending danger without any false alarm is of high importance in this era. Advancements in information technology (IT) and the internet of things (IoT) have provided platforms suitable for the control, monitoring, safety, and security of homes and properties. The internet of things is the interconnection of several physical devices, network connectivity, and communication media to transmit information between devices and devices to humans [2, 3]. The IoT technologies have led to communication, interaction, and data exchange among devices, sensors, and appliances.

Moreover, the overall remote control of the home over the internet, Wi-Fi, Bluetooth, or other network modules via smartphones, laptops, tablets, and other related devices is now made possible, thus giving the desired comfortability to people in the applied area of IoT. Activities in the home and environmental conditions can be monitored and controlled through IoT technologies and are often referred to as smart home automation. According to [4], a smart home (SH) is a place of abode incorporated with a communication network that connects the major home appliances and services utilized for remote control, monitoring, and access from within or outside the premises. A smart home automation system is projected to deliver safety and home security services aside from regular home control. The smart home environment is embedded with intelligence to reduce human efforts in controlling and monitoring the house and its appliances. With ambient intelligent home control and security mechanism, the occupants can have a conducive environment to live in, excluding incessant notifications and alarms.

Artificial intelligence (AI) is contributing immensely to independent automated systems in the field of IoT. Home security is an essential aspect of smart home automation for convenient living. With the aid of AI models and specific IoT technologies, home premises and their conditions can be remotely monitored, controlled, and surveyed. With machine learning and deep learning models, systems can automatically make intelligent decisions on behalf of humans for enhanced quality living. Classification and detection of objects are the key strengths of machine learning and deep learning models. The application of AI in home automation will assist in the classification and detection of intruders in a smart home environment. Studies have shown that deep learning and machine learning models have been efficiently applied in smart home automation for object detection and recognition, human activity detection, facial recognition, intelligent control of appliances, energy efficiency, home monitoring, safety, and security [5–10]. A deep learning algorithm is a machine learning model that adopts the structure of a human brain to conclude data analysis with a given logical structure [11]. Although several AI models exist for the classification, detection, and prediction of objects, the deep learning model has been widely used and proved efficient for providing solutions to engineering, classification, and detection challenges [12].

A convolutional neural network (CNN) model is a subclass of artificial neural networks for image processing and recognition. The CNN model is widely used to solve image-based problems and is efficient with several convolution layers

[13]. The security mechanism that can be adopted in a home environment is based on motion recognition and surveillance. With the CNN model, images captured by the surveillance camera can be processed based on the area of interest for detection. A deep learning model-based intelligent detection can enhance a smart home automation system to classify detected motions into home occupants or intruders before sending an alarm to the user. Such an efficient home automation system design reduces stress, wastage of basic amenities such as electricity and water, and enhances the life quality. The smart home automation system controls lights, entertainment systems, environmental conditions, and other home appliances through ambient intelligence [14]. However, there are still some challenges in the smart home automation domain, such as wide-range connection for the control of the home, intelligent decision making by the system, storage, precise motion detection, and real-time data storage for future prediction, analysis, and decision making. To this end, we propose a real-time cloud-based, low-cost smart home automation system based on an Android mobile application. Note that by low cost, we mean to say that the prototype implementation of the proposed system using IoT hardware such as microcontroller board, sensors, cameras, and other components was set up using low-cost IoT hardware, which by implication is affordable and easy to configure.

To enhance security in a smart home automation environment and make meaningful research contributions, we propose a deep learning model (CNN) to classify human movement patterns as a security mechanism for the identification and classification of humans. The proposed classification of the movement patterns is to distinguish between regular home occupants and intruders in a smart home environment. The patterns of regular home occupants are prerecorded in the system, and once a motion pattern is detected, the system compares it with the existing motion patterns on the system and classifies it as either a regular home occupant or an intruder before raising the alarm. The approach aims to overcome the challenges such as false detection due to the use of masks and distortion due to weather or lighting associated with the biometrics approach in the existing literature. The home is controlled via an Android smartphone, and all the conditions of the home can be viewed on the GUI of the smart home automation application. The mobile application uses a platform as a service for its real-time storage of data generated by sensors and displays a graphical output of the ecological readings. In addition, the IoT hardware choice for the prototype implementation is based on low-cost, extensible, and accessible devices. In summary, the technical contribution of this paper is highlighted as follows:

- (i) A proposal for a deep learning (CNN) algorithm for intrusion detection in a secured smart home automation environment
- (ii) A design and development of an Android-based smart home automation system for the control and monitoring of electrical home appliances and environmental conditions

- (iii) A prototype implementation of an IoT, smart home system for home control, monitoring, and security
- (iv) Experimentation of a CNN-based deep learning model for classifying human walking patterns to detect intruders in a smart home environment

The remainder of this paper is structured as follows: a review of related works is presented in Section 2; Section 3 gives the proposed system architecture and functionalities and discusses the CNN-based home security concept. In Section 4, we present the experimentation and prototype implementation. Finally, the conclusion and future work direction are given in Section 5.

2. Related Literature

Research in the smart home automation domain has witnessed impressive progress, innovations, experimentation, and implementation. With the advancement in IT and IoT technologies, smart home automation systems now offer services beyond the initial home control and environmental monitoring. Machine learning and deep learning algorithms have been deployed to improve the intelligence responses in smart home automation systems. Some research lapses observed are the use of Bluetooth, GSM, SMS, Zigbee, and other communication modes for control which has the drawback of limited range coverage. Other research problems noted are the use of web-based design and SMS as means of appliance control. Lastly, few existing research works incorporated intelligent decisions into the systems. This section presents a review of related and relevant literature works in smart home automation and artificial intelligence. In contrast, Table 1 presents a summary of related works in smart home automation, focusing on the functionality, use of cloud computing, and other features compared to our work.

Xiaodong and Jie [15] presented the design of embedded smart home control and monitoring system using an STM32 microprocessor. Their system was designed for monitoring indoor environmental conditions (temperature and humidity) and home control. The authors used a GSM and GPRS module to communicate remote control and Zigbee terminal communication between the home devices. The implementation of their system was achieved through coordinated remote control and feedback on household appliances and the μ COS-II as an embedded real-time operating system. Bimenyimana et al. [16] designed and implemented a web-based home control system to control and monitor electrical appliances. Their system allows the user to manage electricity usage in the home by remotely turning off an appliance when not in use. A prototype implementation was displayed using Arduino board, Node MCU, Light-Emitting Diodes (LED), relays, and other hardware. A motor DC used in the design controls the light and reduces the energy consumption. However, the control is web-based, taking longer control time than a mobile-based application.

Gunawan et al. [17] presented a smart home control system for controlling home appliances via a website. The system uses temperature, smoke, and gas sensors to monitor environ-

mental conditions such as temperature and humidity, detect gas leakage, and ensure security using a PIR sensor for intrusion detection at the home gate. Communication between home appliances, sensors, and the Arduino board was established through an APC220 Wireless RF module. At the same time, an Ethernet Shield connects the board to the website interface to control the home. The wireless modules give a wider communication range, but a website is not as efficient for the control of the home as a mobile application in terms of time consumption. Similarly, a Wi-Fi-based smart home automation system was proposed by Singh and Ansari [18]. The system controls home appliances, monitors environmental parameters (humidity and temperature), and maintains home security through an alarm buzzer. A prototype implementation of the proposed method was presented using an Arduino microcontroller, ESP8266 Wi-Fi technology alongside sensors, and other IoT hardware. An application named Blynk provided the platform for control. In their system, home control access is granted by a designated administrator of the system.

Adel and Ali [19] presented a design and prototype implementation of a low-cost smart home system. The system was designed to control windows, doors, lights, electricity, and temperature in the home. A prototype implementation of the system was displayed using an Arduino board, servo motors, LED lights, temperature, and motion sensors. The door and windows were controlled using servo motors, a PIR motion sensor for tracking movement in the house, the INA219 high-side DC sensor to monitor the drop and supply of voltage, and a DHT11 temperature and humidity sensor was used for measurement of the temperature and humidity in the home. The system used Bluetooth to communicate, which is a limitation in terms of communication range.

An optimal and automatic control system for the control of electrical home appliances was proposed by Parsa et al. [20]. The proposed system's design is to control power consumption by automatically turning on and off the smart plugs connected to specific devices in the home at the appropriate time. An optimization approach was used to determine the appropriate time of use before the automatic switch was implemented according to the predefined constraints. The designed system favors the supplier more than the electricity consumer in the home.

Classification of human activities in smart homes using different deep learning models was presented by Liciotti et al. [6]. The deep learning models were applied sequentially to determine a specific action carried out by the home occupant at a particular time. The long short-term memory (LSTM) was applied for modeling temporal sequences in long-term dependency situations. The LSTM is combined with other machine learning algorithms to learn, recognize, and predict human actions such as bathing, walking, eating, relaxing, and sleeping and in a smart home environment. Similarly, Manu et al. [21] proposed an LSTM deep learning algorithm for performing predetermined tasks based on human activity recognition. The LSTM algorithm is integrated with IoT technologies to predict human activity towards home control, safety, and comfort for users.

The support vector machine (SVM) algorithm was implemented by Majeed et al. [22], for intelligent decision

TABLE 1: Summary of related works in smart home automation.

	Wireless technology	Functionality	Scalability	Cost-effective	Cloud computing platform	Notification on the lighting condition of the home	Intelligent decision making
Taiwo et al. [29]	Bluetooth and Zigbee	Control of home appliances with Android-based smartphone	No	Yes	No	No	No
Taiwo and Ezugwu [30]	Not discussed	Smart home automation	No	Yes	No	No	No
Majeed et al. [22]	Not discussed	Control light and home temperature with the use of a sensor via an Android-based smartphone	Yes	Yes	Yes	No	Yes
Alam et al. [31]	No	Control of lights and sensors in the home through a computer interface	No	Yes	No	No	No
Naing and Hlaing [32]	No	Control and monitoring of light through sensors	No	Yes	No	Yes	No
Jena et al. [33]	No	Control of lights through voice recognition	No	Yes	No	No	No
Taiwo and Ezugwu [34]		Controls home appliances and detects motion through an Android-based app	Yes	Yes	Yes	Yes	Yes
Mahmud et al. [35]	Wi-Fi	Control of home appliances and electronic machines through a web interface	No	Yes	No	No	No
Liao et al. [36]	Wi-Fi	Control of home appliances, monitoring of environmental conditions in the home, security of the home, and measurement of gas level in the home	Yes	Yes	Yes	No	No
MufHAS	Wi-Fi	Controls home appliances, monitors environmental conditions, and detects movements	Yes	Yes	Yes	Yes	Yes

making in a smart home. The intelligent control and status of home appliances were based on the SVM algorithm. Secured communication is also ensured using blockchain technology. An Android application was developed for the remote control of home appliances. Brenon et al. [23] presented a voice-controlled, context-aware-based home automation system enhanced with a deep reinforcement learning model. The reinforcement learning model is used for context extraction from graphical representation. Their approach also used a CNN for decision making in the home environment. Jaihar et al. [7] presented an approach that combines three AI models for security, intelligent control, and decision making in a smart home automation system. The authors used a CNN model, fisher face classifier, and SVM approach for facial recognition and detection in a smart home environment. Also, Neverova et al. [24] presented a DL approach for detecting human identification from motion patterns. The approach presented compares several DL models to learn human motion patterns efficiently for multimodal authentication systems. Feature and latency extraction was used for the learning representations. The comparison of the CNN-based models presented shows that motion patterns are effective for the authentication and identification of humans.

The existing literature has presented home security and safety mechanisms, models, frameworks, and experimentation. For instance, Saravanan et al. [25] proposed a Bluetooth-based safety and security system for home automation. A smartphone application was designed to control home devices, doors, and overall monitoring. Their system automatically switches off the lights at night, detects gas leakage or smoke, and controls home appliances. The system also locks and unlocks the door in the home automatically via an authentication module. Communication with the user's smartphone and the hardware is via Bluetooth. The method used a secured mechanism. However, the coverage range of Bluetooth is lower compared to Wi-Fi-based technologies. Jabbar et al. [26] proposed a Wi-Fi-based intelligent home automation system for home control, monitoring, and security. The system measures ecological factors around the home and displays the value, controls electrical appliances, and ensures home security through motion detection. An alarm is raised by the system when a motion is detected to alert the user of an intruder. Also, the system can support a reduction in electricity bills through remote control of appliances. However, there is no intelligent module to distinguish between the detected motions.

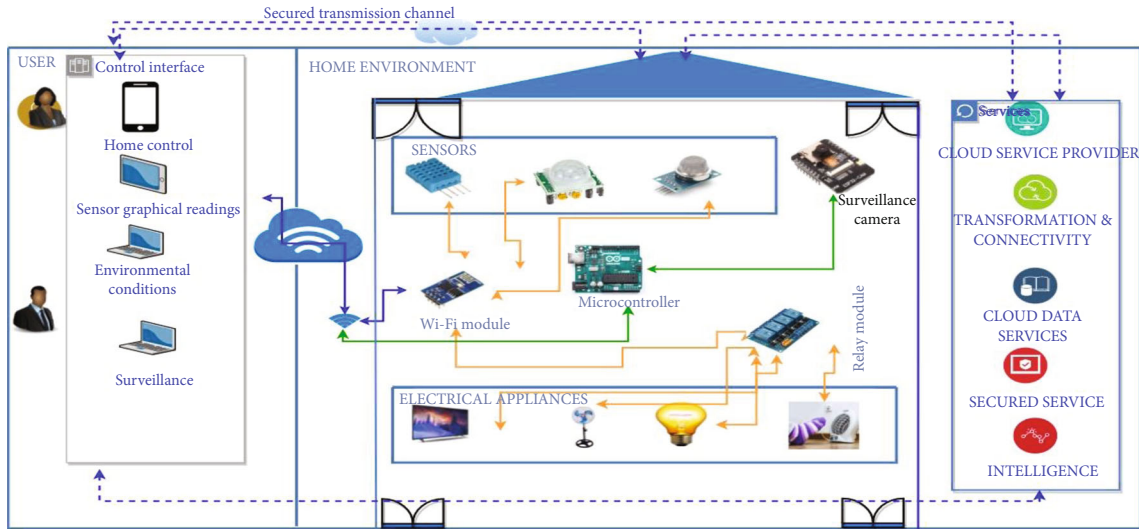


FIGURE 1: The MufHAS Architecture.

A security system for a smart home automation system was presented by Ajao et al. [27]. The system secures entry points of the home and grants access through authorization and authentication. An in-app message notifies the user of access or denial of access. Sensors are also embedded in the system for motion detection. The home security system is controlled through wireless IoT communication and Android mobile application. The presented approach is easy and flexible to use, but capturing an image of intruders is not included for surveillance. Singh et al. [28] proposed a smart home automation system for the safety and control of electrical appliances, doors, and movement detection in a house. The system could also send an alert to the user in response to sensors detecting low levels of gas in a cylinder or the presence of a human. A prototype implementation of the system was carried out using an Arduino Uno board, a Node MCU ESP8266, IR, and LDR sensor modules. A summary of the studies reviewed in this section is listed in Table 1. A comparative presentation of features, methods, and functionalities was included in the summary.

3. System Architecture

The reviewed literature has clearly shown that artificial intelligence greatly contributes to the enhancement of intelligent response, detection, and decision making in smart home automation systems. However, there are still associated with exact motion detection, aversion of false and frequent notifications by home automation systems, especially in the security domain of the field. Therefore, we have been motivated to design and develop an intelligent smart home automation system with various functions and a multifunctional smart home automation system (MufHAS). Also, it is expected that an efficient home automation system must be intelligent to make decisions on behalf of the user to reduce disturbance and human interference. Therefore, a CNN deep learning model is proposed to enhance our system's home security by detecting abnormal intruder movement. In this

section, we describe the architectural design of our MufHAS to control home appliances, monitor and measure environmental conditions, and detect movement in the home. An effective smart home automation system is known to offer assistance to the disabled, sick, or elderly [30], control home appliances remotely [37, 38], manage energy [39, 40], and measure and monitor environmental conditions [41, 42]. It also provides home security to its user by detecting unauthorized movement or intrusion [43, 44]. One of the major objectives was to use the emerging IoT technologies to provide a suitable medium to achieve most of the vital roles a smart home automation system must offer its users. Smart home automation systems' major functional components are smart home technologies, sensors, appliances, microcontrollers, and control mechanisms. An IoT smart home automation and its components are networked together either wired or wirelessly for automated control via a specified medium. With the help of specific IoT sensors, convenience, the safety of lives and properties can now be ensured in a smart home automation environment. For example, motion sensors can alert the user via the system about movements in the home. The user can conveniently switch off lights in the house from any location with the aid of an intelligent notification from the system, thus conserving energy. In recent times, an effective smart home automation system has improved beyond the conventional control of heating, ventilation, and air conditioning appliances in the home to overall automation with the incorporation of home safety and security mechanisms. The overall architectural design of the home is depicted in Figure 1.

3.1. MufHAS Modules. The system has three major modules, the user's side, the home environment, and the backend that comprises the database module for storing data generated from devices in the home and the intelligent module for ensuring security in the home. The user communicates wirelessly with the home environment over the internet using Wi-Fi. The home environment comprises sensors, detectors,

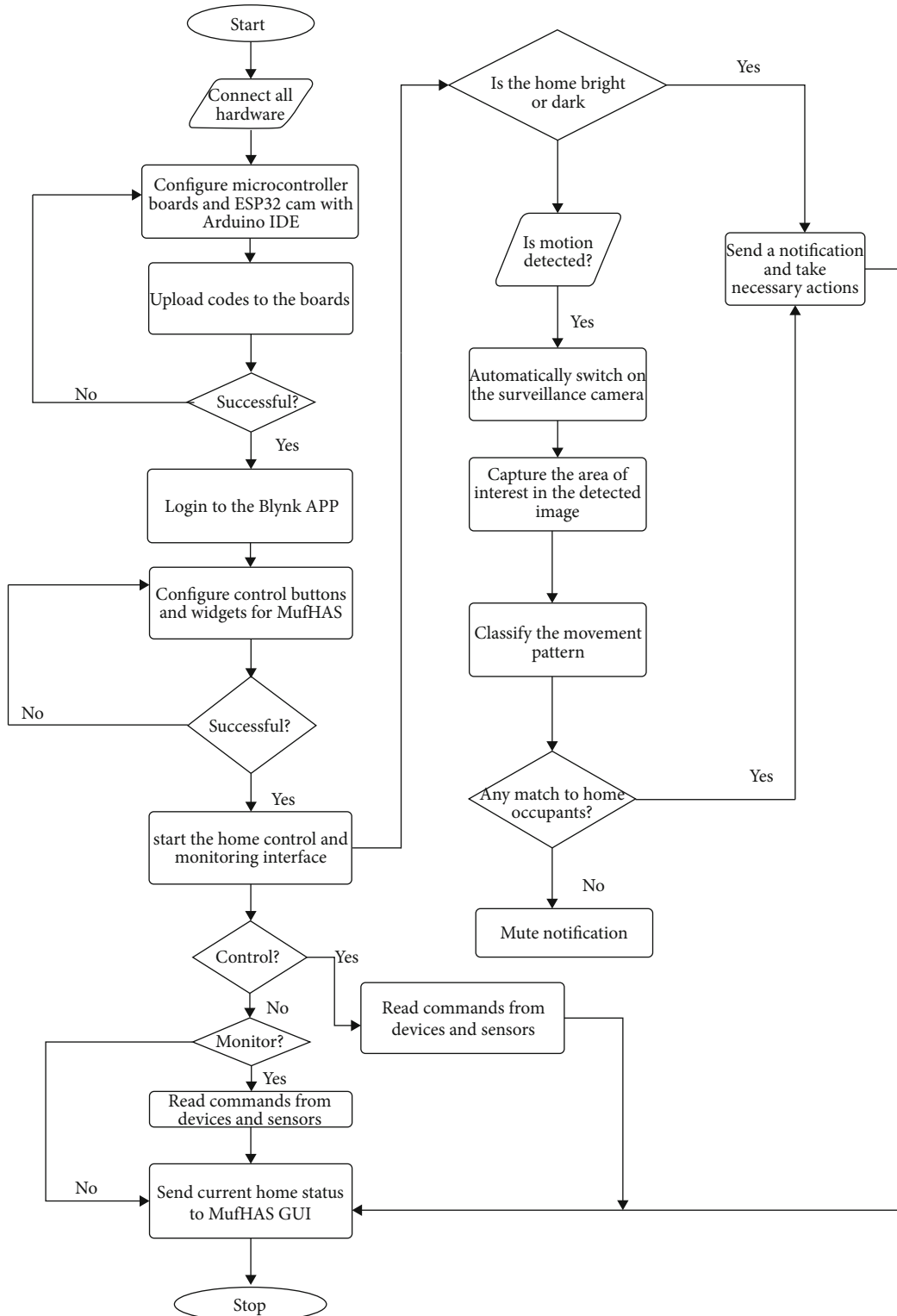


FIGURE 2: Flowchart of MufHAS.

home appliances, a surveillance camera, and a communication gateway. The architectural design is further explained with the prototyping in Section 4. The microcontroller board is responsible for delivering services issued through a command from the user's phone to the home environment. A

real-time database hosted in the cloud is used to store data generated by devices and sensors in the home. The services rendered by cloud computing platforms complement the smart home automation system. Thus, data generated from an intelligent home environment should be stored to know

the performance of the installed devices, especially sensors. Also, prediction and analysis can be carried out based on the data generated in a smart home to enhance systems to be developed and to know the changes in the home's environmental conditions. The process workflow of MufHAS is presented in Figure 2.

3.1.1. User's Module. On the user's side of the system, control and monitoring of the home take place via the GUI of an Android-based smartphone. The interaction between the user and the home takes place at this end. With the designed Android application, users can easily have a real-time view of environmental conditions (temperature and humidity), change the status of an appliance (on or off), live stream the home's surveillance, and perform other tasks as desired. With the cloud computing platform integrated into the mobile application, users can view a graphical chart of the sensors in the home environment. The data generated by the sensors are also stored in the cloud for future use. Communication between the user and the home environment is over the internet. Control of the home is done over the mobile application. However, the surveillance can be streamed from laptops, tablets, or desktop computers.

3.1.2. The Home Environment. The home environment comprises smart home appliances, sensors, a Wi-Fi module, and a surveillance camera. Bulbs, fans, television, and sockets are the primary home appliances considered in our design. The IoT devices incorporated into the design are the temperature and humidity sensor, motion sensor, and camera. The home appliances, devices, and sensors are connected to a microcontroller for communication with each other and the outside environment over a wireless network. The ESP8266 functions as the microcontroller and Wi-Fi module. The communication protocols used are Wi-Fi, one of the primary operating standards for home automation technology, TCP/IP, and HTTPS/IP. The ESP32 camera module is interfaced with an Arduino board as it requires its own microcontroller for proper functionality and power supply. The ESP32 camera board has an in-built Wi-Fi chip for wireless connectivity. The ESP32 and ESP8266 are connected to the same network for seamless communication.

3.1.3. The Backend Module. The backend comprises the intelligent module that oversees the security of the home and the cloud computing platform for storage. The home's security is enhanced with a deep learning model (CNN) for the detection, classification, and notification to the user about the presence of a human or other objects in the home environment. The deep learning model classifies the detected object as a regular home occupant or an intruder based on the movement pattern. The motive behind the classification is to avert frequent notifications and false alarms. Based on the intelligent classification of the detected object, a notification is sent to the user for necessary action. Steps involved in the connection, communication, operations, and services of the MufHAS are presented in Algorithm 1.

3.2. MufHAS' Functionality. This section explains the components of our home automation system. As previously

stated, our system is made up of home appliances, sensors, and an Android smartphone to control and monitor the home. The user communicates and monitors the house over the internet via an Android smartphone. The major functions of our designed smart home automation system are control and monitoring of essential home appliances, measurement of environmental conditions, and motion detection. The proposed system is based on wireless network connectivity, and all its functionalities are further explained in the next subsections.

3.2.1. Control and Monitoring of Home Appliances. The designed MufHAS allows essential home appliances to be controlled within and without the home premises. The user controls the home over the internet outside the home and remotely while within the home. This design consideration makes the system function in a dual mode (remote and global control). Home appliances factored into our system's design are bulbs, air conditioners, sockets, fans, heater, television, fridges, and switches. Control of the home is through an Android mobile application. The user registers to be on the platform to generate login credentials for subsequent home communication. Control over the internet while away from home will reduce the consumption level of amenities and thus impact the economy. For instance, if the user forgets to switch off the light before leaving home, the light can be switched off from any location with global control. The remote control is factored into our consideration to assist the user in reducing the amount spent on internet data as our system is designed to be low cost.

3.2.2. Control and Measurement of Home Conditions. Sensors are installed in the home to measure environmental conditions within the home premises. A DHT11 is a sensor that covers a wider range of temperature and humidity. It is also accurate and more precise in reading. In our design consideration, if the home's temperature is higher than the desired level, the fan is automatically switched on, and if the temperature is lower than the desired level, the heater is switched on. For enhanced healthy living, smoke and fire detectors, carbon monoxide detectors, gas sensors, and air quality sensors are considered. Carbon monoxide is odorless and colorless; thus, it is termed a silent killer. The carbon monoxide detector measures the level of carbon monoxide in the home environment. If a higher concentration is detected, a warning is sent to the user via the mobile application to avert inhalation of poison that can harm the body system. The designed system not only caters to convenience but also for enhanced healthy living.

3.2.3. Security and Detection of Movement. The security of the home against intrusion is of high importance. With the installation of door and window sensors, users can have the home secured against break-ins. MufHAS has sensors installed to detect movement in the home environment. An IoT camera serves as a surveillance camera for capturing intruders before notifying the user to improve the home's security. The user is notified about leaving home, and the doors or windows are unlocked. The system performs an

```

1: Begin
2: Define  $N_c$  parameters
3: Initialize EHA and HSD
4: Establish and confirm the status of  $N_c$ 
5: If  $N_c = 1$ 
6:   Evaluate the initial state of  $Ha$ ;  $\forall EHa \in N_c$ 
7:   if  $Ha = n$  (where  $n$  = number of configured home appliances)
8:     Start MufHAS
9:   Else, go to step 4
10: End if
11: if not ( $N_c \&\& MufHAS = 1$ )
12:   go step 4
13: Evaluate the initial state of  $Hs$ ;  $\forall Hs \in N_c$ 
14: If  $Hs = n$  (where  $n$  = number of home sensors and detectors)
15:   Connect MufHAS to the internet
16:   Acquire sensor data
17: Else, go to step 4
18: If is_connected(MufHAS)
19:   Get the values for  $T$ ,  $H$ , and motion
20:   Upload data to CS via MufHAS
21:   Update status of  $Hs$  in MufHAS
22:   Display graphical status of  $Hs$  in MufHAS
23:   Synchronize data to CS
24: Else, go to step 12
25: End if
26: Case 1: (LDR)
27: if ( $D=1$ ), then
28:   Notify the user, "It's DARK, Turn on the LIGHTS."
29: Else
30:   Notify the user "It's BRIGHT, Turn off the LIGHTS."
31: break;
32: Case 2: (Home security)
33: Ensure the camera is ON
34: If  $M$  is detected,
35:   Notify via iHOCS and apply SVM
36: If  $M \in (HWp_1, HWp_2, HWp_3, \dots, HWp_n)$ 
37:   Mute alarm
38: Else,
39:   Notify user via email "TOSIN: Motion detected"
40:   Raise alarm and send picture to email
41: end if
42: User monitors  $Ha$  and  $Hs$  via MufHAS app
43: Remotely control the home
44: End

```

ALGORITHM 1: MufHAS (home control, monitoring, and security) algorithm

intelligent check with the aid of a CNN model by cross-checking the detected movement with the predefined movements in the database. If the movements are that of home occupants, then the system does not raise the alarm. Else, an alarm is sent to the user with the captured image. The PIR sensor has trigger sensitivity with an adjustable duration of the trigger signal, allowing the user to set the desired trigger sensitivity. Motion recognition and pattern could help identify suspicious and malicious activities [45], thus preventing intrusion.

3.3. Deep Learning-Based Home Security. Home security and safety are of great importance to the well-being of home occupants. Therefore, a smart home automation system

should have a management scheme for home security via the control module with alarms or alerts [46]. With motion recognition and detection, the smart home environment can be secured against intruders while minimizing false alarm rates from the system. We propose a home security module for movement recognition, classification, and detection. Human movement patterns can uniquely identify a person as walking patterns differ from person to person [47, 48]. IoT sensors are widely used for the collection of data from the environment. Motion detectors, sensors, and cameras are applicable in gathering information about a human's motion pattern and activity. Thus, biometric verification of humans is possible through their movement patterns [49]. A smart home environment can therefore be secured using

motion sensors and surveillance cameras. Our work proposes a home control and security system based on motion recognition. We used this concept because it is faster and discrete for identification as against other biometric means of identification [49]. Also, other human detection mechanisms are based on images that compare skin tones, skin color, eye color, and other facial attributes. Although facial detections prove effective, several factors can influence the accuracy of facial detection in a home security domain. Such factors are lighting, weather, brightness, use of facial masks, hoods, and so on [50]. With motion sensors, movement in an environment can be silently detected, captured, and verified. The proposed security module of our system performs operations by capturing movements in the home and comparing them with the predefined motion patterns (home occupants) before raising the alarm.

A CNN is a network of convolutional and pooling layers to extract main features from an input towards the desired output. CNNs are applicable in domains such as facial recognition [51], image classification, image and video recognition, motion recognition [52] recommender systems, medical image analysis [53], and other classification and decision-making domains. The mathematical definition [54] for the image classification of the CNN model represented as a tensor is given as follows:

$$\dim(\text{image}) = n_H n_W n_C, \quad (1)$$

where n_H is the size of the image height, n_W is the size of the image width, and n_C is the number of channels.

The filter is calculated as

$$\dim(\text{filter}) = (\mathfrak{f}, \mathfrak{f}, n_C). \quad (2)$$

\mathfrak{f} denotes an odd dimension.

From equations (1) and (2), the filter of an image is

$$\text{conv}(\mathcal{I}, \mathcal{K})_{x,y} = \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} \sum_{k=1}^{n_C} \mathcal{K}_{i,j,k} I_{x+i-1, y+j-1, k}. \quad (3)$$

The CNN is combined with a Softmax classifier for the classification of the extracted feature. A Softmax classifier is a linear classifier used in deep learning for the classification of vectors and to determine the probability of the extracted [55]. The Softmax is also known as multinomial regression and can be used for mutually exclusive multiclass classification. It has been widely used in deep learning, yielding excellent performance [56, 57]. The mathematical definition of Softmax is given as

$$\sigma\left(\vec{\mathcal{Z}}\right) = \frac{e^{\mathcal{Z}_i}}{\sum_{j=1}^{\mathcal{K}} e^{\mathcal{Z}_j}}, \quad (4)$$

where (\mathcal{Z}) is the input vector to the Softmax function made up of $(z_0 \cdots z_k)$, \mathcal{Z}_i values are the elements of the input vector to the Softmax function that can take any real value, $e^{\mathcal{Z}_i}$ is the standard exponential function applied to each element

```

1:  Begin
2:  Define  $N_c$  parameters
3:  Initialize  $M_s$  and  $Cam$ 
4:  Establish and confirm the status of  $N_c$ 
5:  If  $N_c = 1$ 
6:    Evaluate the initial state of the home
7:    If the camera and motion sensor are active
8:      Start MufHAS security
9:    Else, go to step 4
10:  end if
11:  while  $M_s$  is ON
12:    for each object detected do
13:      if cam is on sleep mode;
14:        trigger ON
15:      else
16:        Continue streaming
17:      end if
18:      Capture image
19:      Extract region of interest
20:      Apply the trained CNN model
21:      if  $O_d \in (HWp_1, HWp_2, HWp_3, \dots, HWp_n)$ 
22:        Save to cloud
23:      else,
24:        Raise an alarm and save the captured picture
25:      end if
26:  End

```

ALGORITHM 2: CNN-based home security algorithm

of the input vector, and \mathcal{K} is the number of classes in the multiclass classifier

Algorithm 2 presents the steps involved in our movement-based object detection and classification using the CNN deep learning model. The workflow of the home security and image processing based on the CNN model is presented in Figure 3. The CNN model is used majorly to classify human and pet movements based on the motion pattern. The object in our context refers to human beings and pets as both occupy a house. Thus, either of these two can trigger an alert while moving. The motion patterns considered are walking, jumping, limping, and running. In our work, the surveillance camera captures and records the activities in the environment. However, the ESP32 camera used in the prototyping does enable a sleep mode if there are no records of activities. Hence, the combination with a PIR motion sensor. If a movement is noticed, the PIR sensor sends a signal to the camera, and it comes on. The camera, in turn, captures the region of interest (movement pattern) in the detected image, classifies and compares it with the existing and predefined set of movement patterns in the database. If the detected pattern matches the existing ones, an alarm is not triggered. Otherwise, the user is notified via the mobile application of the house's situation.

The CNN architecture applied for detecting intruders in the smart home framework is presented in Figure 4. Four (4) convolutional blocks are composed to build the CNN architecture, interleaved with zero padding and max pooling operations. This composition allows for a comprehensive and fine-grained feature extraction procedure, leading to

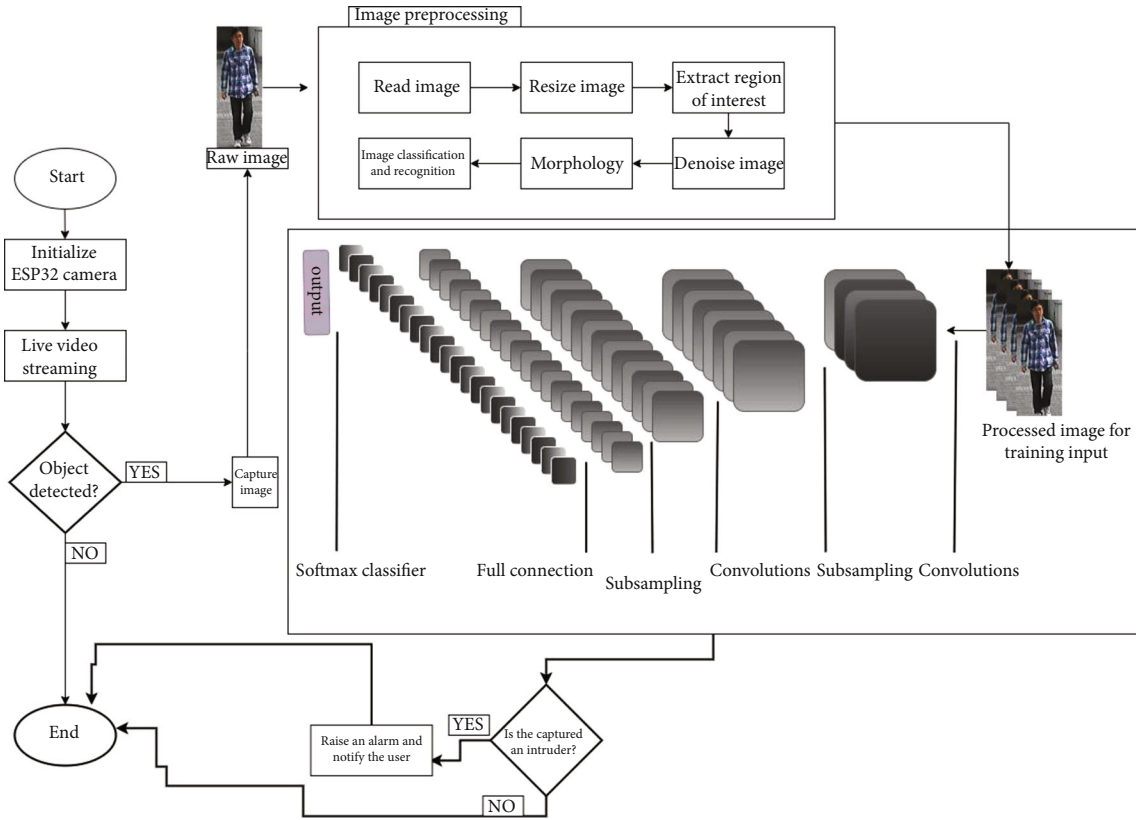


FIGURE 3: Flow process for CNN-based home security.

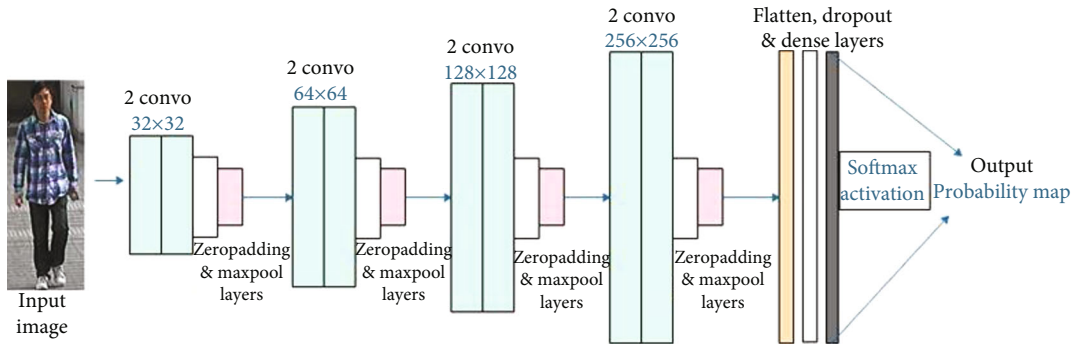


FIGURE 4: An illustration of the CNN architecture for intruder detection in the MufHAS architecture.

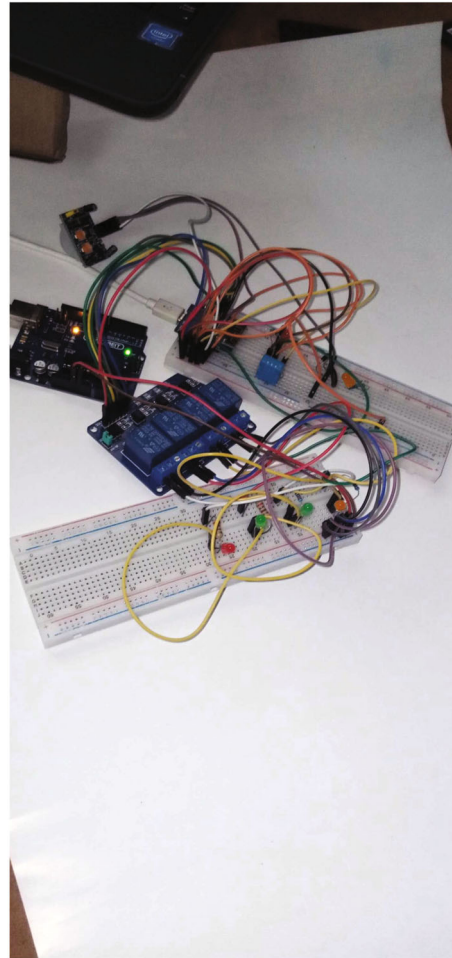
the Softmax activation classification process. The preprocessing of images is an essential aspect of making images fit for the modeling process. The initial process involves a range of steps applied to the input to achieve an acceptable form for the feature extraction. Preprocessing is very important because feature extraction might not yield desired results if the inputs are not properly preprocessed [58]. As a result, inputs were preprocessed using the denoising and CLAHE operations before being passed into the CNN architecture. The former preprocessing operation allows noise removal while the latter improves image quality through a contrast improvement strategy.

Filter counts of 32, 64, 128, and 256 were used for the convolutional layers' first, second, third, and fourth blocks.

Kernel sizes of 3×3 were used in each convolutional operation spanning the four blocks. After each block of the convolutional operation, we zero-padded the output before passing it to the max pooling operation. The peculiarity of the problem being addressed in this study showed that the max pooling operation demonstrated better performance than the average pooling operation. The padding size of 1×1 was used in the zero-padded layer, and the kernel size of 2×2 , 3×3 , 2×2 , and 3×3 was used for the max pooling layers of first, second, third, and fourth convolutional blocks, respectively. The strides of 1×1 were applied to all convolutional operations while the size of 1×1 and 2×2 was alternated in the max pooling layers across the architecture. We applied some regularization techniques to the CNN model



(a)



(b)

FIGURE 5: (a) Home control interface. (b) Initial state of the demo.

to eliminate overfitting the model. First, we used the L2 with a value of 0.0002 in each convolutional operation and dense layer and added a dropout layer of 0.5 rates after the flatten layer. Considering the multiclass nature of the dataset to which the CNN architecture was applied, we used the Soft-max activation function to obtain the probability map for each class of image samples in the dataset.

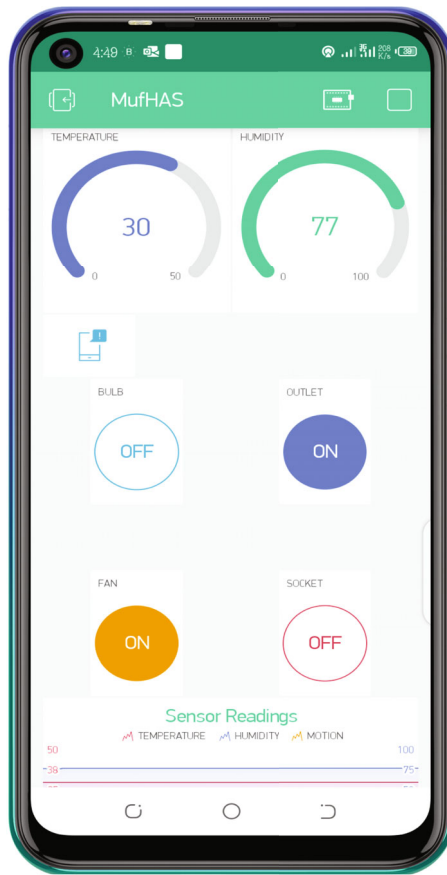
4. Implementation Details

Our system was implemented using a prototype setup. The prototyping was carried out with the use of hardware and software components. The software components we used are the Blynk software and Arduino IDE. The Android-based mobile application was designed and configured with the Blynk platform. The Blynk platform is an IoT platform for designing and configuring mobile applications for IoT systems. We used the Blynk platform to design the GUI of our mobile application and configure the pins of our hardware for home control and monitoring. It supports cloud data storage, display, and visualization of real-time data generated by sensors. The instructions for the configuration of the microcontroller board, sensors, and other home appli-

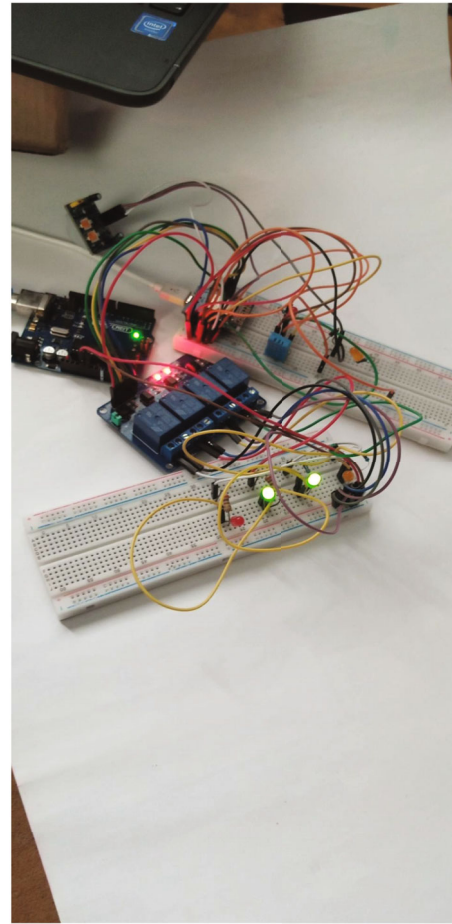
ances were developed using the Arduino IDE. The Arduino IDE is a cross-platform application for Windows, Linux, and MAC operating systems. Commands to control home appliances were also developed and uploaded to the board via the Arduino IDE.

4.1. Hardware Components. The hardware components of our system are an ESP8266 Wi-Fi module, an ESP32-CAM module, an Arduino UNO board, a 5 V four-channel relay module, jumper cables, breadboards, USB cables, LEDs, resistors, and sensors. Jumper cables (male to female and female to female) were used to connect two or more hardware to each other. The resistors were used to reduce current flow, while the breadboards were used to link the hardware components. The major hardware components are further explained.

4.1.1. ESP8266 Wi-Fi Module. The ESP8266 Wi-Fi module is a self-contained system on chip (SOC) with an integrated TCP/IP protocol stack. It is a low-cost Wi-Fi module that gives most microcontrollers access to Wi-Fi networks, thereby enabling the connectivity of several home automation technologies and devices. It allows the developed system to communicate with the internet and grants access to the



(a)



(b)

FIGURE 6: (a) GUI's command to switch on two appliances. (b) System's response to power on two appliances.

home to the user from anywhere. The storage capability of the ESP8266 Wi-Fi module allows it to be integrated with sensors and other home devices. The ESP8266 served a dual function of the microcontroller board and Wi-Fi module in our work.

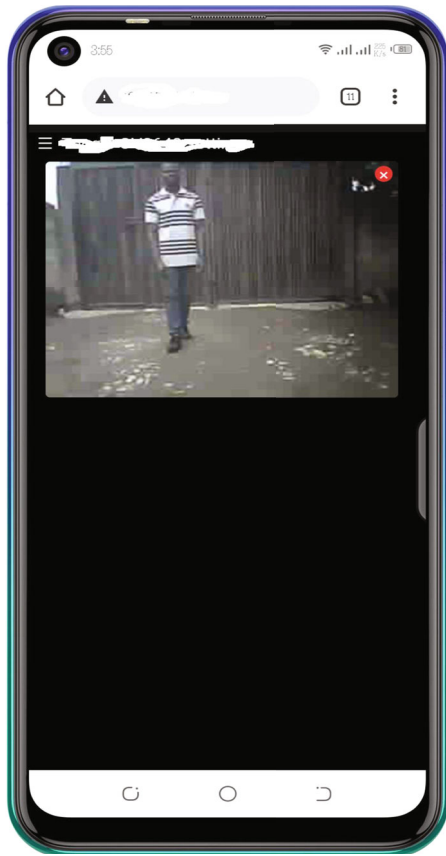
4.1.2. Relay Module. A 5V four-channel relay module was used in building the prototype MufHAS. A relay is an electrical device that controls devices and can switch much higher voltages and currents to normal microcontroller output using low voltage as input. The relay module is used to control a large current from the microcontroller. Each channel of the relay module has an LED indicator to show that the specified port is on.

4.1.3. DHT11 Temperature and Humidity Sensor. The DHT11 sensor was used to measure the home's temperature and humidity levels. The DHT11 has a fast response and interference ability. A DHT11 sensor is a low-cost, embedded sensor that serially provides temperature and humidity values via wire protocol. The humidity range for a DHT11 sensor is 20% to 80% relative humidity, but there are claims

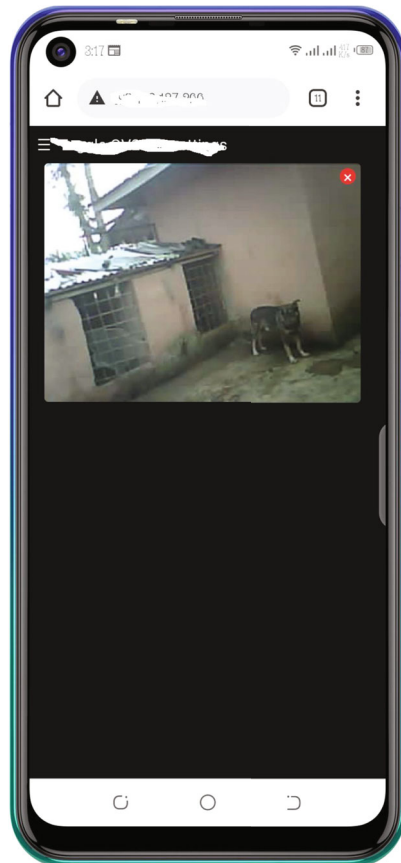
of up to 95% in some datasheets. The temperature range for the DHT11 sensor is 0 to 50°C [59].

4.1.4. HC-SR501 PIR Motion Sensor. An HC-SR501 is a motion-detecting sensor used in an IoT environment for detection motion and used in security systems, garage doors, gates, and automatic switching of lights in some systems. We use the PIR sensor to detect movements in our system, and it sends a notification to the user immediately for an action to be taken. It is a low-cost, sensitive, easy-to-use, and accurate sensor for motion detection. It supports a wide-range communication that has an in-built voltage regulator.

4.1.5. ESP32-CAM. An ESP32 camera module is a low-cost, low-energy development board for video streaming and image capturing in IoT prototyping. It has a combined Wi-Fi and Bluetooth chip based on a 32-bit CPU. The camera module has a resolution of up to 1600 by 1200 and supports up to 4 GB storage with an SD card. We use the ESP32-CAM to ensure security in prototyping our work through video streaming. The user can view the situation of the home once an alert of motion is received via the phone interface. To program the camera board, an Arduino UNO board was

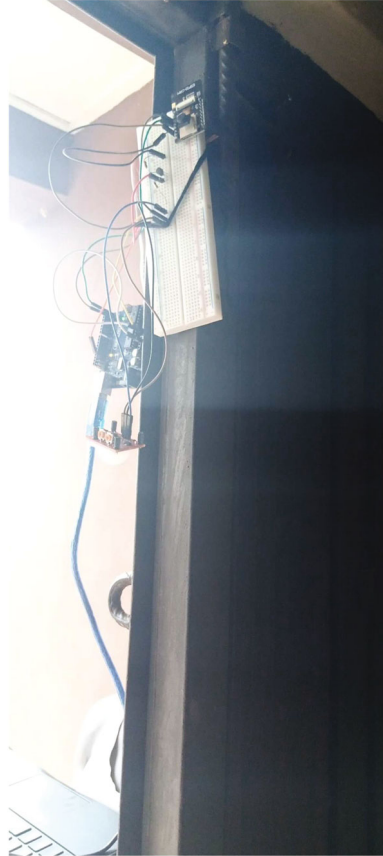


(a) Video streaming interface (human captured)



(b) Video streaming interface with an animal captured

FIGURE 7: Continued.



(c) Surveillance setup

FIGURE 7

used for the USB-to-TTL module because the ESP32-CAM does not have a programmer chip.

4.2. Datasets and Training of the CNN Architecture. The home training dataset consisting of about 4000 samples of different human postures was applied to the CNN for the training process. We found four major postures suggestive of what the camera module in the MufHAS architecture is expected to capture for image recognition. For each posture, samples were drawn into groups demonstrating an intruder and the home occupant. This resulted in eight classes for the multiclass classification used in the CNN architecture. Using the ratio of 80 : 20, samples were split for training and testing the CNN model. Adam optimizer was then used to train the CNN model with the configuration of 0.001 learning rate, 0.9 for beta1, 0.999 for beta2, and epsilon of $1e-8$. The training of the CNN model was carried out for 100 epochs under the computational environment with the configuration of Intel (R) Core i5-7500 CPU 3.40 GHz, 3.41 GHz; RAM of 16 GB; 64-bit Windows 10 OS.

4.3. Results and Discussion. We present a detailed explanation of our system's setup, configuration, and experimentation in this section. A prototype implementation using IoT hardware was used to test the functionality of our system.

The designed MufHAS was tested using an ESP8266 Wi-Fi module, a PIR motion sensor, an ESP32-CAM module, an Arduino UNO board, breadboards, a 5 V four-channel relay module, a DHT11 sensor, and an Android-based smartphone. The microcontroller board, camera module, and mobile application were programmed to communicate on the Arduino IDE. Network credentials for connectivity were declared in the coding phase of the configuration. A wireless connection was established through the Wi-Fi module. The prototype system detects movement, has a video streaming module, and measures the home environment's temperature, and humidity controls a bulb, fan, and socket, and an extra point was left out for scalability. The home appliances were interfaced with the relay module to have the appropriate current flow. The mobile application receives a signal from the system through the Wi-Fi module. The initial state of the system and mobile application are presented in Figures 5(a) and 5(b). All appliances were yet to be powered on. The system starts recording the measured environmental conditions once communication is established.

When the user issues a command on the mobile application, the corresponding home appliance gives an indicator on the LEDs and relays board to depict the switched device on or off. Figure 6(a) gives the GUI of the mobile application, indicating a command to switch on two of the

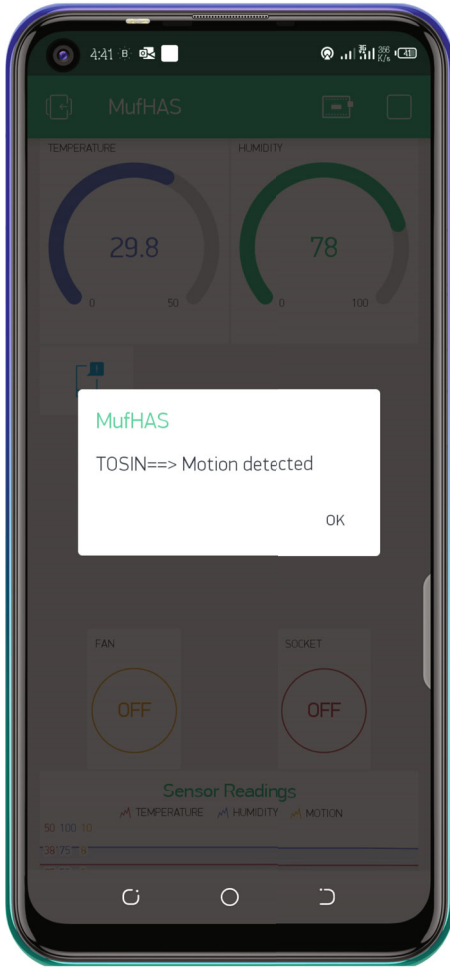


FIGURE 8: Movement notification.

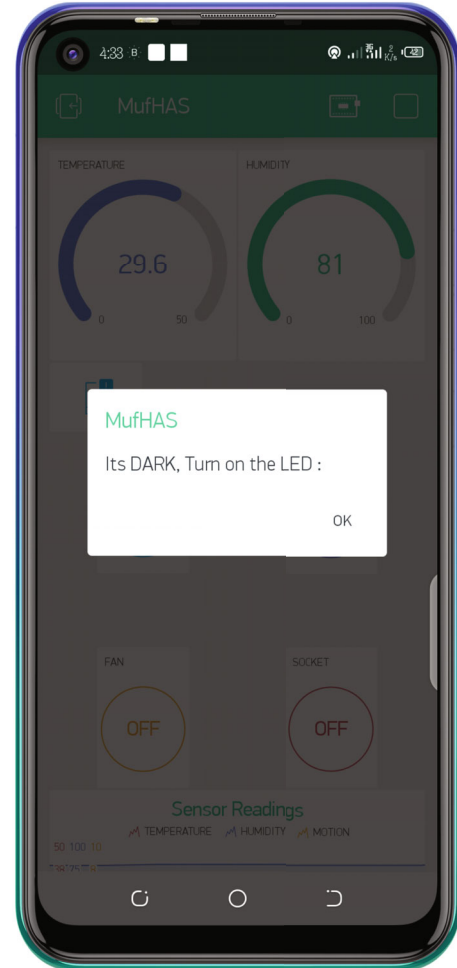


FIGURE 9: Intelligent notification about the light.

appliances, while Figure 6(b) shows the prototype's indicator. Time to establish communication between the mobile application and the smart home shows the system is effective in terms of connectivity even when not close to the system. Also, the humidity and temperature levels are measured at intervals by the system. The mobile application indicator reveals that the user can gather information about the home's environmental conditions and the level of the measured parameters. The user monitors the home environment over the internet via the video-streaming interface. The camera is placed at the entrance of the home to monitor the surroundings, record, and capture events. With the aid of the SD card module embedded in the camera module, recordings of the home environment are stored. Also, the user can capture images as they appear on the screen. The video streaming interface of the security module of our system, with a human and an animal captured, is shown in Figures 7(a) and 7(b), while Figure 7(c) shows the hardware setup. As explained earlier, movements are detected in the house, and a notification is sent to the user for necessary

action. Figure 8 shows the GUI of the mobile application with the displayed notification about a movement detected in the house. Our system also gives an intelligent notification to the user to switch on the light when it is dark or switch off the light when it is bright. This is to ensure energy efficiency in the home. The system was not designed to automatically turn on or off the light because the user might be on the premises and want the light on at that instance. Figure 9 shows the GUI of the notification to switch on the light. The mobile application displays a graphical reading of the sensors installed in the home for the user to visualize the trend of the environmental conditions. These readings are displayed based on real-time data acquired from the sensors and stored in the cloud platform of the Blynk application. The data generated can be used for weather prediction and analysis of the environment in the future. The graphical display is presented in Figure 10.

4.4. CNN Image Classification Experimentation. The proposed CNN model is tested to evaluate the feasibility and efficiency of our proposed method for home security. The

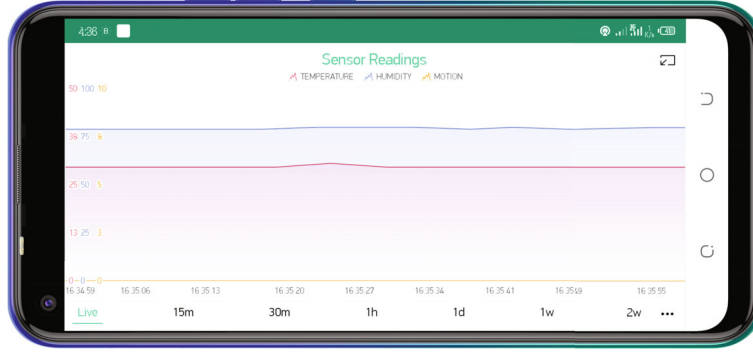


FIGURE 10: Graphical display of sensor readings.

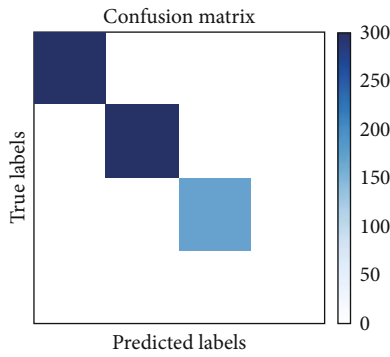


FIGURE 11: Confusion matrix of the trained CNN model when applied for prediction on some samples for the classification problem of detecting home occupants and intruders.

dataset for the training was downloaded from the CV image database [60], a dataset containing 3,884 images of the human motion of 972 pedestrians. Each person in the dataset was captured from four different angles to show different walking patterns. We divided the dataset into four groups: walking, jumping, limping, or running. Each group contained similar captured angles of different people in the same position. After the grouping, we separated the groups into two other subgroups that we termed home occupants and intruders for the classification process.

The CNN model yields an accuracy of 98%. A performance measurement required in DL classification is the confusion matrix. It presents an output table for the classification model, combining the predicted and actual values. The confusion matrix is shown in Figure 11. The accuracy of the CNN deep learning experiment shows that smart homes can be made more intelligent for the security of home occupants and their properties. The classification proves that motion patterns can be used to distinguish and identify entities in the home environment discretely and efficiently.

The performance of the CNN architecture during training is captured in Figures 12 and 13. The curve of the loss values obtained for the training and validation samples during the 100 training epochs is graphed in Figure 12. In the first 20 epochs, we see the unstable pattern of the valida-

tion samples as applied to the CNN model while the training dropped steadily. Interestingly, the result obtained in epochs 40–100 demonstrates that the loss values for the training and evaluation phases are in tune and that the CNN model effectively learns the classification problem. Similarly, the accuracy of both the training and evaluation phases rose consistently in alignment with the loss values sustained. The accuracy values for the two phases showed a stable output around 60-100 epochs of the training process.

Using the trained model, test samples were served to it, and the model's performance was evaluated using the accuracy, precision, recall and F1-score, specificity, and Cohen's kappa metrics. Table 2 presents the evaluation table using the precision, recall, F1-score, specificity, and Cohen's kappa values to compare the classifiers that were compared.

The precision evaluates the model's relevancy and quantifies the prediction about the number of positive values in the model. It is calculated as

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \cdot (5)$$

The percentage of rightly classified actual positive predictions in the model is evaluated by calculating the recall as

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \cdot (6)$$

The F1-score is a measure of accuracy based on the precision and recall values of a test. It is calculated as

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \cdot (7)$$

Specificity is used to ascertain the proportion of actual negative cases that the model rightly predicted. Specificity is calculated as follows:

$$\text{Specificity} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}} \cdot (8)$$

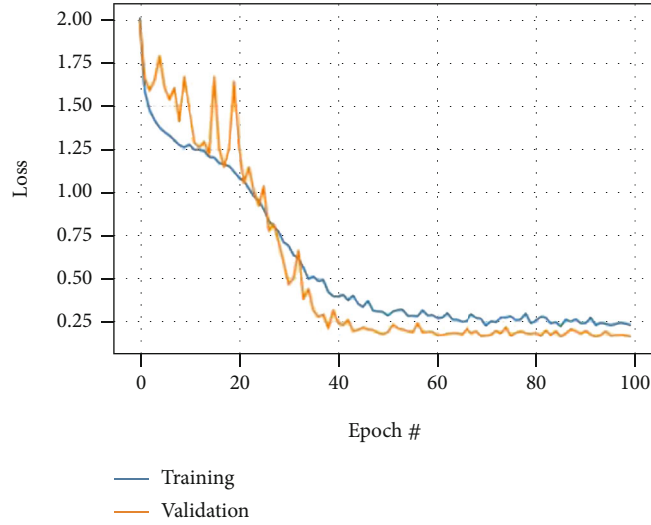


FIGURE 12: Training and validation loss obtained during training of the CNN architecture in the classification of walk postures of home occupants from intruders.

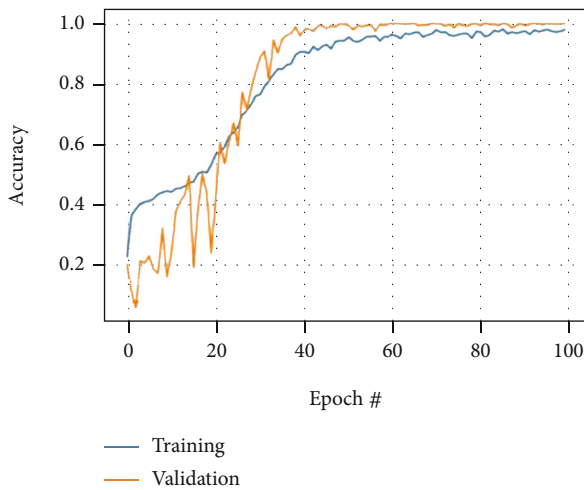


FIGURE 13: Training and validation accuracy obtained during training of the CNN architecture in the classification of walk postures of home occupants from intruders.

Cohen's kappa (K) is a measure of agreement between two raters in a test and is also used to assess a classification model's performance. It is calculated as follows:

$$K = \frac{\rho_0 - \rho_e}{1 - \rho_e}, \quad (9)$$

where ρ_0 is the overall accuracy of the model and ρ_e is the measure of the agreement between the model's prediction and the actual class values.

The prediction procedure on the trained CNN model was achieved using 775 samples, and the confusion matrix obtained is shown in Figure 12. The y -axis of the figure shows the true labels, while the x -axis is the predicted labels for the samples drawn from the test set. The diagonal of our confusion matrix confirms the CNN architecture's impres-

sive performance, which is almost similar to what is obtained in a well-performing model where only the diagonals contain values. On row 1 of the matrix, we have 300, 0, 0, and 0, which means our CNN model successfully classified the 300 samples for the class represented by the row. The same follows for the second row, which has 0, 300, 0, and 0. The third row shows the values of 0, 0, 174, and 1, and the fourth output is 0, 0, 0, and 0 since no sample was passed for this class. The CNN prediction accuracy yielded a value of 0.998 as compared to the SVM, KNN, and complex decision tree yielding an accuracy of 0.900, 0.867, and 0.830, respectively, and demonstrating a very good performance of the CNN model in learning to classify home occupants from an intruder in the proposed MufHAS architecture. This performance also rippled into the values obtained for the precision, recall, F1-score, specificity, and Cohen's kappa, as listed in Table 2.

We compare our work with an existing deep learning-based smart mat monitoring system for identity information, stepping position, and activity status in a smart home or building environment [61]. The authors used a CNN model to predict valid users in a room, the authors tested their model using a dataset of 1000 samples, and the results of the CNN training model yielded an accuracy of 96%. MufHAS-CNN training yields an accuracy of 99.8%. However, it is noteworthy that the comparison with [60] was restricted to only one metric because the authors of the existing work focused their performance evaluation only on the prediction accuracy metric measure to evaluate their model. Also, the authors evaluated their implemented model based on triboelectric output signals and output voltages of the individual walking pattern, which they also applied in training their proposed CNN model, basically for recognition accuracy. The input values are different from the current implementation discussed in this study, hence the limitation in comparison. Moreover, the dataset used in [60] was not accessible to run or replicate any further analysis and comparison using other stated performance metrics.

TABLE 2: Results obtained from the prediction for the classification problem of house occupant and intruders using the proposed CNN architecture.

Metrics	Accuracy	Precision	Recall	F1-score	Specificity	Cohen's kappa
CNN	0.998	0.998	0.998	0.997	1.000	0.998
SVM	0.900	0.833	0.714	0.769	0.957	0.706
KNN	0.867	0.667	0.857	0.750	0.870	0.661
Complex decision tree	0.830	0.667	0.571	0.615	0.913	0.734

We also compared the proposed CNN model with the models presented in [24]. More so, similar to the above-highlighted limitation in metric comparisons with some related existing studies, the comparison with the model implementation in [25] is limited to accuracy because it is the only similar metric that was employed in the previous works. Table 3 compares our CNN model and previous works using the accuracy. This proves that the proposed CNN model can enhance home security in IoT smart home automation. With the proposed CNN models, smart home automation applications can be enhanced to detect intruders based on motion patterns. The users can also identify intruders based on the surveillance camera and models for detection, classification, and differentiation of motion patterns. The smart home applications' security notifications and alerts will also be based on detected motion patterns. Thus, home security is enhanced in the smart home environment.

4.5. Home Temperature and Humidity Evaluation. The temperature and humidity sensor was connected to the system and placed in a strategic location in the home to measure both the humidity and temperature values and display them on the mobile application. The mobile application was refreshed hourly, and the output displayed on the application indicates that the sensor receives a signal from the environment. Also, the values displayed correlated with the weather forecast for the periods. This shows that communication between the home environment and the mobile application is effective. The readings were taken hourly for seven days, and the graph was plotted using MATLAB. Figure 14 presents the graph for the temperature readings of MufHAS, while Figure 15 shows the MufHAS' graph of the humidity readings. From the graphs presented, $n = 1, 2 \dots, N$ is the n th hour collected, and the range for temperature readings is between 0°C and 30°C . The range of humidity is between 60 and 95%. Figures 16(a) and 16(b) show the output of readings from MufHAS plotted against online daily weather readings for temperature and humidity, respectively.

Based on the graphs in Figures 16(a) and 16(b), we calculated the mean percentage error to measure how close the designed MufHAS mobile application readings are to the daily weather forecast. The mean percentage error of the readings was calculated as follows:

$$\%Error = \frac{1}{N} \sum_{i=1}^N \left(\frac{|T_i^M - T_i^W|}{T_i^W} \right), \quad (10)$$

TABLE 3: Comparison with previous works.

Model	Accuracy (%)
CNN [61]	96.00
CON-RNN [24]	91.79
Small Conv-LSTM [24]	91.50
Large Conv-LSTM [24]	91.50
Conv-CWRNN [24]	92.38
Conv-DCWRNN [24]	93.02
Proposed CNN-MufHAS	99.80

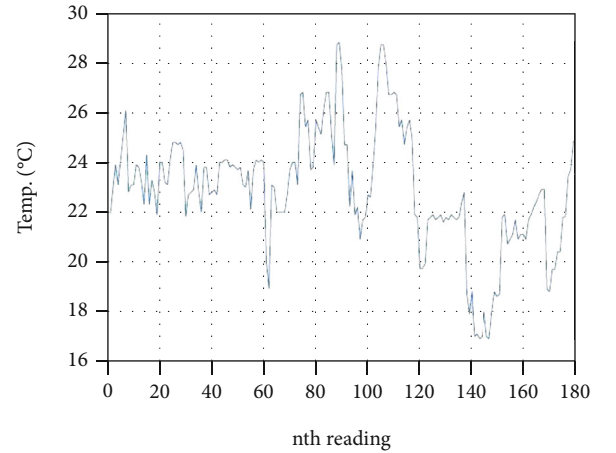


FIGURE 14: Graph of temperature readings.

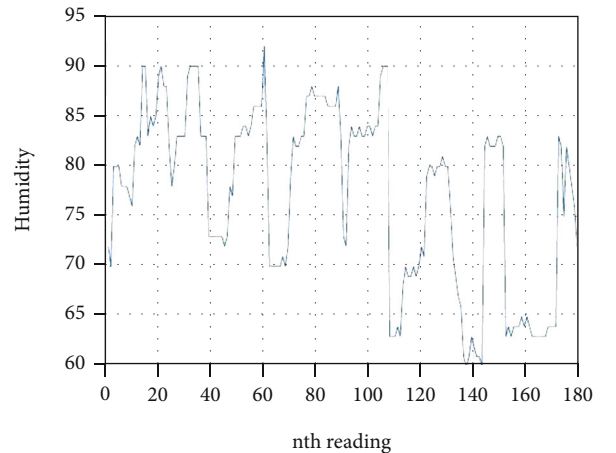


FIGURE 15: Graph of humidity readings.

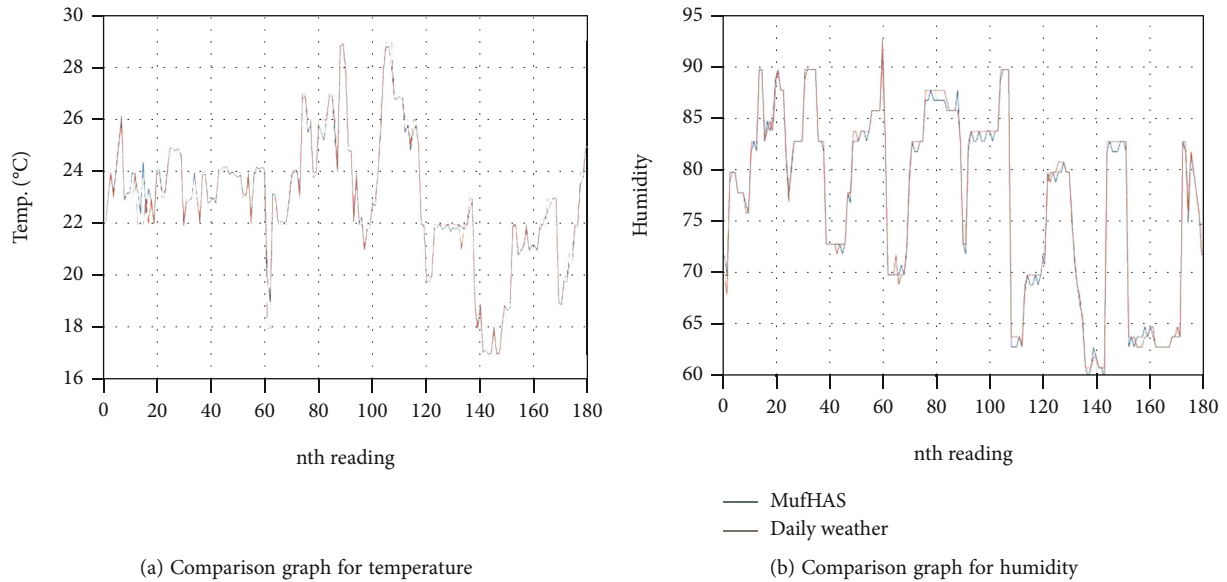


FIGURE 16

where T_i^M is the MuShAS temperature readings in $^{\circ}\text{C}$ for day i and T_i^W is the online delay weather temperature readings in $^{\circ}\text{C}$ for day i .

The temperature calculated percentage error for $N = 225$ measurements is 0.9057%, and the humidity's calculated percentage error for $N = 180$ measurements is 0.6256%. These show that the readings of the MufHAS mobile application are of good accuracy and be relied on to measure environmental conditions in the home with good results.

5. Conclusion and Future Recommendation

Improved convenient living, a healthy lifestyle, comfortability, and home security are areas of interest and development. The elderly, handicapped, and sick need to reduce daily activities that can stress them and negatively impact their health. To this end, a smart home automation system that can facilitate local and global monitoring, control, and safety of the home was developed. This work contributes to the existing research in home automation with the design and development of a multifunctional Android-based mobile application for the smart home automation domain. We have proposed an approach to enhance home security using the CNN deep learning model to classify and detect intruders in the home. The detection is based on the identification of motion in the home environment. Using this method shows that users will have enhanced security of their houses while having minimal disturbance from notifications.

The proposed method intends to eliminate frequent notifications and false notifications in a smart home automation system. The drawback of our proposed method is the detection of multiple movements at a time. The training and classification models were based on the movement of a person at a time. The feature extraction, classification, and accurate prediction of multiple movements either from animals or humans are not yet established. Therefore, the motion detection mod-

ule can be further developed for accurate classification and prediction of the exact motions detected when several movements are detected in the system. Another limitation of the system is not considering emotions as an influence on the movement of human beings. Emotions can make an individual's walking pattern differ from the regular movement, which may cause a false classification. Also, data generated from sensors can be used for analysis and prediction that can enhance the home automation system's functionality and improve users' lifestyles.

For the next phase of this research, we plan to implement more sensors in the system and test over a couple of weeks to generate data from the home environment to be used for analysis. A PIR motion sensor and ESP32-CAM attribute value to detected objects, and we intend to gather real-time motion images using our IoT devices via the cloud database to form a dataset for training and classification using the proposed support vector machine algorithm. A web interface is used for video streaming in our current work. As part of the future work, a mobile application module for video streaming will be incorporated, supporting real-time, cloud-based storage of media. Also, the deep learning model will be integrated into the mobile application for full functionality. As future work, a mechanism to classify friends and extended family members will also be considered for the system not to classify such a category of people as intruders. Electrical appliances are also considered to be fully implemented in the next phase. For future work, we intend to incorporate sensors in smart home healthcare into the system to assist in the remote monitoring of patients and use deep learning algorithms for our classification.

Abbreviations

Ha:	Home appliances
Nc:	Network connectivity
Hc:	Home control

Hs: Home sensor
 Ms: Motion sensor
 Ho: Home occupant (where the Ho is determined by the walking, running, limping, or jumping patterns of humans or pets in the home)
 Wp: Motion patterns (walking, jumping, limping, or running)
 HWp: Home occupant's walking pattern
 Hp: Home pet
 In: Intruder
 Cs: Cloud storage
 M: Motion
 D: Darkness
 LDR: Light-dependent resistor
 Cs: Cloud storage.

Data Availability

All data used in the study can be found in the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] B. Cele, *Quarter One Crime Statistics*, South African Government, Pretoria, 2021.
- [2] K. Ashton, "That 'internet of things' thing," *RFID Journal*, vol. 22, no. 7, pp. 97–114, 2009.
- [3] O. Taiwo, L. A. Gabralla, and A. E. Ezugwu, "Smart home automation: taxonomy, composition, challenges and future direction," in *International Conference on Computational Science and Its Applications*, Cagliari, 2020.
- [4] N. King, "Smart home-a definition," 2003, http://www.housinglin.org.uk/_assets/Resources/Housing/Housing_advice/Smart_Home_-_A_definition_September_2003.pdf.
- [5] F. Mehmood, I. Ullah, S. Ahmad, and D. Kim, "Object detection mechanism based on deep learning algorithm using embedded IoT devices for smart home appliances control in CoT," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2019.
- [6] D. Liciotti, M. Bernardini, L. Romeo, and E. Frontoni, "A sequential deep learning application for recognising human activities in smart homes," *Neurocomputing*, vol. 396, pp. 501–513, 2020.
- [7] J. Jaihar, N. Lingayat, P. S. Vijaybhai, G. Ventakesh, and K. P. Upla, "Smart home automation using machine learning algorithms," in *International Conference for Emerging Technology (INCET)*, Belgaum, India, 2020.
- [8] D. Popa, F. Pop, C. Serbanescu, and A. Castiglione, "Deep learning model for home automation and energy reduction in a smart home environment platform," *Neural Computing and Applications*, vol. 31, no. 5, pp. 1317–1337, 2019.
- [9] Z. Peng, X. Li, and F. Yan, "An adaptive deep learning model for smart home autonomous system," in *International Conference on Intelligent Transportation, Big Data and Smart City (ICITBS)*, Vientiane, Laos, 2020.
- [10] S. K. Shah, Z. Tariq, and Y. Lee, "Audion IoT analytics for home automation safety," in *IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, 2018.
- [11] A. Oppermann, "Towards data science - what is deep learning and how does it work?," 2019, September 2021, <https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb69ac>.
- [12] M. Khan, J. Seo, and D. Kim, "Towards energy efficient home automation: a deep learning approach," *Sensors*, vol. 20, no. 24, p. 7187, 2020.
- [13] M. Sahu and R. Dash, "A survey on deep learning: convolution neural network (CNN)," in *Intelligent and Cloud Computing, Smart Innovation, Systems and Technologies*, 2021.
- [14] G. Lobaccaro, S. Carlucci, and E. Löfström, "A review of systems and technologies for smart homes and smart grids," *MPDI Energies*, vol. 9, no. 5, p. 348, 2016.
- [15] Z. Xiaodong and Z. Jie, "Design and implementation of smart home control system based on STM32," in *30th Chinese Control and Decision Conference (2018 CCDC)*, Shenyang, 2018.
- [16] S. Bimenyimana, A. Ishimwe, G. N. O. Asemota, C. M. Kemunto, and L. Li, "Web-based design and implementation of smart home appliances control system," *IOP Conference Series: Earth and Environmental Science*, vol. 168, article 012017, 2018.
- [17] T. S. Gunawan, R. H. I. Yaldi, M. Kartiwi et al., "Prototype design of smart home system using internet of things," *Science*, vol. 7, no. 1, pp. 107–115, 2017.
- [18] U. Singh and M. A. Ansari, "Smart home automation system using internet of things," in *2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)*, Greater Noida, 2019.
- [19] H. Aadel and J. Ali, "Design and implementation prototype of a smart house system at low cost and multi-functional," in *Future Technologies Conference (FTC)*, San Francisco, 2016.
- [20] A. Parsa, T. A. Najafabadi, and F. R. Salmasi, "Implementation of smart optimal and automatic control of electrical home appliances (IOT)," in *IEEE 2017 Smart Grid Conference (SGC)*, Tehran, Iran, 2017.
- [21] R. D. Manu, S. Kumar, S. Snehashish, and K. S. Rekha, "Smart home automation using IoT and deep learning," *International Research Journal of Engineering and Technology*, vol. 6, no. 4, pp. 1–4, 2019.
- [22] R. Majeed, N. A. Abdullah, I. Ashraf, and Y. B. Zikria, "An intelligent, secure, and smart home automation system," *Scientific Programming*, vol. 2020, 14 pages, 2020.
- [23] A. Brenon, F. Portet, and M. Vacher, "Arcades: a deep model for adaptive decision making in voice controlled smart-home," *Pervasive and Mobile Computing*, vol. 49, pp. 92–110, 2018.
- [24] N. Neverova, C. Wolf, G. Lacey et al., "Learning human identity from motion patterns," *IEEE Access*, vol. 4, pp. 1810–1820, 2016.
- [25] S. K. Saravanan, A. M. Nainar, and S. V. Marichamy, "Android based smart automation system using multiple authentications," *IRE Journal*, vol. 3, no. 6, pp. 60–65, 2019.
- [26] W. A. Jabbar, M. H. Alsibai, N. S. Amran, and S. K. Mahayadin, "Design and implementation of IoT - based automation system for smart home," in *International Symposium on Networks, Computers and Communications*, Rome, Italy, 2018.
- [27] L. A. Ajao, J. G. Kolo, E. A. Adedokun, O. M. Olaniyi, O. C. Inalegwu, and S. K. Abolade, "A smart door security-based home automation system: an internet of things," *SciFed Journal of Telecommunication*, vol. 2, no. 2, 2018.

- [28] H. Singh, V. Pallagani, V. Khandelwal, and U. Venkanna, "IoT based smart home automation system using sensor node," in *Fourth International Conference on Recent Advances in Information Technology (RAIT-2018)*, Dhanbad, India, 2018.
- [29] O. Taiwo, A. E. Ezugwu, N. Rana, and S. M. Abdulhamid, "Smart home automation system using Zigbee, Bluetooth, and Arduino technologies," in *Computational Science and Its Applications-ICCSA 2020*, pp. 587–597, Springer, Cham, 2020.
- [30] O. Taiwo and A. E. Ezugwu, "Smart healthcare support for remote patient monitoring during Covid-19 quarantine," *Informatics in Medicine Unlocked*, vol. 20, no. 100428, pp. 1–12, 2020.
- [31] M. U. Alam, R. Tiwari, and J. Sharma, "Design and implementation of smart home automation system using LabVIEW," *International Journal for Scientific Research and Development*, vol. 8, no. 5, pp. 515–520, 2020.
- [32] M. Naing and N. S. Hlaing, "Arduino based smart home automation system," *International Journal of Trend in Scientific Research and Development*, vol. Volume-3, no. Issue-4, pp. 276–280, 2019.
- [33] S. R. Jena, R. Shanmugam, A. Tyagi, and R. K. Patjoshi, "Implementation of voice recognized LED using Arduino Uno Micro," *International Journal of Future Generation Communication and Networking*, vol. 13, no. 3, pp. 2233–2241, 2020.
- [34] O. Taiwo and A. E. Ezugwu, "Internet of things-based intelligent smart home control system," *Security and Communication Networks*, vol. 2021, 17 pages, 2021.
- [35] S. Mahmud, S. Ahmed, and K. Shikder, "A smart home automation and metering system using internet of things (IoT)," in *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, Dhaka, Bangladesh, 2019.
- [36] L.-D. Liao, Y. Wang, Y.-C. Tsao et al., "Design and validation of a multifunctional Android-based smart home control and monitoring system," *IEEE Access*, vol. 7, pp. 163313–163322, 2019.
- [37] L. Smirek, G. Zimmermann, and M. Beigl, "Just a smart home or your smart home – a framework for personalized user interfaces based on eclipse smart home and universal remote console," *Procedia Computer Science*, vol. 98, pp. 107–116, 2016.
- [38] M. Wang, G. Zhang, C. Zhang, J. Zhang, and C. Li, "An IoT-based appliance control system for smart homes," in *Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, Beijing, China, 2013.
- [39] J. Chhabra and P. Gupta, "IoT based smart home design using power and security management," in *International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, Noida, 2016.
- [40] Y. Li, "Design of a key establishment protocol for smart home energy management system," in *Fifth International Conference on Computational Intelligence, Communication Systems and Networks*, Madrid, 2013.
- [41] H. Ghayvat, S. Mukhopadhyay, X. Gui, and N. Suryadevara, "WSN- and IOT-based smart homes and their extension to smart buildings," *Sensors*, vol. 15, no. 5, pp. 10350–10379, 2015.
- [42] R. Piyare, "Internet of things: ubiquitous home control and monitoring system using Android based smart phone," *International Journal of Internet of Things*, vol. 2, no. 1, pp. 5–11, 2013.
- [43] R. K. Kodali, V. Jain, S. Bose, and L. Boppana, "IoT based smart security and home automation system," in *International Conference on Computing, Communication and Automation (ICCCA)*, Noida, 2016.
- [44] A. Anitha, "Home security system using internet of things," *IOP Conference Series: Materials Science and Engineering*, vol. 263, article 042026, 2017.
- [45] M. Barnachon, S. Bouakaz, B. Boufama, and E. Guillou, "Ongoing human action recognition with motion capture," *Pattern Recognition*, vol. 47, no. 1, pp. 238–247, 2014.
- [46] R. El-Azab, "Smart homes: potentials and challenges," *Clean Energy*, vol. 5, no. 2, pp. 302–315, 2021.
- [47] R. J. Orr and G. D. Abowd, "The smart floor: a mechanism for natural user identification and tracking," in *CHI 00: extended abstracts on human factors in Computing Systems*, The Hague The Netherlands, 2000.
- [48] J. Suutala and J. Roning, *Towards the Adaptive Identification of Walkers: Automated Feature Selection of Footsteps using Distinction Sensitive LVQ*, International Workshop on Processing Sensory Information for Proactive Systems (PSIPS), 2004.
- [49] N. Rahman, K. Kalpoma, and T. Hasan, "Automated person identification system using walking pattern biometrics," *International Journal of Scientific and Engineering Research*, vol. 6, pp. 98–102, 2015.
- [50] H. Sidenbladh, "Detecting human motion with support vector machines," in *17th IAPR International Conference on Pattern Recognition*, Cambridge, 2004.
- [51] S. Khan, M. H. Javed, E. Ahmed, S. A. A. Shah, and S. U. Ali, "Facial recognition using convolutional neural networks and implementation on smart glasses," in *2019 International Conference on Information Science and Communication Technology (ICISCT)*, Karachi, Pakistan, 2019.
- [52] H.-R. C. Cheng, G.-Z. Cao, C.-H. Li, A. Zhu, and X. Zhang, "A CNN-LSTM hybrid model for ankle joint motion recognition method based on sEMG," in *International Conference on Ubiquitous Robots (UR)*, Kyoto, Japan, 2020.
- [53] O. N. Oyelade and A. E. Ezugwu, "Deep learning model for improving the characterization of coronavirus on chest X-ray images," medRxiv, 2020.
- [54] M. Ismail, "Towards data science - convolutional neural networks' mathematics," 2020, September 2021, <https://towardsdatascience.com/convolutional-neural-networks-mathematics-1beb3e6447c0>.
- [55] X. Qi and J. Liu, "Comparison of support vector machine and softmax classifiers in computer vision," in *Second International Conference on Mechanical Control and Computer Engineering*, Harbin, China, 2017.
- [56] Y. Li, T. Wang, B. Kang et al., "Overcoming classifier imbalance for long-tail object detection with balanced group softmax," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020.
- [57] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 39, pp. 1137–1149, 2016.
- [58] S. Ahlawat and A. Choudhary, "Hybrid CNN-SVM classifier for handwritten digit recognition," *International Conference on Computational Intelligence and Data Science*, vol. 167, pp. 2554–2560, 2020.

- [59] Chewett, “The Chewett blog, electronics,code and contraptions,” 2017, November 2020, <https://chewett.co.uk/blog/309/dht11-sensor-review-code-examples/>.
- [60] “CVonline: image databases,” 2019, August 2021, <https://homepages.inf.ed.ac.uk/rbf/CVonline/Imageabase.htm#people>.
- [61] Q. Shi, Z. Zhang, T. He et al., “Deep learning enabled smart mats as a scalable floor monitoring system,” *Nature Communications*, vol. 11, no. 4609, pp. 1–11, 2020.

Research Article

Analysis and Research on Digital Reading Platform of Multimedia Library by Big Data Computing in Internet Era

Wanxia Zhang ¹, Bo Liu,¹ and Sang-Bing Tsai ²

¹Jiangsu Urban and Rural Construction College, Jiangsu 213000, China

²Regional Green Economy Development Research Center, School of Business, WUYI University, China

Correspondence should be addressed to Wanxia Zhang; zhangwx0628@126.com

Received 13 November 2021; Revised 10 December 2021; Accepted 20 December 2021; Published 10 January 2022

Academic Editor: Chao-Yang Lee

Copyright © 2022 Wanxia 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.

Digital reading promotion service is a service way for libraries to provide readers with a series of digital resources, enjoy the service functions, and share the experience of using them in various digital reading platforms, which is to meet the reading interests and reading needs of more readers, and is also the focus of the current library work. In the era of new media, the characteristics of digital reading are subtly changing the readers' needs for reading environment, reading content, and reading style. Libraries should keep pace with the development of the times and provide readers with diversified, intelligent, and targeted digital reading platforms. The digital reading platform should continuously improve the digital reading service functions, broaden the service scope and dissemination channels, and finally realize the diversification, interest, and intelligence of digital reading service. This paper takes the digital reading platform of libraries in the region as the research theme and carries out research work on libraries. The province is divided into three regions according to the geographical map: southeastern region, central region, and northwestern region. The digital reading platforms of 14 prefecture-level public libraries and 58 libraries of higher education institutions in each region were accessed. Firstly, we check the construction of digital resources within the library websites, secondly, we count the opening of digital reading platform functions, and finally, we check the opening of digital reading platforms. Through the research, it is found that there are problems of unbalanced distribution of digital reading resources in regional libraries; unattractive design of readers' interface and inadequate reading service functions; lack of continuous training of readers' guidance; insufficient publicity and promotion; low efficiency of staff in responding to consultation; and low degree of platform openness and weak awareness of sharing. Finally, the problems found in the research are summarized, and the solution measures for the regional digital reading platform are proposed. Libraries in the digital era should give priority to systems that can manage all library resources comprehensively and effectively, adapt to more flexible library workflows, and enable libraries to provide better services to users.

1. Introduction

Along with the rapid development of digital technology and network technology, more and more digital reading information has become a resource that can be described and accessed on the network [1]. Reading resources have evolved from the initial single paper text content to today's situation where audio, images, video, and hyperlinks coexist. Diversified, personalized, and intelligent network technologies advance the development and innovation of new digital reading services [2]. Profit-oriented digital commercial reading platforms are being constantly improved and updated,

while libraries are still lacking in digital reading promotion activities and open sharing of resources. The traditional stacked system faces problems such as inconvenient data interaction and poor scalability in the way of providing reading services, which restrict the development of digital reading promotion work [3]. At present, libraries are in the stage of transformation to digital services and need to make a change to realize the popularization and promotion of digital resources and services, so it is timely and extremely necessary to optimize digital reading promotion services. With the increasing proportion of digital resources in university library collections, the services provided by university

libraries for users are becoming more and more diversified, and library management and services are being expanded and extended; the traditional business workflow and service methods of university libraries are increasingly unable to meet the management and service requirements under the current digital conditions, so university libraries are facing transformation and change [4]. However, the transformation and change of libraries cannot be achieved without the support of library management system.

Due to the changes in the nature of collection resources and service contents and methods, the traditional integrated library management system has become increasingly difficult to perform the management services of modern libraries [5]. In other words, many university libraries are still using products developed more than ten years ago, and although they have been upgrading their systems, they are only upgrading traditional functions or integrating new system modules, such as Huifen and ILAS, which can solve part of the business and service problems by introducing auxiliary products to integrate with their management systems, but they cannot change the underlying architecture of the system and cannot solve the problem at root [6]. In recent years, the rapid development of network technology and the exponential growth of electronic resources and digital content have increasingly exposed the weaknesses of traditional integrated systems. In order to deal with the increasing electronic and digital resources, many professional management systems have come out one after another, such as electronic resource knowledge base, electronic resource management system, digital resource management system, and resource discovery system [7]. However, due to the lack of flexibility, interoperability, and efficiency, these systems can hardly meet the needs of both internal and external users of libraries, and university libraries urgently need new systems to replace the original system. In the digital era, users' information needs have changed significantly, and university libraries are undergoing a comprehensive transformation and development from resource construction to service mode [8]. This paper analyzes the main library integration systems at home and abroad by sorting out the development of library integration systems, comparing the similarities and differences in technology and functions of each system at home and abroad and comprehensively evaluating each type of system in different periods. On the basis of understanding and mastering the functional characteristics of the systems, the corresponding selection principles and selection strategies are proposed, so as to guide the university libraries in China to carry out the selection work better [9]. By systematically sorting out the development history of integrated library systems, this paper summarizes the commonalities that exist between libraries and library systems, and by studying the new concepts and technologies of foreign products, it helps domestic absorption and reference of foreign advanced development concepts and can further enrich the theoretical system of library automation systems [10].

By analyzing the system architecture and functional characteristics of mainstream products in the current automation market, we aim to understand the differences and

connections between different systems and provide feasible suggestions for the introduction of suitable systems in university libraries. With the development of information technology and the increasing digitalization of libraries, the industry environment in which college libraries are located has changed dramatically, yet many college library systems are still using products developed in the previous generation, and the outdated business processing methods and management concepts have hindered the development and changes of college libraries. Therefore, the renewal of university library management systems becomes an urgent problem. In this paper, by comparing different products developed by different automation system vendors, we find out the similarities and differences in the functions of each system, which is a reference for domestic university libraries to select a new system that suits their development concept and service needs in a targeted way. The main research of this paper is to investigate the design concepts, functions, and technical characteristics of traditional library integration system, new generation library integration system, and next-generation library service platform as well as the current situation of domestic university libraries' applications and to compare and analyze them, compare the advantages and disadvantages of different products, and accordingly propose the selection principles and selection strategies applicable to different types of university libraries in China. At the same time, libraries increase their investment in digital reading resources to meet the reading needs of different types of readers. Libraries need to continuously improve digital technology, speed up the construction of digital reading platform, increase the number of digital reading services, and improve the service experience of readers.

2. Related Work

The Digital Age, also known as the Post-Information Age and the Bit Age, was first proposed in the book *Digital Survival* made by MIT professor Negroponte. Subsequently, Wikipedia defined it as the computer age or the information age. In the digital age, people will have more choices, more diverse lifestyles, more widespread cooperation, and sharing, and everyone will have the freedom to transmit information and the ability to access it at the right time, which was difficult or impossible to do in the past. In such an era, people are more inclined to solve problems through the Internet [11]. At the same time, the iteration of technology has promoted the deeper development of user needs. Among them, user needs in the digital era show four main characteristics: diversification, integration, personalization, and interactivity. Based on the above concepts, the author believes that the digital era can be simply understood as the era of meeting people's information needs with the help of digital technology [12].

The essential feature of the digital era library is to use modern digital information technology and network communication technology to process and transform all kinds of traditional media documents into digital information. In addition, libraries in the digital era should have the equipment and ability to store and protect large amounts of information [13]. With the advent of digital era, the collection of

university libraries gradually changes from physical resources to digital resources. In order to meet the teaching and research needs of university teachers and students, university libraries expand their collections by purchasing a large number of electronic and digital resources on the one hand and actively participate in self-built and co-built resources on the other hand, which realize the digitization and sharing of local resources [14]. In the future university libraries, digital resources and traditional print resources will jointly constitute the university library collection resource system and will occupy a dominant position. The current library services are changing from the traditional way to the networked and digital way. Traditionally, university libraries mainly focus on circulation and reading of paper documents, and their service mode is also focused on providing paper resources collection service. This passive and mechanical way determines that users need to go to the library to enjoy services such as lending and returning services, searching services, and consulting services [15]. It also determines that users can only use the library's collection and services, and it is difficult to share resources with other universities or institutions. However, the emergence of mobile Internet has broken this limitation. When users want to receive library services, they only need a cell phone to access the library system so that they can enjoy the services brought by the library, such as collection inquiry, book renewal, online consultation, network information retrieval, and other services, and at the same time, they can easily realize resource sharing with other university libraries or institutions through the Internet platform. With the advent of the digital era, the proliferation and high-speed utilization of information no longer restrict users to the collection resources, which enriches access but also makes it difficult for users to accurately extract useful information from the vast amount of resources [16].

This has eventually evolved into a situation where the accuracy of people's access to the information they need is decreasing with the growth of information resources. As a result, users expect libraries to provide more personalized services to improve the utilization of information resources [17]. First, users want to have mobile access to resources. Users have a strong demand for off-campus (off-library) access to library resources, and they want to be able to access all kinds of resources without going to the library [18]. The popularity of mobile communication technology and wireless network technology enables users to access library resources anytime and anywhere, which greatly improves the efficiency of accessing resources. Secondly, users want to customize their services and organize content according to their needs and preferences, so that libraries can provide targeted services. Finally, users have a high level of demand for personalized community communication platform and want to have a platform for community communication. As an important information resource center of the university, university libraries should not only have high-quality information resources but also have easy access to external information resources, so as to better provide high-quality literature and information services for teachers, students, and scholars [19], although the proportion of digital

resources in university collections is getting higher and higher and can meet the needs of the vast majority of users. However, traditional printed documents still occupy a large proportion in university libraries and have many advantages that cannot be replaced by digital resources, so for a long time, digital resources and paper resources will coexist and develop in the collection resources through the way of complementary advantages [20]. On the other hand, although traditional libraries also advocate the sharing of resources, the degree of sharing is greatly restricted by the limitation of the technical power of traditional libraries [21]. In the digital era, libraries can cooperate with library consortia and CALIS to use interlibrary loan and document delivery, so that users can download and deliver documents through a library webpage, thus reducing the duplicate purchase of resources, realizing the effective use of documents, and solving the contradiction between the expensive cost of using documents and the limited acquisition funds of libraries.

3. Integrated Library Digital Reading System Platform

3.1. Library Integration System. First of all, the cooperative alliance between libraries needs to develop policies on digital reading services and be sure to implement the content of the policy requirements to achieve win-win cooperation in the alliance. The integrated library system is an integrated system that makes the main functions of the library (including interviewing, cataloging, circulation, continuous publication management, and public search) all realized on the basis of one bibliographic database. With the use of the integrated system, library staff can always know which status the library collection is in, which facilitates the daily management of librarians and improves the efficiency of library operations. The library connects users to various shared resources through the management system, providing fast and powerful search options, and users can retrieve the collection resources at any time, enhancing the user experience. With the development of the times and the changing needs of users, integrated library systems are always in a dynamic development. "Library Automation System," "Library Integrated System," "Next Generation Library Integrated System," "Next Generation Library Service Platform," and other names are all more relevant kind of names or terms in a certain period of time. In short, no matter how the names evolve, the author believes that they are essentially systems that use advanced technologies to provide better management and services for libraries and users. Cloud computing refers to the delivery of computing and storage capacity as a service to end users, where end users access cloud-based applications through a web browser or mobile application, while business software and data are stored on remote servers.

The next-generation service platform in the cloud computing environment uses a microservice architecture that allows various library resources, such as electronic and digital resources, and network resources, to be stored in the cloud in an open manner. In addition, it allows users to access and utilize resources anytime and anywhere with

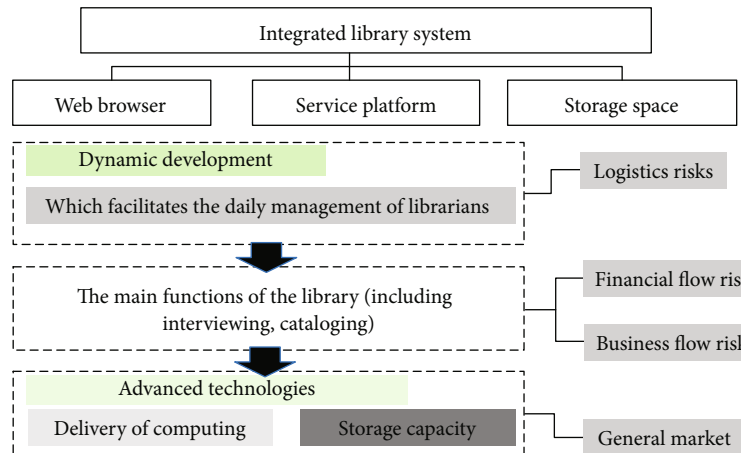


FIGURE 1: Integrated library system.

lower cost and storage space. Also, interoperability is enhanced, and libraries can read the featured data on the Internet directly through API interface, as shown in Figure 1. The maintenance of the system after cloud-based will not have to be done through the local client but can be updated and maintained on the cloud platform, and the future data migration and system replacement can be done through the cloud, which not only reduces the cost of system migration but also improves the overall efficiency of the library. See Figure 1.

Open source systems developed based on the open source concept represent a spirit of openness. From the perspective of Lib2.0, open source software is not only the openness of source code and content but also the openness of application programming interface (Open API) and Open Link standard specifications such as OpenURL. Koha open source system in foreign libraries proves that open source system is a viable option to break the monopoly of commercial systems. Before the emergence of open source systems, the monopoly of automation system companies and data developers on library integration systems and databases made libraries lack voice and bargaining power over the construction of their management systems, which greatly limited the development of libraries. Therefore, the next-generation LSP must be an open system. Although the next-generation systems that have emerged in recent years also emphasize openness, they only open part of their interfaces, which is still too closed compared with open source systems that advocate open source. The next-generation LSP is completely open source, which breaks through the traditional application interface model and opens up the library resources and services to the outside world in a more effective way, and also introduces external open data and services into the library system, which can not only improve the openness of library resources but also make up for the lack of resources and services in the library. In addition to the openness and access to content, the next-generation service platform is also committed to putting all open source content including source code on the community platform, driven by the community and open to the community, so that library staff, third-party developers, and users can par-

ticipate in the community, build the library service platform together, and provide technical support and services for the construction of the next-generation LSP.

The next-generation LSP takes a more modular approach, packaging the system into individual ILS modules and ERMS modules. These modules will be published as applications in the app store, and libraries or users can pick and choose functional modules according to their needs, just like a smartphone downloading an app from the app store, and then combine them into an optimized new system. The modular approach makes interaction between applications both easy and fast, meeting the need for libraries to utilize application components and services more efficiently. Under the influence of data globalization, data accessibility and system scalability are important features combined in the next-generation LSP. The next-generation service platform should have good scalability to expand various application systems as needed, such as campus network management system, institutional knowledge base, and resource discovery system.

3.2. Library Digital Reading Construction. Alma is a next-generation integrated library system developed by Ex Libris, a provider of library automation systems, and according to the product roadmap provided by Ex Libris on its website, Alma embodies Ex Libris' "Unified Resource Management" concept. Alma is designed from the ground up to be deployed on a multitenant platform with a browser-based interface that supports the complete library operations of resource selection, access, metadata cataloging and management, digital resource management, and user services, regardless of collection size and library size, as shown in Figure 2; processes were personalized. To support Alma's globally dispersed customer base, Alma provides support services through data centers in multiple countries on multiple continents, including the U.S., Singapore, and the Netherlands. The next-generation LSP is completely open source, which breaks through the traditional application interface model and opens up the library resources and services to the outside world in a more effective way, and also introduces external open data and services into the library

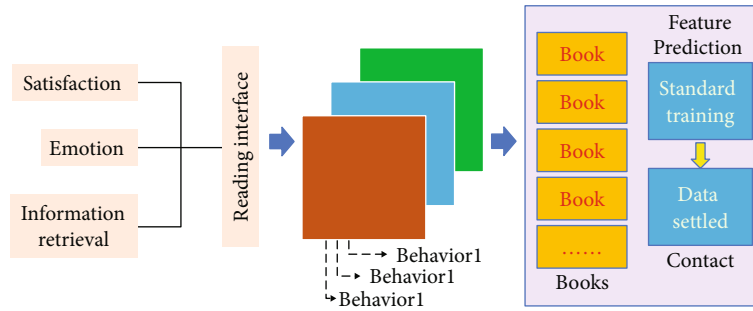


FIGURE 2: Library digital reading interface settings.

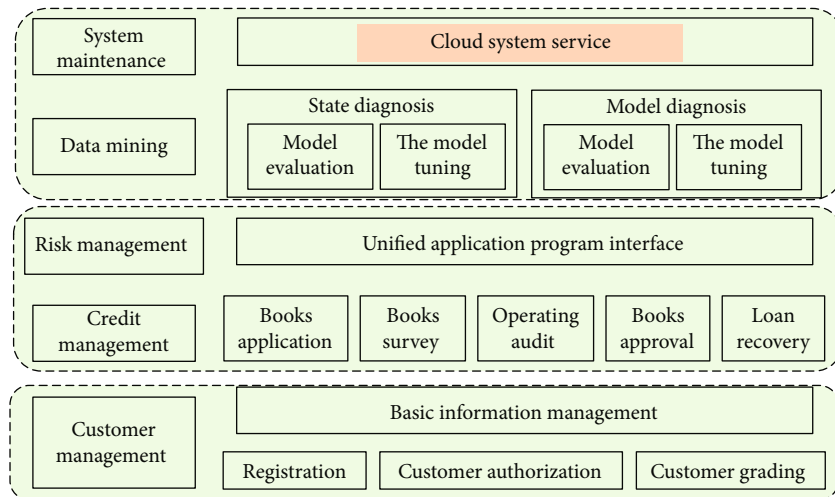


FIGURE 3: Cloud system service experience.

system, which can not only improve the openness of library resources but also make up for the lack of resources and services in the library. In addition to the openness and access to content, the next-generation service platform is also committed to putting all open source content including source code on the community platform, driven by the community and open to the community and library staff. Alma has extended its integration interfaces and application programming interfaces (APIs) to allow access to the data and workflows stored in the platform. Since its introduction, the Alma system has attracted significant interest from university libraries, with 1,808 libraries worldwide having installed Alma to date, according to the Library Technology Guide website, including 1,501 university libraries. According to the Library Technology Guide website, 1,808 libraries worldwide have installed Alma, and 1,501 of them have chosen Alma as their system platform, accounting for 83%. See Figure 2.

The new generation of integrated systems offers Software as a Service (SaaS) in a multitenant model, where SaaS can host services through a cloud platform without the need for local client installation. On the other hand, such a setup can also increase the library’s dependence on the Internet. When the Internet or local network is disconnected, the library will be disconnected from its system. Both Alma and WMS support cloud platform deployment; through cloud computing technology, the system can be deployed

more rapidly, with version upgrades, hardware updates, patching vulnerabilities, and backing up system data; online cloud upgrades can be handled directly without the need for interruption of service, as shown in Figure 3 (high litigation). However, cloud services also bring certain risk, such as data being held by the service provider; there is a risk of leakage. Open interface is one of the features that distinguish the new generation of integrated systems from traditional integrated systems. Open and unified application program interface (API) can more easily access each other with other systems, realize the sharing of resources, and better meet the needs of users for open access to resources in the digital era. Alma, WMS, and Sierra all provide open API interfaces to access third-party applications and integrate with the system. Alma is a commercial system, but it has open APIs for almost all functional areas, allowing libraries to access the system’s data and functionality through external programs and applications. For example, these APIs make it possible to access Alma through other discovery layers outside of Primo. The library will also be able to use the APIs to interconnect Alma with campus infrastructure such as learning management systems, financial systems, and research management systems, thus embedding library services into the university’s teaching and research. The library will be able to access and process the data in Sierra through SQL. See Figure 3.

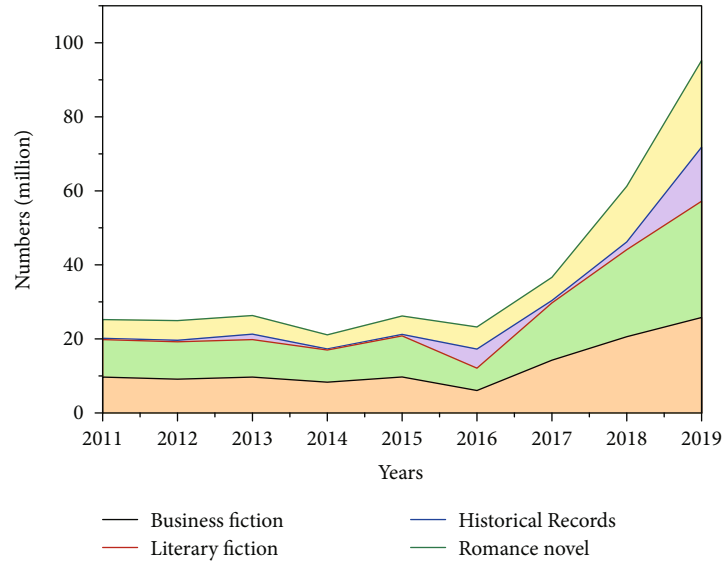


FIGURE 4: The relationship between reading ability and fragmented reading.

3.3. Interactive Experience of Digital Reading. The ability to read is an ability that people must acquire in the process of socialization in the digital age, and it is also an ability that is necessary to achieve digital survival. Children's reading ability is an important means for children to perceive the world, build their emotional structure, enhance their sense of identity, and thus build their self-worth system. Paper books are a traditional linear text medium that exhibits a linear temporal character from the past to the future, while digital media are characterized by virtual temporality and spatial mobility. This difference leads to the transformation of deep reading in the paper era into fragmented reading in the digital era. Digital media often make children more easily immersed in fragmented reading because of the wonderful sensory enjoyment they bring, and although this approach can somewhat promote the development of children's sensory organs and symbolic thinking, deep reading is a multidimensional perception and exploration of reading content, and the absence of deep reading will restrict the balanced development of deep reading, and fragmented reading can fundamentally improve children's reading ability, as shown in Figure 4. And children can avoid getting into trouble and develop good digital reading habits by receiving proper guidance from parents in parent-child reading for digital reading. However, the author found from interviews with parents of three groups of parent-child families that parents often failed to play the role of a guide in parent-child digital reading together, which transformed parent-child reading into general digital reading for children. See Figure 4. The benefits of clientless and cloud-based systems are obvious; i.e., system maintenance and upgrades can be achieved without downtime, and library staff can freely access the system and perform work from any network location.

In the interviews, we asked about the status of digital reading with parents and children in the library and found that parents choose digital reading together mostly because

of its rich and interesting resources, but in the actual reading together process, we found that children only like to watch video content and lack communication with parents, which leads them to feel that the meaning of reading together is lost, and parents hope that the digital reading all-in-one machine can help them arrange a reasonable setup. Parents hope that the digital reading machine can help them arrange a reasonable parent-child reading mode, so that they are more willing to use the digital reading machine. Parent-child digital reading and children's general digital reading are different from the participation of parents, so the product should meet the basic digital reading needs, but we also should think about how to help parents can better assist children to read. In terms of the content of parent-child reading, parents need to have a certain understanding of the reading material, find the reading value of it, make them interested in the reading content through wonderful explanation and interpretation, and also take the initiative to think and understand the content, and when children have doubts and confusion in the reading process, parents can use the correct expression to explain to children, making the whole reading process more in-depth and meaningful, and also show respect. It is also a sign of respect for children. However, most parents, due to their work, cannot guarantee that they will be able to preview the reading content when they read with their children, driving them to unconsciously read in pieces with their children. Therefore, when the library digital reading terminal can let parents understand the general content and connotation of the required reading materials in a short time, it can make up for the parents' lack of time and avoid impatience, so that they can guide children more patiently in parent-child reading together.

4. Analysis of Digital Reading Experience

In order to meet the requirements of university library services in the digital age and be able to meet the needs of

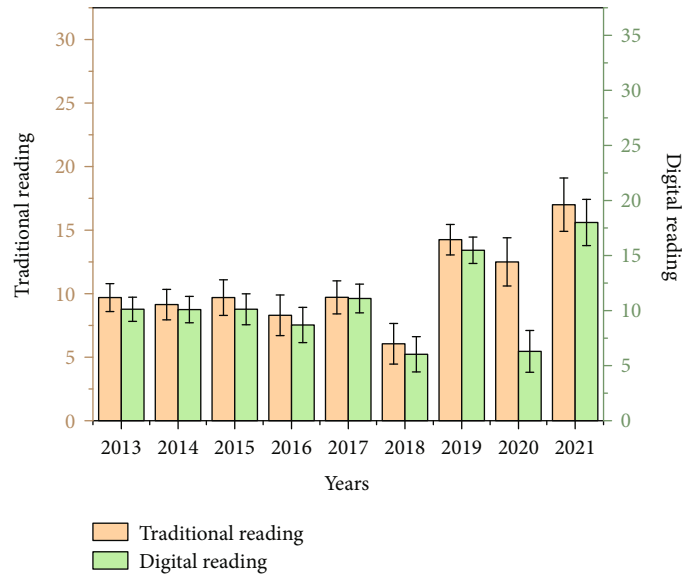


FIGURE 5: Change in reading volume for multi-channel reading system.

future development of university libraries, the advancement of technology should be ensured. Library integration systems generally adopt Client/Server (client/server), Browser/Server (browser/server), or a mixture of both architectures, and although these two architectures can meet the needs of traditional libraries, they cannot keep up with the development of the times. Therefore, the new generation system becomes a key concern. The new generation system realizes a service-oriented architecture, supports multitenant cloud deployment, and supports multiple communication standards and resource description standards, which can well meet the management and service needs of university libraries, as shown in Figure 5. From the long-term development of libraries, it is undoubtedly more advantageous to choose a technologically advanced management system. The integrated library system itself is to meet the daily work needs of the library, so basing on the actual needs of the library is the most fundamental selection principle. First of all, university libraries should evaluate the development concept and management requirements of the library and examine whether the functions of the system meet the resource management and service requirements of the library: for example, whether the resource type of the library is mainly traditional paper or digital literature and whether the professional ability and technical strength of the librarians can provide deeper knowledge services under the new technology conditions. Secondly, we should also consider whether the system provider can give good after-sales service and technical support, including system update and maintenance. Therefore, university libraries should choose a suitable integrated system according to their own needs. See Figure 5. There is a shift in the reading medium used for parent-child reading, which in turn has a different impact on children's reading skills.

Economics is also an important factor to consider when choosing a library integration system for a university library. As a nonprofit organization, university libraries are mainly

funded by their parent organizations. Different types of university libraries have different funding, and universities with sufficient funding usually consider the system with excellent performance first and less consider the cost performance of the system, while universities with insufficient funding often consider the affordability of the library. When selecting a system, university libraries must consider the initial purchase cost, annual subscription cost, additional costs for system updates, and training costs, based on the overall system functionality and the manufacturer's technical support and after-sales service. It is advisable for the library community to plan to join the construction of FOLIO-type next-generation service platform projects in the form of regional or industry consortia, so that libraries can transform from system users to platform participants and plan the layout for the eventual construction of an open-source, scalable, and community-integrated service platform. On the other hand, even if the plan of building the next-generation service platform has not yet been implemented, the libraries can take the lead in changing the concept by participating in the FOLIO project, providing useful references for the design of the next-generation service platform system and at the same time cultivating technical talents for the next-generation service platform, so as to lay a solid foundation for the implementation of the next-generation service platform. Secondly, multiple university libraries form an alliance and choose the next-generation integrated system as a regional common platform. The alliance members work on the same platform and unify the workflow to achieve real resource and service sharing in the region and industry. At present, the Alma system has been successfully practiced to realize the sharing of front and back office systems among alliance members, which facilitates the connection of systems and makes it possible for members to carry out regional joint work and services more smoothly and better realize the sharing of literature and information resources among universities, as shown in Figure 6. Finally, if the

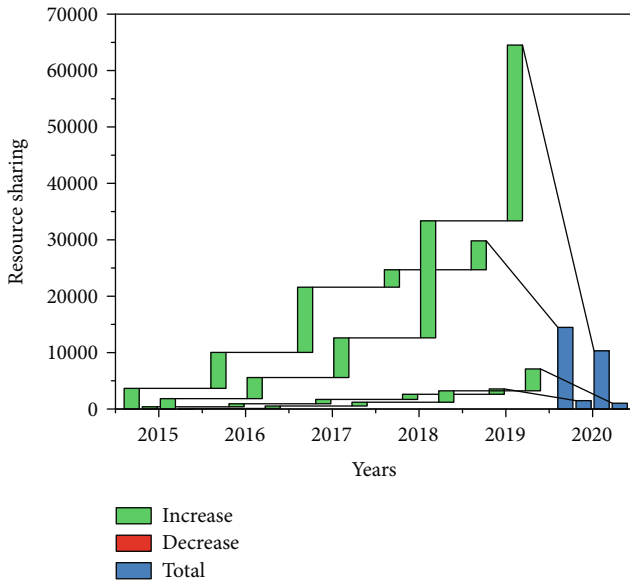


FIGURE 6: Changes in the resource sharing system.

next-generation service platform cannot be implemented for various reasons and the introduction of an integrated system continues to be chosen, then, it is better for all university libraries in the same region or under the same alliance to choose the same type of system. The most typical examples are Tianjin Higher Education Literature Information Center and Jiangsu Higher Education Digital Library, both of whose member libraries uniformly adopt the same type of integrated system, the former using Unicorn system imported from abroad and the latter using Huifen system developed independently in China. See Figure 6.

The scale of collection varies greatly among different types of universities; especially with the increase of electronic resource possession, the goal of collection resource construction in each university library is obviously differentiated, some universities focus on the construction of digital resources, while some universities still focus on paper documents, so the demand for library system is not the same. Before preparing to replace the system, university libraries need to have a clear positioning of the library, consider the philosophy of the library and the structure of the collection and the strength of talents, and understand the needs of the library while clarifying what kind of changes the new system can bring. In the digital era, universities of different levels and types have different orientations. The libraries of universities with “double first-class” construction should have a high level of service capability and provide teachers, students, scholars, and other people with massive literature resources and a more intelligent service environment, so they should mainly consider the next-generation open source library service platform with unified resource management and certain social functions or introduce the new generation of ILS. The libraries of general undergraduate universities focus on the development of key disciplines and the construction of special collections, so the next-generation ILS can meet their needs, as shown in Figure 7. If you want to customize the service independently, then

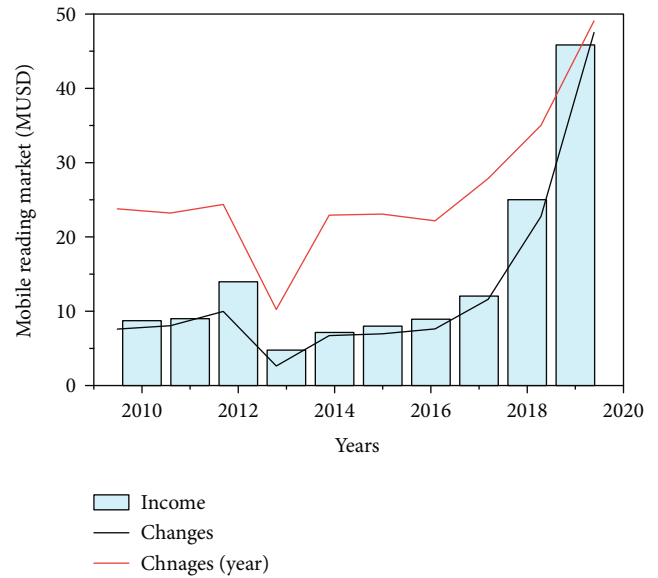


FIGURE 7: Mobile reading market size change.

you can also consider the open source system if the conditions allow. The focus of resource construction in higher education libraries is on paper and electronic resources, so the combination of traditional integrated library system plus electronic resource management system and digital resource management system can also meet most of their needs. See Figure 7.

In order to retain market share and grow further, ILS companies are constantly adopting industry mergers and technological innovations to enhance their competitiveness as competition becomes increasingly fierce around the world. Through a series of acquisitions, foreign ILS companies have not only strengthened their R&D strength and expanded their global market share but also strengthened the competitiveness of their products. However, domestic ILS companies not only do not have the phenomenon of mergers and acquisitions but also have the phenomenon of internal management and technical R&D cadres leaving the original company to start new companies due to the development direction and policy system, which, on the whole, not only weaken the original technical strength of the company but also indirectly take away the original market share of the company. For example, the technical and management team separated from Shenzhen Ketu Company set up Guangzhou Tucson Company, and the technical managers separated from Beijing Xiyang Company set up Qingda Xinyang Company. Researching the current market situation is to grasp the market dynamics and development trend, collect relevant information about the system, and fully understand the functional characteristics of the relevant products. At this stage, we can enter the website of system developers by means of network survey, investigate the basic information of system developers and the overall usage of their products, and record the technical parameters and functional characteristics of the system in detail, carry out in-depth and detailed comparative analysis of each system, summarize their respective advantages and shortcomings,

and make selection evaluation with the support of data, which has the advantage of avoiding blindness at the early stage of selection.

Commercial products and open source products have their own advantages and disadvantages, and all university libraries have to do is to consider them according to the actual situation of the library. The strength of university libraries can be considered in terms of economic strength and librarians' ability. Open source products have good generality, openness, and scalability, but they also require high technical level of librarians. If a university library has a high demand for improving its personalized services and has a stable and high level of technical staff, then open source products are definitely worth considering first. However, for many university libraries, where the library technology is generally low and the librarians' computer skills are average, it is a better choice to choose mature commercial products. Mature commercial products have perfect service and product training system, and there is no particularly high requirement for the comprehensive quality of librarians, which can meet the work requirements of a large proportion of university libraries; after all, the purpose of introducing the system in most libraries is to help them improve their work efficiency.

5. Conclusion

After more than 30 years of development and application, the potential of traditional ILS has almost reached its limit, and the conflict between traditional ILS and the needs of library users has become more intense. Today's libraries have a wider range of options than ever before, ranging from improving legacy ILS to introducing next-generation ILS and participating in the development of next-generation service platforms. This paper focuses on comparative and survey research, comparing and analyzing library integrated systems and library service platforms separately on the one hand and, on the other hand, investigating and analyzing the current application status of library systems in some universities in China. The selection principles and selection strategies are proposed through comparison and summary, so as to provide a reference basis for the selection work of domestic university libraries. College libraries should choose the system suitable for their own needs. Different university libraries differ greatly in terms of collection structure, workflow, and capital budget, so it is important to understand the philosophy and service needs of the library, which helps university libraries to clarify the selection principles and enhance the target of selection at the early stage of selection. Libraries in the digital era should give priority to systems that can manage all library resources comprehensively and effectively adapt to more flexible library workflows and enable libraries to provide better services to users, although traditional ILS has become increasingly difficult to meet the needs of university libraries in the digital age, given that many universities in China still manage print-based documents primarily. Therefore, the traditional library integration system can still meet the needs of some university libraries after improvement and upgrading. In the future, it

is advisable for the library community to plan to join the construction of next-generation service platform projects in the form of regional or industry consortia.

Data Availability

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

Conflicts of Interest

No competing interests exist concerning this study.

Acknowledgments

This study was supported by the School-Level Project of Jiangsu Vocational College of Urban and Rural Construction (2020KYC015).

References

- [1] N. Ahammad, "Open source digital library on open educational resources," *The Electronic Library*, vol. 37, no. 6, pp. 1022–1039, 2019.
- [2] K. Ahmad, Z. Jian Ming, and M. Rafi, "Assessing the digital library research output: bibliometric analysis from 2002 to 2016," *The Electronic Library*, vol. 36, no. 4, pp. 696–704, 2018.
- [3] A. I. Alzahrani, I. Mahmud, T. Ramayah, O. Alfarraj, and N. Alalwan, "Modelling digital library success using the DeLone and McLean information system success model," *Journal of Librarianship and Information Science*, vol. 51, no. 2, pp. 291–306, 2019.
- [4] O. M. Ataeva and V. A. Serebryakov, "Information model of LibMeta digital library," *Lobachevskii Journal of Mathematics*, vol. 40, no. 7, pp. 861–875, 2019.
- [5] T. Bogaard, L. Hollink, J. Wielemaker, J. van Ossenbruggen, and L. Hardman, "Metadata categorization for identifying search patterns in a digital library," *Journal of Documentation*, vol. 75, no. 2, pp. 270–286, 2019.
- [6] L. R. Borrero, P. Schwindingera, A. Castelli et al., "Seeking a common ground for the Nautical Archaeology Digital Library (NADL). Reflections on science, method, theory and templates," *Virtual Archaeology Review*, vol. 12, no. 24, pp. 11–24, 2021.
- [7] J. S. Downie, S. Bhattacharyya, F. Giannetti, E. D. Koehl, and P. Organisciak, "The HathiTrust Digital Library's potential for musicology research," *International Journal on Digital Libraries*, vol. 21, no. 4, pp. 343–358, 2020.
- [8] K. Fenlon, M. Senseney, M. Bonn, and J. Swatscheno, "Humanities scholars and library-based digital publishing: new forms of publication, new audiences, new publishing roles," *Journal of Scholarly Publishing*, vol. 50, no. 3, pp. 159–182, 2019.
- [9] P. Gooding, M. Terras, and L. Berube, "Identifying the future direction of legal deposit in the United Kingdom: the Digital Library Futures approach," *Journal of Documentation*, vol. 77, no. 5, pp. 1154–1172, 2021.
- [10] L. Han, G. Zhang, B. Yong, Q. He, F. Feng, and Q. Zhou, "Statistical study of characteristics of online reading behavior networks in university digital library," *World Wide Web*, vol. 22, no. 3, pp. 1175–1187, 2019.

- [11] J. Jun, J. Song, and C. Kim, "A near-threshold voltage oriented digital cell library for high-energy efficiency and optimized performance in 65nm CMOS process," *IEEE Transactions on Circuits and Systems I Regular Papers*, vol. 65, no. 5, pp. 1567–1580, 2018.
- [12] G. Kortemeyer and S. Dröschler, "A user-transaction-based recommendation strategy for an educational digital library," *International Journal on Digital Libraries*, vol. 22, no. 2, pp. 147–157, 2021.
- [13] S. Li, Z. Hao, L. Ding, and X. Xu, "Research on the application of information technology of big data in Chinese digital library," *Library Management*, vol. 40, no. 8-9, pp. 518–531, 2019.
- [14] Y. Li and C. Liu, "Information resource, interface, and tasks as user interaction components for digital library evaluation," *Information Processing and Management*, vol. 56, no. 3, pp. 704–720, 2019.
- [15] G. M. O’Gara, L. Woolcott, E. Joan Kelly, C. Muglia, A. Stein, and S. Thompson, "Barriers and solutions to assessing digital library reuse: preliminary findings," *Performance Measurement and Metrics*, vol. 19, no. 3, pp. 130–141, 2018.
- [16] A. R. Abdul Rahman and S. Mohezar, "Ensuring continued use of a digital library: a qualitative approach," *The Electronic Library*, vol. 38, no. 3, pp. 513–530, 2020.
- [17] S. Rathee, A. Kumar, S. Kaushik, E. Kazimieras Zavadskas, A. Banaitis, and J. A. Garza-Reyes, "An MCDA cause-effect factors model for the implementation of Greenstone Digital Library software," *Management Decision*, vol. 58, no. 11, pp. 2543–2564, 2020.
- [18] L. Rutherford, A. Singleton, L. A. Derr, and M. K. Merga, "Do digital devices enhance teenagers' recreational reading engagement? Issues for library policy from a recent study in two Australian states," *Public library quarterly*, vol. 37, no. 3, pp. 318–340, 2018.
- [19] Y. Shi and Y. Zhu, "Research on aided reading system of digital library based on text image features and edge computing," *IEEE Access*, vol. 8, pp. 205980–205988, 2020.
- [20] M. Terras, J. Baker, J. Hetherington et al., "Enabling complex analysis of large-scale digital collections: humanities research, high-performance computing, and transforming access to British Library digital collections," *Digital Scholarship in the Humanities*, vol. 33, no. 2, pp. 456–466, 2018.
- [21] L. Wenige and J. Ruhland, "Retrieval by recommendation: using LOD technologies to improve digital library search," *International Journal on Digital Libraries*, vol. 19, no. 2-3, pp. 253–269, 2018.

Research Article

Frequent-Pattern-Based Broadcast Scheduling for Conflict Avoidance in Multichannel Data Dissemination Systems

Chuan-Chi Lai ¹, Yu-De Lin,² and Chuan-Ming Liu ³

¹Department of Information Engineering and Computer Science, Feng Chia University, Taichung 40724, Taiwan

²Wistron (Taiwan), Taipei 11469, Taiwan

³Department of Computer Science and Information Engineering, National Taipei University of Technology, Taipei 10618, Taiwan

Correspondence should be addressed to Chuan-Ming Liu; cmliu@ntut.edu.tw

Received 6 September 2021; Revised 11 November 2021; Accepted 30 November 2021; Published 14 December 2021

Academic Editor: Chao-Yang Lee

Copyright © 2021 Chuan-Chi Lai 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 popularity of mobile devices, using the traditional client-server model to handle a large number of requests is very challenging. Wireless data broadcasting can be used to provide services to many users at the same time, so reducing the average access time has become a popular research topic. For example, some location-based services (LBS) consider using multiple channels to disseminate information to reduce access time. However, data conflicts may occur when multiple channels are used, where multiple data items associated with the request are broadcasted at about the same time. In this article, we consider the channel switching time and identify the data conflict issue in an on-demand multichannel dissemination system. We model the considered problem as a data broadcast with conflict avoidance (DBCA) problem and prove it is NP-complete. We hence propose the frequent-pattern-based broadcast scheduling (FPBS), which provides a new variant of the frequent pattern tree, FP*-tree, to schedule the requested data. Using FPBS, the system can avoid data conflicts when assigning data items to time slots in the channels. In the simulation, we discussed two modes of FPBS: online and offline. The results show that, compared with the existing heuristic methods, FPBS can shorten the average access time by 30%.

1. Introduction

With advances in wireless communications technologies, mobile devices deeply affect our daily lives, such as notebooks, smart phones, and tablets. Users can easily access various information services, such as online news, traffic information, and stock prices. Recently, wireless data dissemination becomes a popular topic [1–3], which can transmit information to a number of users simultaneously. In comparison with the conventional end-to-end transmission (or client-server) model, wireless data dissemination can make use of wireless network channels to reduce the delivery time for obtaining information. Wireless data broadcasting is well-suited to the location-based services (LBS) in an asymmetric communication environment, where a large number of users are interested in popular information such as news [4], traffic reports [5], and multimedia streams [6, 7].

In general, wireless data dissemination can be classified into two modes: push-based and pull-based (on-demand).

In push-based wireless data dissemination environments [8–10], data items are disseminated cyclically according to a predefined schedule. In fact, the access pattern of data items may change dynamically, and the broadcast frequency of popular data items may be lower than the broadcast frequency of unpopular data items. Such a case will result in a poor average access latency. In view of this, pull-based wireless data dissemination [11–13] that disseminates data items timely according to the received requests was proposed to overcome the aforementioned drawback. In the pull-based mode, the users first upload their demand information to the server through the uplink channel, and then, the relevant information will be immediately arranged into the broadcasting channels for disseminating data to users. In wireless data dissemination environments, a way of judging the quality of a scheduling approach is to measure the access time of the generated schedule. The access time is a measured time period from starting tuning the channels to obtaining all the requested information. Thus, it is

important to have a better broadcasting schedule for shorter access time.

1.1. Motivation. In early literature, some conventional works [14–16] focus on how to maximize the bandwidth throughout or minimized the access time in single channel environments. Recently, with the advance on antenna techniques, most of works [17–19] has shifted their focus on the similar issues in multiple channel environments. In general, a multichannel wireless data dissemination system can provide a more network bandwidth and a shorter access time for data dissemination than a single-channel wireless data dissemination system can.

However, one new issue, data conflict [20–22], emerges while each client retrieves data items on multiple channels with channel switching in push-based broadcasting environments. Two types of conflicts may occur in multichannel dissemination systems. The first type of conflict is that two required data items are allocated on the same time slot of different channels, so the client cannot download the required data items simultaneously. The second type of conflict occurs if two required data items are allocated on the t and $(t + 1)$ time slots of different channels, respectively. In such a scenario, the client cannot download both required data items during the time period $[t, t + 1]$. The 1st conflict type is obvious. The reason of the 2nd conflict type is that switching from any channel to a different channel takes time. A client cannot download data at time slot $t + 1$ from one channel if it was downloading data item from another channel at time slot t , because a time slot is already the smallest unit for data retrieving. Note that a client is allowed to access one channel at one time.

Such a data conflict issue makes a client miss its needed data items during the time period for channel switching, thereby leading to a worse access time. On the one hand, some works [20–22] provide some solutions from the client’s point of view. These solutions can make each client schedule itself for retrieving the data items on channels efficiently. On the other hand, only one work [13] provides a server-side scheduling algorithm with consideration of the data conflict issue in on-demand multichannel environments. The provided algorithm considers the associations between data items and requests while allocating data items on multiple channels and this provides a conflict-free schedule.

Most broadcast scheduling techniques in on-demand multichannel data dissemination environments do not consider the time requirement for channel switching, thereby leading to data conflicts or long access time. This phenomenon motivates us to propose a more efficient server-side scheduling method with conflict avoidance using frequent pattern mining technique, thereby shortening the average access time.

1.2. Contribution. In this study, we discuss how to shorten the average access time on a multichannel wireless data dissemination environment under the data conflict conditions. The contributions of this work are listed as follows:

- (1) Identify the data broadcast with conflict avoidance (DBCA) problem in on-demand multichannel wireless data dissemination environments and prove the considered DBCA problem is \mathcal{NP} -complete
- (2) We propose a heuristic approach, frequent-pattern-based broadcast scheduling (FPBS), for providing an approximate schedule in polynomial time. Inspired by frequent-pattern tree (FP-tree), we suggest a new tree, FP*-tree, for FPBS to schedule the requested data items with the consideration of channel switching
- (3) We analyze the time complexity and average access time of FPBS in both average case and worst case
- (4) We verify the performance of FPBS which achieves a shorter average access time in comparison with the existing method, UPF [13]

The rest of this paper is organized as follows. Section 2 gives the background and reviews related research in the literature. Section 3 defines the DBCA problem and proves that the DBCA problem is \mathcal{NP} -complete. Section 4 explains the proposed approach with examples and algorithms in detail. In Section 5, we discuss the time complexity and access time of the proposed approach in worst case. Section 6 presents the experimental simulation results and validates the correctness and effectiveness of the proposed methods in various situations. Finally, we conclude this work in Section 7.

2. Related Work

In the multichannel dissemination environments, many related research works focused on data scheduling to improve the access time performance [17, 18] from the perspective of spectrum utilization. Yee et al. [17] proposed a greedy algorithm to find the best way to distribute data items into the channels, allowing users to access requested data in a limited time. Zheng et al. [18] considered the data access frequency, data length, and channel bandwidth into a model and proposed a two-level optimization scheduling algorithm (TOSA) to find an appropriate schedule. They also showed that the schedule of TOSA is approximate to the best average time. Yi et al. [19] proposed a method to allow replicating multiple copies of a data item in a broadcasting channel. If there are multiple copies of a popular data item in the channel, the average access time can be effectively reduced.

In addition to the above methods, some works considered the priority of incoming queries and found ways to reduce the access time [12, 14, 15, 23]. Lu et al. [14] proposed some algorithms to schedule data for maximum throughput request selection (MTRS) and minimum latency request ordering (MLRO) problems in a single-channel environment and proved that both problems are \mathcal{NP} -hard. Xu et al. [15] proposed a SIN- α algorithm with a set of priority decisions based on the ratio of the length of the expiration time over the amount of information. Lv et al.

[23] proved that minimizing access time in the broadcasting scheduling of multi-item requests with deadline constraint in a single-channel environment is an \mathcal{NP} -hard problem. The authors provided a profit-based heuristic scheduling algorithm to minimize the request miss rate (or delivery miss rate) considering the access frequency of data. Liu and Su [12] focused on reducing the demand for the loss rate and shortening the access time. Two kinds of algorithms, most popular first heuristic (MPFH) and most popular last heuristic (MPLH), were proposed to solve the problems and they also analyzed differences between the online version (the user demands continuously come in the system, so the scheduling task needs to wait until it starts receiving information of the demands) and offline version (the system already has all the information of demands).

Some works had found that the dependency between requested data items may greatly influence the performance of multichannel data broadcasting. Lin and Liu [24] considered the dependencies among data items as a directed acyclic graph (DAG). They proved that finding the best schedule preserving dependencies between each data item is an \mathcal{NP} -hard problem and proposed some heuristics for the problem. Qiu et al. [25] proposed a three-layer on-demand data broadcasting (ODDB) system for enhancing the uplink access capacity by introducing a virtual node layer. Each virtual node can merge duplicated requests and help the server reduce huge computational load, there by improve the broadcasting efficiency.

Lu et al. [20–22] firstly defined two types of well-known data conflicts in multichannel broadcast applications. They proved the client-side retrieval scheduling problem is \mathcal{NP} -hard and provided some client-side data retrieval algorithms for helping clients to retrieve data within multiple channels efficiently. Liu et al. [26] firstly proposed a server-side heuristic data scheduling algorithm, dynamic urgency and productivity (DUP), for on-demand multichannel systems with consideration of the request conflict (or request overlapping) issue and the dependency between requests for scheduling at the request level and giving higher priorities to the requests which are close to their deadlines. Such an approach provided a counteracting effect to the request starvation problem and improved the utilization of broadcasting bandwidth. However, they did not consider two types of data conflicts. He et al. [13] proposed a server-side heuristic scheduling approach, most urgent and popular request First (UPF), with the consideration of two types of data conflicts in on-demand systems. Except for UPF method, the hardness of data scheduling problem considering two types of data conflicts from the server perspective is seldom discussed.

The comparisons of the existing works and this paper are summarized in Table 1. In this work, we propose a new server-side heuristic scheduling approach for providing a conflict-free multichannel data broadcast service with a better performance on the average access time.

3. Problem Description

The length of a broadcasting cycle is an important factor which is normally predefined in the wireless data dissemina-

tion applications. Most of existing data scheduling strategies focus on investigating how to efficiently schedule data items in each broadcasting cycle. To validate the performance of a scheduling strategy, average access time (or average latency), is the commonly and widely used metric. If the average access time is shorter, users generally can obtain all the requested data in a shorter time, meaning that the used scheduling strategy is more efficient. In the following subsections, we will describe the considered system model, define the considered scheduling problem, and then prove the hardness of this problem.

3.1. System Model. In this work, the considered on-demand multichannel data dissemination system is shown in Figure 1 and we only consider the one-hop broadcasting scenario. The considered data dissemination system uses $|C| + 2$ antennas with *orthogonal frequency division multiplexing* (OFDM) technique [27] to provide $|C|$ downlink broadcast channels, 1 downlink index channel, and 1 uplink request channel, where $C = \{c_1, c_2, \dots, c_{|C|}\}$ and $|C| > 1$. The downlink index channel and request uplink channel are denoted as c_{index} and c_{uplink} , respectively. Each user device has two antennas with one for receiving data over the downlink broadcast channels and one for transferring requests via the uplink request channel. We assume that each user device can only access one channel at one time. We assume that all the channels are nonoverlapping, synchronous and discretized into fixed-duration slots. The broadcasting server puts the requests coming from the uplink channel into a buffer with *first-come-first-serve* (FCFS) strategy and handles all the received requests in a batch manner. In this work, we only focus on the efficiency of (application-layer) data/packet scheduling for users to retrieve the requested data items by accessing the downlink channels.

We assume that all the requested data items are in a dataset $D = \{d_1, d_2, \dots, d_{|D|}\}$, where $|D|$ is the size of D , and the length of a broadcasting cycle is $L = |D|$ in default. Suppose that there are n queries, $Q = \{q_1, q_2, \dots, q_n\}$, and each query q_i requests k data items from the dataset D , where $i = 1, 2, \dots, n$ and $k = 1, 2, \dots, |D|$. We let $q_i = \{d_1^i, d_2^i, \dots, d_k^i\}$ and all the data items have the same data size, where $d_j^i \in D$, $j = 1, 2, \dots, k$, and $\cup_{i=1}^n q_i \subseteq D$. Thus, the system has to arrange the requested data items into $|C|$ broadcasting channels. Note that each time slot on a broadcasting channel can contain at most one data item and data replication is only allowed on different channels. That is, multiple copies of one data item may be placed within a broadcasting cycle. Suppose L is the cycle length, each index I_t at time slot t records the informations about all the data items in time slot t' and the corresponding requests of these data items, where t' is obtained by

$$t' = \begin{cases} (t + 2) \bmod L, & \text{if } t + 2 > L, \\ t + 2, & \text{otherwise.} \end{cases} \quad (1)$$

When a client tunes in the channel, it will access the

TABLE 1: Comparisons of related works and the proposed method.

Related works	System model	Criterion	Method	Data popularity	Request dependency	Channel switching	Data conflict
[17–19]	On-demand	Latency	Server-side	No	No	No	No
[12, 14, 15]	On-demand	Latency	Server-side	Yes	No	No	No
[21]	Push-based	Latency	Client-side	No	No	Yes	Yes
[22]	Push-based	Throughput	Client-side	No	No	Yes	Yes
[23]	On-demand	Request miss rate	Server-side	Yes	No	No	No
[24, 25]	On-demand	Latency	Server-side	No	Yes	No	No
[26]	On-demand	Latency	Server-side	Yes	Yes	No	No
[13]	On-demand	Request miss rate	Server-side	Yes	Yes	Yes	Yes
Our work	On-demand	Latency	Server-side	Yes	Yes	Yes	Yes

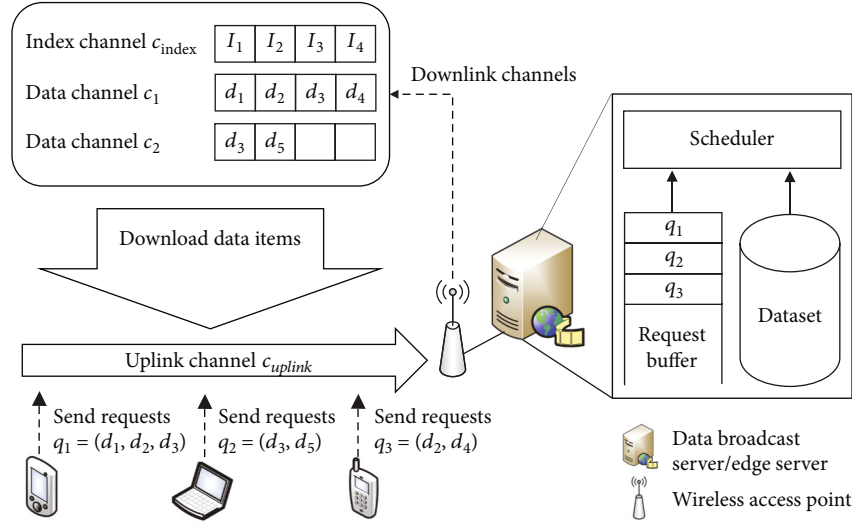


FIGURE 1: The considered on-demand multichannel wireless data dissemination environment.

index channel in advance until obtaining information about the first required data item.

3.2. Problem Formulation. The considered scheduling problem can be treated as a mapping \mathcal{M} that data items associated with all the queries to $|C|$ broadcasting channels. For each data item $d_j^i \in D$ associated with a query $q_i \in Q$, let $\text{pos}(q_i, d_j^i) = (c_j^i, p_j^i)$ be the position of data item d_j^i in the broadcast, where c_j^i is the channel number, $1 \leq c_j^i \leq |C|$, and p_j^i is the location of d_j^i on that channel, $1 \leq p_j^i \leq |D|$. Such a mapping $\mathcal{M} : Q \times D \rightarrow \mathbb{N} \times \mathbb{N}$ is a 1-to-1 mapping.

Since there are multiple channels and each user can only tune into one broadcasting channel at one time instance, each user may switch channels many times for retrieving all the requested data items on different channels. In general, channel switching is a relatively fast operation (in the micro-second range) [28, 29]. For simplicity, we follow the similar assumptions about channel switching in [22], and each channel switching takes one time slot in the considered data dissemination environment. Figure 2 shows an example of the channel switching. However, channel switching may cause a new problem, data conflict, in multichannel wireless data dissemination systems. For example, if one of requested

data items for request q_i is placed at the previous, the same, or the later location of a scheduled data item which is also associated with q_i on different channels, a data conflict occurs. An example of data conflicts is presented in Figure 3. The data conflict may result in a longer access time and can be defined as Definition 1.

Definition 1 (data conflict). For a query q_i , two requested data items d_j^i and $d_{j'}^i$, $1 \leq j \neq j' \leq k$, if $c_j^i \neq c_{j'}^i$, the conflict occurs when $p_j^i = p_{j'}^i$ or $|p_j^i - p_{j'}^i| = 1$.

Let loc_{\min}^i denote the minimum value of all the locations of the data items associated with q_i and loc_{\max}^i is the maximum value of all the positions of the data items associated with q_i . In other words, $\text{loc}_{\min}^i = \min_{1 \leq j \leq k} p_j^i$ and $\text{loc}_{\max}^i = \max_{1 \leq j \leq k} p_j^i$. The access time of query q_i , $\text{acc}(q_i)$, can be defined as $|\text{loc}_{\max}^i - \text{loc}_{\min}^i|$, while the search starts from the beginning of the broadcasting cycle. The average access time for a mapping \mathcal{M} is thus $\text{acc}_{\mathcal{M}} = \sum_{i=1}^n \text{acc}(q_i)/n$.

In summary, the problem we want to solve in this work is *data broadcast with conflict avoidance (DBCA)* problem which can be defined as follows.

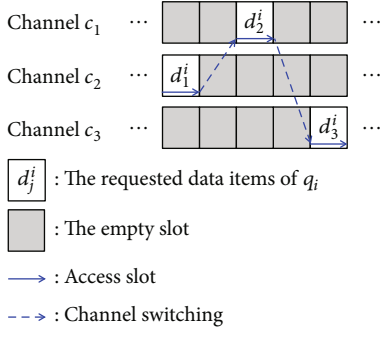


FIGURE 2: An example of channel switching, where the data items d_1^i , d_2^i , and d_3^i are requested by q_i .

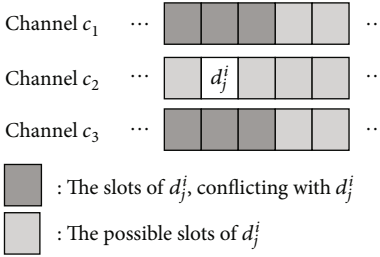


FIGURE 3: An example of the data conflict problem.

Definition 2 (DBCA problem). Suppose all the notations are defined as above. The DBCA problem is to find a mapping $\mathcal{M} : Q \times D \rightarrow \{1, \dots, |C|\} \times \{1, \dots, L\}$ such that

- (1) there is no data conflict for each query in the mapping, i.e., w.r.t. query q_i , for each pair of data items d_j^i and $d_{j'}^i$, $1 \leq j \neq j' \leq k$, we have $|p_j^i - p_{j'}^i| > 1$ when $c_j^i \neq c_{j'}^i$; and
- (2) the average access time of \mathcal{M} , $\text{acc}_{\mathcal{M}} = \sum_{i=1}^n \text{acc}(q_i)/n$, is minimized.

3.3. NP-Completeness. To the best of our knowledge, most of the existing works only considered the schedules without data replication in a broadcasting cycle. They did not discuss and analyze the schedules with conflict avoidance problem on multichannel dissemination environments in detail. Conversely, our proposed approach, FPBS, considers a multichannel dissemination environment which allows replicating data items on different channels of a broadcasting cycle. In such a scenario, we investigate the data conflict problem and propose a new approach to avoid this problem. In this subsection, we will prove DBCA problem is \mathcal{NP} -complete.

In the definition of DBCA problem, the first objective indicates that the broadcasting schedule avoids the data conflict problem. The second objective is to minimize the average access time. Since the server has no prior knowledge about the coming requests, the process for scheduling the broadcasting is made in an online fashion. We first look at the offline version of the DBCA problem in the follow-

ing and it refers to conflict-free data broadcasting with minimum average latency (CDBML) problem and define it as below.

Definition 3 (CDBML problem). Instance: There are $|C|$ data broadcasting channels with cycle length L , a set of $|D|$ data items $D = \{d_1, \dots, d_{|D|}\}$, and a set of n requests $Q = \{q_1, \dots, q_n\}$. Each request q_i , $1 \leq i \leq n$, is associated with k data items, $d_1^i, d_2^i, \dots, d_k^i$, where $d_j^i \in D$, $1 \leq j \leq k \leq |D|$. Any two data items associated with two different requests are different, and every data item needs an unit time u_t to be broadcast. Let loc_{\min}^i and loc_{\max}^i be the start time and finish time of q_i , respectively.

Question: Does there exist a mapping $\mathcal{M} : Q \times D \rightarrow \{1, \dots, |C|\} \times \{1, \dots, L\}$ such that

- (1) For two data items d_j^i and $d_{j'}^i$ associated with q_i , $|p_j^i - p_{j'}^i| > 1$ and
- (2) the average access time, $\sum_{i=1}^n |\text{loc}_{\max}^i - \text{loc}_{\min}^i|/n$, is minimized

In the definition of CDBML problem, the first objective indicates that the broadcasting schedule avoids the data conflict problem. The second objective is to reduce the average access time and all of the data items associated some request q_i should be broadcasting before the end of the broadcasting cycle. W_i is an indication function used to present if a request is served or not. To show further that the CDBML problem is \mathcal{NP} -complete, we consider a special case of it, where the number of data items associated with each request is the same and equal to the number of channels. That is, we consider the case $k = |C|$. The data items associated with different requests are all different. The following gives the definition of the decision problem for the above special case.

Definition 4 (CDBML ρ problem). Instance: There are $|C|$ data broadcasting channels with cycle length L , a set of $|D|$ data items $D = \{d_1, \dots, d_{|D|}\}$, a set of n requests $Q = \{q_1, \dots, q_n\}$, and an integer h . Each request q_i , $1 \leq i \leq n$, is associated with $|C|$ data items, $d_1^i, d_2^i, \dots, d_{|C|}^i$, where $d_j^i \in D$, $1 \leq j \leq |C| \leq |D|$. Any two data items associated with two different requests are different, and every data item needs an unit time u_t to be broadcast. Let loc_{\min}^i and loc_{\max}^i be the start time and finish time of q_i , respectively.

Question: Does there exist a mapping $\mathcal{M} : Q \times D \rightarrow \{1, \dots, |C|\} \times \{1, \dots, L\}$ such that

- (1) For two data items d_j^i and $d_{j'}^i$ associated with q_i , $|p_j^i - p_{j'}^i| > 1$ and
- (2) $\sum_{i=1}^n \text{acc}(q_i)/n \leq h$, where $\text{acc}(q_i) = |\text{loc}_{\max}^i - \text{loc}_{\min}^i|$

To show that the CDBML ρ problem is NP-complete, we reduce the minimizing mean flow time in unit time open

shop (MMUOS) scheduling [30] problem with preemption ($O | p_{i,j} \in \{0, 1\}; \text{pmtn} | \Sigma C_i$) to the CDBML ρ problem. [30] has proved such a problem ($O | p_{i,j} \in \{0, 1\}; \text{pmtn} | \Sigma C_i$) is \mathcal{NP} -hard by the reduction from the *graph coloring* problem, and thus, the CDBML problem is \mathcal{NP} -hard. The MMUOS problem is defined as follows.

Definition 5 (MMUOS problem). Instance: Given m machines, a set of n jobs $J = \{J_1, J_2, \dots, J_n\}$, a set of $|O|$ unit operations $O = \{o_1, o_2, \dots, o_{|O|}\}$, and an integer T . Each job J_i , $1 \leq i \leq n$, consists of m unit operations o_i^j , where $o_i^1, o_i^2, \dots, o_i^m$. The j th operation, $1 \leq j \leq m$, has to be processed on the j th machine. Job J_i will be processed in a window defined by a release time r_i and a finish time c_i .

Question: Does there exist a mapping $\mathcal{M} : Q \times D \rightarrow \{1, \dots, |C|\} \times \{1, \dots, L\}$ such that

- (1) For operations o_i^j and o_i^l in job J_i , $\mathcal{M}(o_i^j) \neq \mathcal{M}(o_i^l)$ and
- (2) $\sum_{i=1}^n C_i/n \leq T$, where $C_i = |c_i - r_i|$

Theorem 6. *The CDBML ρ problem is \mathcal{NP} -complete.*

Proof. It is easy to see that the CDBML ρ problem is in \mathcal{NP} , since validating the existence of an given conflict-free schedule simply needs polynomial time. In order to prove the CDBML ρ problem is \mathcal{NP} -hard, a reduction from the MMUOS problem can be made. Suppose that I' is an instance of the MMUOS problem. A corresponding instance I of the CDBML ρ problem can be constructed from I' as follows:

- (1) An unit operation time is equal to the unit time slot to broadcasting a data item
- (2) Let a job J_i correspond to a request q_i , $1 \leq i \leq m$ and operations o_i^j in J_i be the data item d_i^j associated with q_i
- (3) Let m machines be the $|C|$ data broadcasting channels (i.e., $m = |C|$)
- (4) Let J_i 's release time r_i be q_i 's start time loc_{\min}^i in the schedule
- (5) Let J_i 's finish time c_i be q_i 's finish time loc_{\max}^i in the schedule
- (6) Let integer T be the integer h in CDBML ρ problem.
- (7) Let the unit time u_i' in MMUOS problem be three times of u_i in CDBML ρ problem ($u_i' = 3 * u_i$)

According to the last step of the construction, the first objective of MMUOS problem can be equivalent to the first objective of CDBML ρ problem and the above construction can be done in polynomial time. It is straightforward to show that there is a solution for an instance I' of the MMUOS problem if and only if there is a solution for the

instance I of the CDBML ρ problem since the reduction is a one-to-one mapping for the variables from the MMUOS problem to the CDBML ρ problem. Hence, the CDBML ρ problem is \mathcal{NP} -complete. \square

Thus, we can conclude the following theorem.

Theorem 7. *The CDBML problem is \mathcal{NP} -complete.*

4. Frequent-Pattern-Based Broadcast Scheduling

In this section, we propose an approach, the *frequent-pattern-based broadcast scheduling* (FPBS), to shorten the average access time per user for the DBCA problem. In FPBS, we construct a new tree with the frequent patterns of queries. This tree is named as FP*-tree. FPBS includes four stages: (1) sorting requested data items, (2) constructing the FP*-tree's backbone, (3) constructing the FP*-tree's accelerating branches, and (4) schedule mapping. In the following, the proposed method will be introduced with a running example in detail.

4.1. Stage 1: Sorting Requested Data Items. We consider a running example which uses two data broadcasting channels c_1, c_2 and an additional index channel c_{index} . The data dissemination server receives five queries $q_1 = \{d_2, d_5, d_7\}$, $q_2 = \{d_2, d_3, d_4\}$, $q_3 = \{d_1, d_3, d_6\}$, $q_4 = \{d_1, d_3, d_4, d_5\}$, and $q_5 = \{d_2, d_5, d_8\}$ and then derives the access frequency f_{d_j} of each data item d_j in these queries. After that, the server sorts all the data items in each query according to the descending order of their access frequencies and also derives the statistical average access frequency f_{q_i} of each query q_i . For example, $f_{q_1} = (f_{d_2} + f_{d_5} + f_{d_7})/|q_1| = (3 + 3 + 1)/3 = 2.33$. Hence, the final result is presented in Table 2.

The detailed process, FPBS_StatisticAndSort(Q), for the first stage is presented in Algorithm 1. Line 2 and Line 3 analyze the received query set Q , derive the statistical information, and save it as a temporary set S . The operations from Line 4 to Line 6 sort every requested data item of each query according to the access frequency of the data item. As the example shown in Table 2, the orders of requested data items in queries q_4 and q_5 change after the sorting. Line 7 and Line 8, respectively, store the results in two lists, $\text{list}_{\text{SortedWithSize}}$ and $\text{list}_{\text{SortedWithFre}}$, in different orders. Finally, the process returns these two lists at Line 9 for the use in following stages.

4.2. Stage 2: Constructing the FP*-Tree's Backbone. After deriving some statistical information and the sorting result in Table 2, the system starts to create the backbone of a FP*-tree. In this stage, the system will always select the query which requests the most number of data items to be inserted into the FP*-tree in advance. If there are multiple queries which request the same number of data items, the system will select the one which has the maximum average access frequency f_{q_i} . Thus, the system select q_4 as the first query to construct the backbone of a FP*-tree and the result

TABLE 2: The sorted result of requested data items.

Query	Requested data items	Sorted result	f_{q_i}
q_1	d_2, d_5, d_7	d_2, d_5, d_7	2.33
q_2	d_2, d_3, d_4	d_2, d_3, d_4	2.67
q_3	d_2, d_5, d_8	d_2, d_5, d_8	2.33
q_4	d_1, d_3, d_4, d_5	d_3, d_5, d_1, d_4	2.5
q_5	d_1, d_3, d_6	d_3, d_1, d_6	2

is shown in Figure 4(a). After adding q_4 to the FP*-tree, the system will update the statistical information of unhandled queries, as shown in Table 3.

After updating the statistical information, the system will select the next query to handle in the same way. In the previously mentioned, both q_1 and q_3 request 2 data items so the system will compare the remaining average access frequencies of q_1 and q_3 ($f_{q_1} = 2, f_{q_3} = 2$) and both values are the same. Then, the query which comes into the system first will be selected, so q_1 becomes the next one in this step. Note that the numbering of q_1 is smaller than q_3 's and it means that q_1 comes into the system earlier. Thus, the handling priority of the remaining queries is $q_1 \rightarrow q_3 \rightarrow q_5$. While adding data item d_2 into the FP*-tree, the system needs to consider the relations between d_2 and the other queries. In this case, q_2 and q_3 also request the data item d_2 . The system then checks the other data items which are in the request list of both queries and have been added into the FP*-tree. Since the level of d_4 is larger than d_3 's level, the system will insert d_2 as d_4 's child. Such a way can avoid increasing the access time of q_4 which has been handled. After handling d_2 , the system handles d_7 in the same way and the result of FP*-tree is shown in Figure 4(b). The system then updates the statistical information which is presented in Table 4.

The next query which will be handled is q_3 . Since there are no other queries relating to the requested data item d_8 , the system needs to add d_8 after d_2 according to the order of q_3 's requested list. However, d_2 is also requested by q_1 , and thus, d_2 already has one branch and the position is occupied by d_7 . Therefore, d_8 can only be scheduled in the level (time slot) after d_2 and d_7 . In this case, the system creates a new branch of d_2 and inserts an empty node between d_2 and d_8 . Note that an empty node is a node without saving any data item. After handling q_3 , the results are shown in Figure 4(c). The last query is q_5 , and there are no other queries relating to d_6 . Hence, the system has to add d_6 after d_1 according to the order of q_5 's requested list. However, d_1 is also requested by q_4 , and thus, d_6 needs to be scheduled after d_4 . In this case, the system creates a new branch of d_1 and inserts an empty node between d_1 and d_6 . Finally, the construction of FP*-tree's backbone is finished and the result is shown in Figure 4(d).

Algorithm 2 presents two functions for the backbone construction. FPBS_CreateBackbone(S) describes the main process of an FP*-tree's backbone construction and FPBS_AddNodeForBackbone(\mathcal{T}, N_p, d) is the function of adding

a node during the backbone construction. From Line 3 to Line 5, the operations initialize an empty FP*-tree \mathcal{T} and create a sorted query table Q_{table} with the derived sorted result in the stage 1. The operations from Line 6 to Line 8 handle each requested data item of the first query in the sorted query set. The first query is the most important and has maximum number of requested data items. As shown as the above example in Figure 4(a), the query q_4 is the first to be handled. At Line 9, the remaining information of unhandled queries and data items in the query table Q_{table} will be updated. From Line 10 to Line 17, the operations continuous inserting the unhandled data items of Q_{table} into the backbone of \mathcal{T} . At Line 13, the operation finds the right position of \mathcal{T} 's backbone to insert the unhandled data item with the consideration of query dependency and the order of data items. The operations from Line 21 to Line 35 presents the detailed process of adding a data node to the backbone of \mathcal{T} . Note that the operation, $\mathcal{T}.\text{isOverload}(N_{\text{temp}}.\text{slot} + 1)$, at Line 26 is used to avoid scheduling data items out of $|C|$ data broadcasting channels. Figures 4(c) and 4(d) are the running examples for such operations.

4.3. Stage 3: Constructing the FP*-Tree's Accelerating Branches. After the construction of FP*-tree's backbone, the system starts to create the accelerating branches to optimize the schedule. The purpose of constructing the accelerating branch is to increase the chance of each user getting the requested data item earlier after switching channels.

In this stage, we propose two different ordering rules, *request-number-first* and *frequency-first*, to insert data items in the FP*-tree's accelerating branches. The priority of a query for the insertion of FP*-tree is decided by following values: number of requested data items, average access frequency, and arrival time. With request-number-first rule, the system will select the query which requests the maximum number of data items to handle first. If multiple queries request the maximum number of data items, the system will select the one of them that has the maximum average access frequency. If multiple queries has the maximum average access frequency unfortunately, the system will select the query according to its arrival order. Conversely, with frequency-first rule, the system will first select the query which has the maximum average access frequency. If multiple queries has the maximum average access frequency, the system will select the one of them that requests the maximum number of data items. If multiple queries requests the maximum number of data items unfortunately, the system will select the query according to its arrival order. Note that the construction of the FP*-tree's backbone always follows the request-number-first rule in our design. The system can use different rules only when constructing accelerated branches of the FP*-tree.

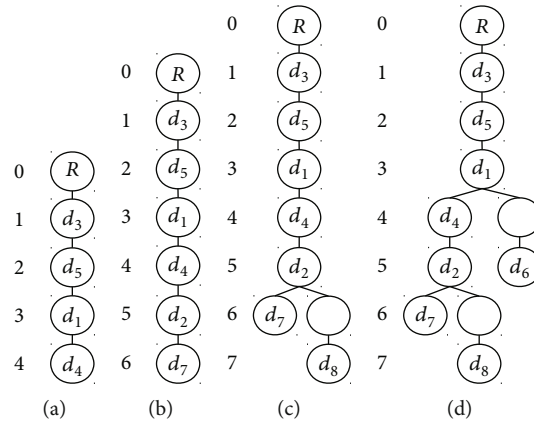
Since different orders of handling queries and data items make the process constructs different accelerating branches of FP*-trees, we will compare the performance results of different schedules generated by using different rules. By default, the system uses frequency-first rule to select the query for constructing the FP*-tree's accelerating branches. Due to limitations on space and the similar process, we only

```

1. Function FPBS_StatisticAndSort ( $Q$ )
   Input: a set of queries (clients)  $Q$ ;
   Output: two lists of sorted queries with sorted requested data,  $\text{list}_{\text{SortedWithSize}}$ ,  $\text{list}_{\text{SortedWithFre}}$ ;
2 create a temporary set  $S \leftarrow \emptyset$ ;
3  $S \leftarrow \text{StatisticDataFrequency}(Q)$ 
4 for each query  $q$  in  $Q$  do;
5    $\text{sortRequiredDataByFrequency}(q, S)$ 
6 end;
7  $\text{list}_{\text{SortedWithSize}} \leftarrow \text{sortQuerySetByQuerySize}(S)$ 
8  $\text{list}_{\text{SortedWithFre}} \leftarrow \text{sortQuerySetByAverageFrequency}(S)$ 
9 return  $\text{list}_{\text{SortedWithSize}}$ ,  $\text{list}_{\text{SortedWithFre}}$ ;
10 end;

```

ALGORITHM 1: Deriving the statistical information and sorted result.

FIGURE 4: Constructing the backbone of a FP*-tree step-by-step: (a) add q_4 , (b) add q_1 , (c) add q_3 , and (d) add q_5 .TABLE 3: Updated result after handling q_4 .

Query	Unhandled data items	Items added in FP*-tree's backbone	f_{q_i}
q_1	d_2, d_7	d_5	2
q_2	d_2	d_3, d_4	3
q_3	d_2, d_8	d_5	2
q_4	\emptyset	d_3, d_5, d_1, d_4	0
q_5	d_6	d_3, d_1	1

TABLE 4: Updated result after handling q_1 .

Query	Unhandled data items	Items added in FP*-tree's backbone	f_{q_i}
q_1	\emptyset	d_5, d_2, d_7	0
q_2	\emptyset	d_3, d_4, d_2	0
q_3	d_8	d_5, d_2	1
q_4	\emptyset	d_3, d_5, d_1, d_4	0
q_5	d_6	d_3, d_1	1

introduce the proposed approach with frequency-first in detail. In this example, the system follows the frequency-first rule and gets the following handling sequence: $q_2 \rightarrow q_4 \rightarrow q_1 \rightarrow q_3 \rightarrow q_5$. Note that the value of f_{q_i} is shown in Table 2.

The system first handles query q_2 and q_2 's sorted requested data items are d_2, d_3 , and d_4 . Hence, the system sequentially schedules d_2, d_3 , and d_4 . When scheduling d_2 , the system temporarily inserts d_2' into level (or slot) 1 and the position is a right child of the root. Then, the system searches d_2 in the backbone and check whether $p_2 > p_2'$ and $p_2 - p_2' > 1$ or not. In this case, $p_2 > p_2'$ and $p_2 - p_2' = 4 > 1$ is hold, so d_2 can be inserted into the position of d_2' .

For the next requested data item d_3 , the system inserts d_3' after d_2 in the accelerating branch and then checks whether the position is legal or not in the same way. In this case, d_3 can be inserted into the position of d_3' . For the last requested data item d_4 by query q_2 , the system tries to temporarily insert d_4' after d_3 in the accelerating branch. However, the system can find d_4 in the backbone that $p_4 - p_4' \leq 1$. Thus, d_4 can not be inserted into the accelerating branch. After handling q_2 , the result of FP*-tree is shown in Figure 5(a).

For the next query q_4 , the system will do nothing in the accelerating branch. The reason is that q_4 is the first query handled in the backbone and the schedule, $d_3 \rightarrow d_5 \rightarrow$

```

1: Function FPBS_CreateBackbone ( $S$ )
   Input: a sorted set of queries (clients)  $S$ ;
   Output: a basic FP*-tree  $\mathcal{T}$ ;
2: create a empty FP*-tree  $\mathcal{T}$  and the root  $R$  of  $\mathcal{T}$ ;
3: set  $S$  into a query table  $Q_{table}$ ;
4: let  $q \leftarrow S.first()$ ;
5: let a temporary pointer  $N_{curr} \leftarrow R$ ;
6:   for each requested data  $d$  in  $q$  do;
7:      $N_{curr} \leftarrow FPBS\_AddNodeForBackbone(\mathcal{T}, N_{curr}, d)$ ;
8:   end;
9: update  $Q_{table}$ ;
10: while  $Q_{table}$  contains any unhandled required data do
11:    $q_{un} \leftarrow$  the query with the maximum number of unhandled data items in  $Q_{table}$ ;
12:   for each unhandled requested data  $d'$  in  $q_{un}$  do;
13:      $N_{d-p} \leftarrow$  find the other queries which also needs data  $d'$  and then choose one of the handled data nodes whose slot is
maximum in  $\mathcal{T}$ 
14:      $FPBS\_AddNodeForBackbone(\mathcal{T}, N_{d-p}, d')$ ;
15:   end;
16:   update  $Q_{table}$ ;
17: end;
18: return  $\mathcal{T}$ ;
19: end;
20: Function FPBS_AddNodeForBackbone ( $\mathcal{T}, N_p, d$ )
   Input: an FP*-tree  $\mathcal{T}$ , the parent node  $N_p$ , and a new data item  $d$ ;
   Output: an added node  $N_d$ ;
21: create a new node  $N_d$  with data item  $d$ 
22: if  $N_p$  has children then
23:   create an empty node  $N_e$ ;
24:    $N_p.addChild(N_e)$ ;
25:   let a temporary pointer  $N_{temp} \leftarrow N_e$ ;
26:   while  $\mathcal{T}.isOverload(N_{temp}.slot+1)$  do
27:     create an empty node  $N_e$ ;
28:      $N_p.addChild(N_e)$ ;
29:      $N_{temp} \leftarrow N_e$ ;
30:   end;
31:    $N_{temp}.addChild(N_d)$ ;
32: else
33:    $N_p.addChild(N_d)$ ;
34: end;
35: return  $N_d$ ;
36: end;

```

ALGORITHM 2: Functions used for the FP*-tree's backbone construction.

$d_1 \rightarrow d_4$, has been optimized. Go on the next step, q_1 is going to be handled and q_1 's requested data items are d_2 , d_5 , and d_7 . Since d_2 has been inserted into the accelerating branch, the system skips d_2 and tries to insert d_5 in this step. According to the order of q_1 's requested list, d_5 needs to be inserted after d_2 . In the accelerating branch, node d_2 already has a child, so the system creates a new branch of d_2 , inserts an empty node as d_2 's right child, and then add temporary d_5' after the empty node. Since there is no d_5 whose $p_5 > p_5'$ in the backbone, it is legal to insert d_5 at the position of d_5' . For the last requested data item d_7 in q_1 , d_7 is inserted in the same way. The system inserts d_7' after d_5 in advance and check whether the backbone contains d_7 or not. Since $p_7 > p_7'$ and $p_7 - p_7' = 2 > 1$, it is legal to insert d_7 at the posi-

tion of d_7' . After handling all the requested data items in q_1 , the result of FP*-tree is shown in Figure 5(b)

After handling q_1 , the system will start to handle q_3 . The sorted requested data items are d_2 , d_5 , and d_8 . Since d_2 has been scheduled at the first slot (level) in the accelerating branch, the system skips d_2 in this step. The next data item d_5 also has been scheduled in the accelerating branch while handling the previous query q_1 . Hence, the system only needs to handle d_8 for q_3 . According to the requested list of q_3 , d_8 needs to be inserted at a position that is after d_2 and d_5 . In the accelerating branch, $p_5 > p_2$ so that d_8 will be inserted under the d_5 . Since d_5 already has a branch, the system creates a new branch of d_5 , inserts an empty node after d_5 , and tries to inserts a temporary d_8' after the empty node

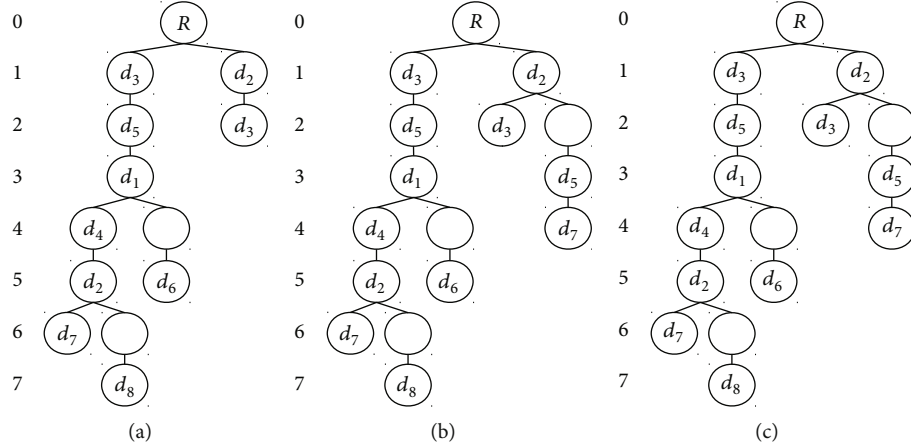


FIGURE 5: Constructing the accelerating branch of the FP*-tree step-by-step: (a) add q_2 , (b) add q_1 , and (c) add q_5 .

(at $p_8' = 5$). However, $C = 2$ and the bandwidth has been occupied by d_2 and d_6 at slot $p_8' = 5$. Then, the system will insert an empty node again and try to add a temporary d_8' at position $p_8' = 6$. Then, the system starts to find d_8 in the backbone and check whether $p_8 > p_8'$ and $p_8 - p_8' > 1$ or not. In this case, $p_8 - p_8' = 1$, so it is illegal to place d_8 at the position of d_8' and the system removes all the empty nodes after d_5 in the accelerating branch. Hence, the final FP*-tree is shown in Figure 5(c).

Algorithm 3 presents the pseudocodes for the functions of accelerating branch construction. FPBS_CreateAcceleratingBranch(\mathcal{T}, S) is the main function for constructing accelerating branch. The process calls the subfunction FPBS_AddNodeForAcceleratingBranch($\mathcal{T}, N_{\text{curr}}, d$) to insert a data item into the accelerating branch of \mathcal{T} at Line 6. Such a process is similar to the function FPBS_AddNodeForBackbone(\mathcal{T}, N_p, d) in the backbone construction. The operation at Line 7 calls another subfunction FPBS_RangeSearch(\mathcal{T}, N_p) to check whether the inserted data item is in the search range (or levels) or not. The insertion will be illegal if the same data item in the backbone of \mathcal{T} locates at one of search levels. If the insertion is illegal, the inserted nodes (including the data item and empty node(s)) will be deleted at Line 47.

4.4. Stage 4: Schedule Mapping. After finishing stage III, the system will map every slot (or level) of FP*-tree into the broadcasting channels using the breadth-first-search (BFS) strategy. The final results are shown in Figure 6. Note that the maximum number of data items in each slot (level) is the number of channels, $|C|$. The mapping process is described as the operations before Line 24 in Algorithm 4. From Lines 25 to 29, the process schedules the index items in index channel and the result is shown in Figure 6. According to the indexing rule defined in (1), the index I_1 records the information about who requests the data items in slot 3 and the index I_6 records the similar information corresponding to the data items in slot 1.

Consider the example of Table 1, for the request $q_2 = \{d_2, d_3, d_4\}$, the final schedule in Figure 6 generated by the

proposed FPBS shows that the user can retrieve all the requested data items d_2, d_3 (on c_2), and d_4 (on c_1) within 4 time slots including a channel switching. If there is no accelerating branch, the user needs 5 time slots to retrieve data items d_2, d_3 , and d_4 on c_1 . This result shows that the proposed FP*-tree can indeed reduce the access time.

5. Analysis and Discussion

In this section, we analyze the performance of FPBS in terms of *time complexity*, *space complexity*, and *access time*.

5.1. Time Complexity. Suppose that the notations are defined as above and the FP*-tree is denoted as \mathcal{T} , then, the time complexity of the \mathcal{T} 's construction will be $\mathcal{O}(nk)$. The idea of FP*-tree design comes up from the FP-tree and only one difference between them is that FP*-tree needs to add an empty node when creating a new branch except for the root node. In the last stage of the proposed method, schedule mapping needs to map all the data nodes of \mathcal{T} to the broadcasting channels and $|\mathcal{T}| \leq nk$, so the time complexity of schedule mapping is also $\mathcal{O}(nk)$. Due to the nature of the FP*-tree which is evolved from FP-tree, FPBS costs $\mathcal{O}(nk)$ in both average case and worst case. In summary, FPBS provides a polynomial algorithm for solving the DBCA problem.

5.2. Space Complexity. After discussing the time complexity of FPBS, we start to analyze the space complexity of FPBS. In this part, we only consider the temporary space for FPBS process. In the stage 1 of FPBS process, the system uses a $\mathcal{O}(nk)$ size table to store the sorted requests and the statistical information. In the stage 2, the system uses the obtained sorted table to construct the backbone of an FP*-tree and it also costs $\mathcal{O}(nk)$ space. In the stage 3, the system constructs accelerating branches of the FP*-tree and it costs $\mathcal{O}(nk')$ space, where $1 \leq k' \leq k$. In the last stage, the system just maps the FP*-tree to the channels and only costs $\mathcal{O}(1)$ additional temporary space for traversing the FP*-tree. That is, the temporary space complexity during the scheduling process is $\mathcal{O}(nk)$.

```

1: Function FPBS_CreateAcceleratingBranch ( $\mathcal{T}, S$ )
   Input: an FP*-tree  $\mathcal{T}$  and a sorted set of queries (clients)  $S$ 
   Output: a final FP*-tree  $\mathcal{T}$ 
2: let a temporary pointer  $N_{curr} \leftarrow R$ ;
3: create a temporary list  $list_q$  and a temporary node  $N_{temp}$ ;
4: for each query  $q$  in  $S$  do
5:   for each requested data  $d$  in  $q$  do
6:      $N_{temp} \leftarrow$  FPBS_AddNodeForAcceleratingBranch ( $\mathcal{T}, N_{curr}, d$ );
7:      $N_{curr} \leftarrow$  RangeSearch ( $\mathcal{T}, N_{temp}$ );
8:      $list_q.add(N_{curr})$ ;
9:     if  $N_{curr}.slot > \mathcal{T}.slot$  then
10:       delete the path of  $list_q$  in  $\mathcal{T}$ ;
11:       break;
12:     end
13:   end
14:    $list_q.clear()$ ;
15: end
16: return  $\mathcal{T}$ ;
17: end
18: Function FPBS_AddNodeForAcceleratingBranch ( $\mathcal{T}, N_p, d$ )
   Input: an FP*-tree  $\mathcal{T}$ , the parent node  $N_p$ , and a new data item  $d$ 
   Output: an added node  $N_d$ 
19: create a new node  $N_d$  with data item  $d$ ;
20: if  $N_p$  has children then
21:   if  $N_p$  has a child  $N_d'$  with  $d$  then
22:     return  $N_d'$ ;
23:   else
24:     create an empty node  $N_e$ ;
25:      $N_p.addChild(N_e)$ ;
26:     let a temporary pointer  $N_{temp} \leftarrow N_e$ ;
27:     while  $\mathcal{T}.isOverload(N_{temp}.slot+1)$  do
28:       create an empty node  $N_e$ ;
29:        $N_p.addChild(N_e)$ ;
30:        $N_{temp} \leftarrow N_e$ ;
31:     end
32:     create a new node  $N_d$  with data item  $d$ ;
33:      $N_{temp}.addChild(N_d)$ ;
34:   end
35: else
36:   create a new node  $N_d$  with data item  $d$ ;
37:    $N_p.addChild(N_d)$ ;
38: end
39: return  $N_d$ ;
40: end
41: Function FPBS_RangeSearch ( $\mathcal{T}, N_p$ )
   Input: an FP*-tree  $\mathcal{T}$  and a search node  $N_d$ 
   Output: a result node  $N_d$  within the search range
42: int  $Num_e \leftarrow$  the number of  $N_d$ 's ancestors which are empty;
43: int  $startSlot \leftarrow N_{temp}.slot - Num_e + 1$ ;
44: int  $endSlot \leftarrow N_{temp}.slot + 1$ ;
45: for  $i \leftarrow startSlot$  to  $endSlot$  do
46:   if find a node  $N_{temp}$  that has the same data as  $N_d$  does at level  $i$  of  $\mathcal{T}$  then
47:     delete the path that contains  $N_d$  and all the empty connected ancestors of  $N_d$ ;
48:     return  $N_{temp}$ ;
49:   end
50: end

```



```

51:   return  $N_{d_i}$ ;
52:   end

```

ALGORITHM 3: Functions used for the construction of the FP*-tree's accelerating branch.

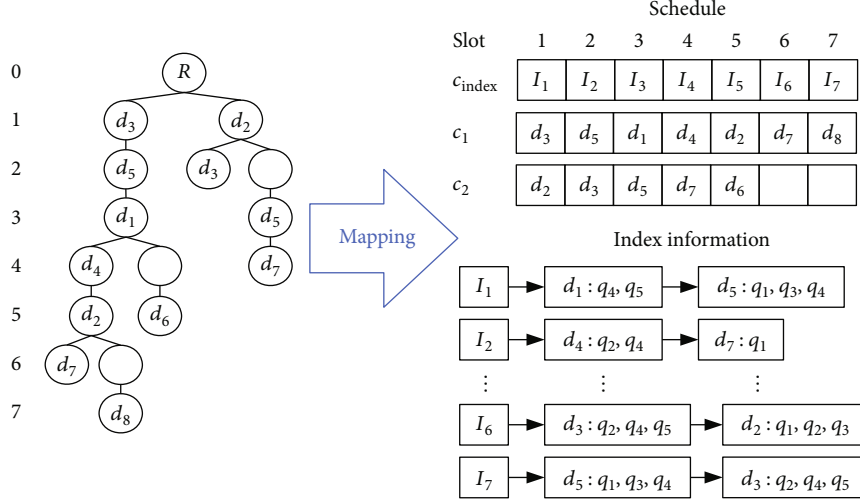


FIGURE 6: Mapping FP*-tree into the broadcasting channels with indexes.

5.3. Access Time. In wireless data dissemination environments, access time (or latency) is an important metric for validating the efficiency of scheduling. In FPBS, the system always first selects the request, whose size and average access frequency are maximum, and then schedules it in the backbone of FP*-tree. We then treat it as the base of schedule. That is, the access time for a request q_i can be formulated as Theorem 8.

Theorem 8. Suppose that \mathcal{F} is the maximal frequent item-set in the first-scheduled request, \hat{t} is the minimum cost for channel switching, and t_{wait} is the average waiting time from tuning into the channel to receiving the first required data item for a request, the access time for a request q_i can be expressed as

$$acc(q_i) = \begin{cases} t_{\text{wait}} + |q_i| + \sigma_1 \hat{t} + \sigma_2, & \text{if } (q_i \subseteq \mathcal{F}) \vee (q_i \cap \mathcal{F} = \emptyset), \\ t_{\text{wait}} + |q_i \cap \mathcal{F}| + |q_i \setminus \mathcal{F}| + \sigma_1 \hat{t} + \sigma_2, & \text{otherwise,} \end{cases} \quad (2)$$

where σ_1 is the frequency of channel switching and σ_2 is the frequency of occupied slot (empty node in the FP*-tree) skipping.

Proof. With the use of index channel in FPBS, the average waiting time can be reduced efficiently. If $q_i \subseteq \mathcal{F}$, it means that all the required data items for q_i can be obtained before the end of broadcasting all the data items in \mathcal{F} . In such a case, the access time for q_i will be $t_{\text{wait}} + |q_i| + \sigma_1 \hat{t} + \sigma_2$, where $|q_i| + \sigma_1 \hat{t} + \sigma_2 \leq |\mathcal{F}|$. If $q_i \cap \mathcal{F} = \emptyset$ (is equivalent to $|q_i \cap \mathcal{F}| = 0$), it means that q_i and \mathcal{F} are two disjoint sets. In this case, the data items requested by q_i only can be allo-

cated after the first-scheduled maximal frequent item-set, so the access time for q_i will be $t_{\text{wait}} + |\mathcal{F}| + |q_i| + \sigma_1 \hat{t} + \sigma_2$. However, the time $|\mathcal{F}|$ can be merged into the average waiting time t_{wait} until accessing the first data item requested by q_i . Otherwise, for the case of $|q_i \setminus \mathcal{F}| > 0$, q_i and \mathcal{F} are two partially overlapping. It means that some required data items for q_i will be scheduled after \mathcal{F} . Hence, the access time for q_i will be $t_{\text{wait}} + |q_i \cap \mathcal{F}| + |q_i \setminus \mathcal{F}| + \sigma_1 \hat{t} + \sigma_2$, where $|q_i \cap \mathcal{F}| + |q_i \setminus \mathcal{F}| + \sigma_1 \hat{t} + \sigma_2 \geq |F|$. \square

After discussing the general case of access time, we also discuss the worst case in following Theorem 9.

Theorem 9. Suppose all the notations are defined as above. The worst case of access time will be

$$acc_{\text{worst}} = t_{\text{wait}} + \left| \bigcup_{i=1}^n q_i \right|. \quad (3)$$

Proof. In general, the worse case is the scenario that a client access the channels from the first time slot to the last time slot. In other words, the worse access time of FPBS will be the height of the FP*-tree. According to the design of FPBS approach, the accelerating branches of FP*-tree is impossible to be longer than the backbone of FP*-tree. Hence, the height of the FP*-tree $\mathcal{H}_{\mathcal{F}}$ will be the height of the backbone, $|\bigcup_{i=1}^n q_i|$. In practice, each client tunes in channel at random time slot, so the access time in worst case acc_{worst} will be $t_{\text{wait}} + |\bigcup_{i=1}^n q_i|$. \square

```

1: Function FPBS_ScheduleMapping ( $\mathcal{T}, S, |C|$ )
   Input: an FP*-tree  $\mathcal{T}$ , a sorted query set  $S$ , and the number of channels  $|C|$ 
   Output: a scheduled channel set  $S_{\text{channel}}$  and an index channel  $I_{\text{channel}}$ 
2: let a list  $\text{list}_{\text{handling}} \leftarrow T.\text{root.children}$ ;
3: let a temporary list  $\text{list}_{\text{next}} \leftarrow \emptyset$ ;
4: create a data channel  $S_{\text{channel}}$  with  $|C|$  data broadcasting channels (or rows);
5: create an index channel  $I_{\text{channel}} \leftarrow \emptyset$ ;
6: int  $i$ ;
7: while  $\text{list}_{\text{handling}}$  is not empty do
8:    $i \leftarrow 1$ ; /* $i$  is used as a pointer to the current channel*/
9:   for each node  $N$  in  $\text{list}_{\text{handling}}$  do
10:    if  $N$  is an empty node then
11:      break;
12:    else if  $N.\text{parent}$  is an empty node then
13:      insert  $N$  into  $S_{\text{channel}}$  whose slot  $N.\text{slot}$  is not occupied;
14:    else
15:      insert  $N$  into the  $i$ th channel;
16:    end
17:    if  $N$  is not a leaf node then
18:      add  $N$ 's children into  $\text{list}_{\text{next}}$ ;
19:    end
20:     $i \leftarrow i + 1$ ;
21:  end
22:  copy every node of  $\text{list}_{\text{next}}$  to  $\text{list}_{\text{handling}}$ ;
23:   $\text{list}_{\text{next}}.\text{clear}()$ ;
24: end
25: for  $j \leftarrow 1$  to  $\mathcal{T}.\text{height}()$  do
26:   for  $i \leftarrow 1$  to  $|C|$  do
27:    Use  $S$  to check who requests the data item in the slot determined by (1) and channel  $C_i$  of  $S_{\text{channel}}$  and then update this
    information to  $I_{\text{channel}}[j]$ ;
28:   end
29: end
30: return  $I_{\text{channel}}, S_{\text{channel}}$ ;
31: end

```

ALGORITHM 4: The function used for the schedule mapping

In FPBS, each data item is not replicated in the FP*-tree's backbone and $|\bigcup_{i=1}^n q_i|$. In this work, we focus on minimizing the average access time and the proposed FPBS approach can effectively shorten the access time of each request using the accelerating branches. In (2), the terms $|q_i \cap \mathcal{F}|$ and $|q_i \setminus \mathcal{F}|$ are uncertain since the relation between request q_i and the maximal frequent item-set F is unpredictable. Hence, FPBS focus on minimizing the frequencies of channel switching or occupied slot (empty node in the FP*-tree) skipping, such as σ_1 and σ_2 in (2). This problem is solved by FP*-tree using the accelerating branches in our proposed approach. In other words, FPBS is proposed for effectively make the upper bound of access time be tighter. Thus, the worst case becomes a very rare occurrence.

6. Simulation Results

We validate and discuss the performance of FPBS in terms of average access time by running the experimental simulations in different scenarios. The unit of time is a time slot. All the simulations are written in C++ and executed on a Windows 7 server which is equipped with an Intel (R) Core (TM) i7-

3770 CPU @ 3.4 GHZ and 12G RAM. We use Quandl databases [31] to extract the U.S stock prices and then use the obtained stock dataset as the input of our simulation.

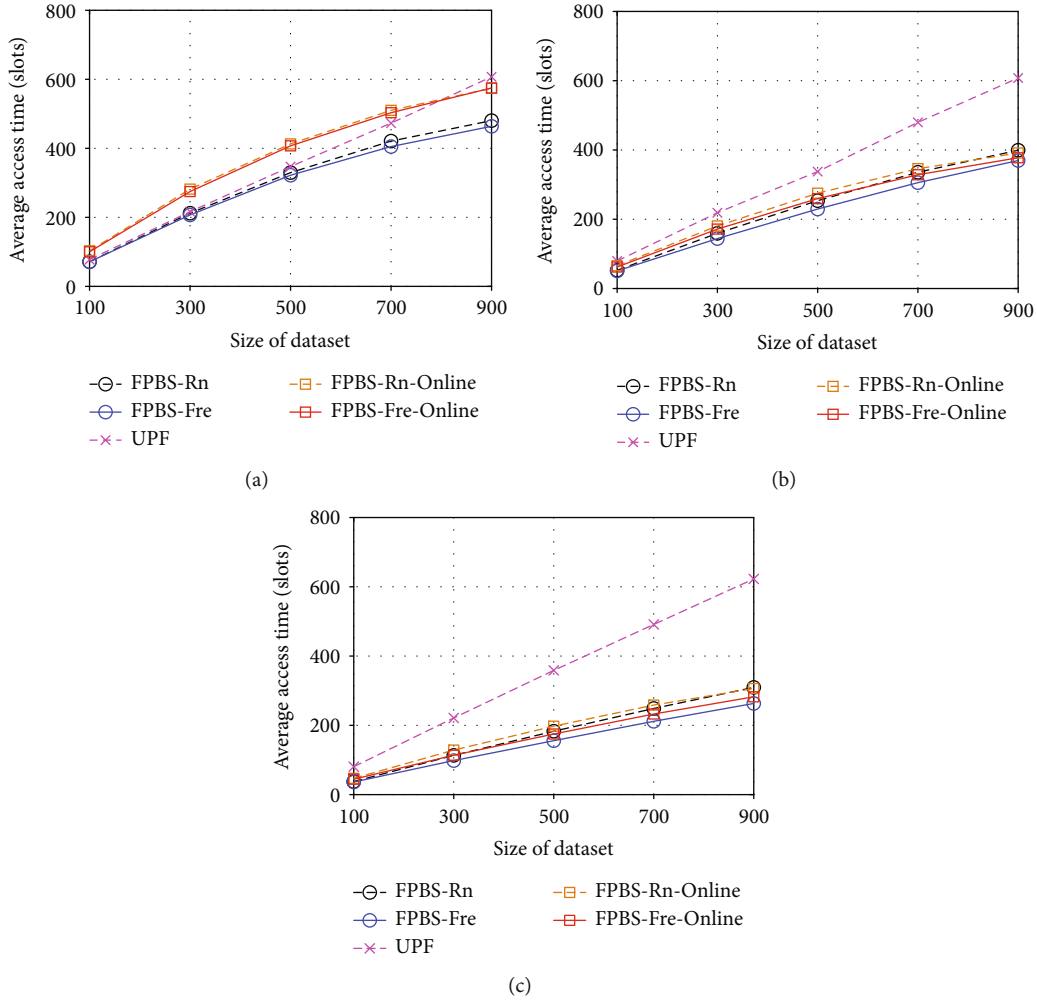
We assume that the maximum number of channels is 10 ($|C| = 2, 3, \dots, 10$) in the simulation. Therefore, we assume that one of the channels is the uplink for receiving the request and the remaining 10 channels are used as the downlink broadcasting channel. The detailed parameters of our simulations are shown in Table 5.

In the simulations, FPBS is conducted in online and offline modes. In the online mode, the system will use a buffer to keep the information of queries and request data items. When the buffer becomes full, the system will start to schedule data into the broadcasting channels. The scheduled data items will be removed from the buffer and new user demands are continuously coming in the buffer. It means that the FP*-tree and schedule may change during the simulation. Conversely, we assume that the system in the offline mode schedules the data after storing all the requested information in the buffer.

Note that there are two selecting strategies during scheduling process of FPBS, request-number-first and frequency-

TABLE 5: Simulation parameters.

Parameter	Default value	Range (type)
Size of dataset, $ D $	500	100, 300, 500, 700, 900
Number of users	5000	—
Maximum number of requested data items, q_{\max}	10	2, 4, 6, 8, 10
Number of downlink broadcast channels, $ C $	6	2, 3, ..., 10
Size of buffer	3000	500, 1000, ..., 4500

FIGURE 7: Effect of the different sizes of dataset with different number of channels: (a) $|C| = 3$, (b) $|C| = 6$, and (c) $|C| = 9$.

first. Request-number-first strategy is to select the query according to the length of its requested data items first and then selecting the query according to its average access frequency if multiple queries request same number of data items. Frequency-first strategy is to select the query according to its average access frequency first and then select the query according to the length of its requested data items if multiple queries have the same average access frequency. Hence, we discuss the above two strategies in online and offline modes, respectively.

To the best of our knowledge, none of existing works model the optimal performance of the multi-item request

scheduling simultaneously considering the channel switching and dependencies between different requests over multi-channel dissemination environments. Only [13] provides a heuristic algorithm, UPF, to discuss the similar problem. This is the reason that we choose UPF as the comparative baseline in the simulations.

6.1. Size of Dataset. In the first simulation, we discuss the performance of FPBS with different sizes of dataset in terms of average access time. Note that the size of dataset indicates the number of different data items stored in the dataset. Figure 7 shows the results in three different cases if the

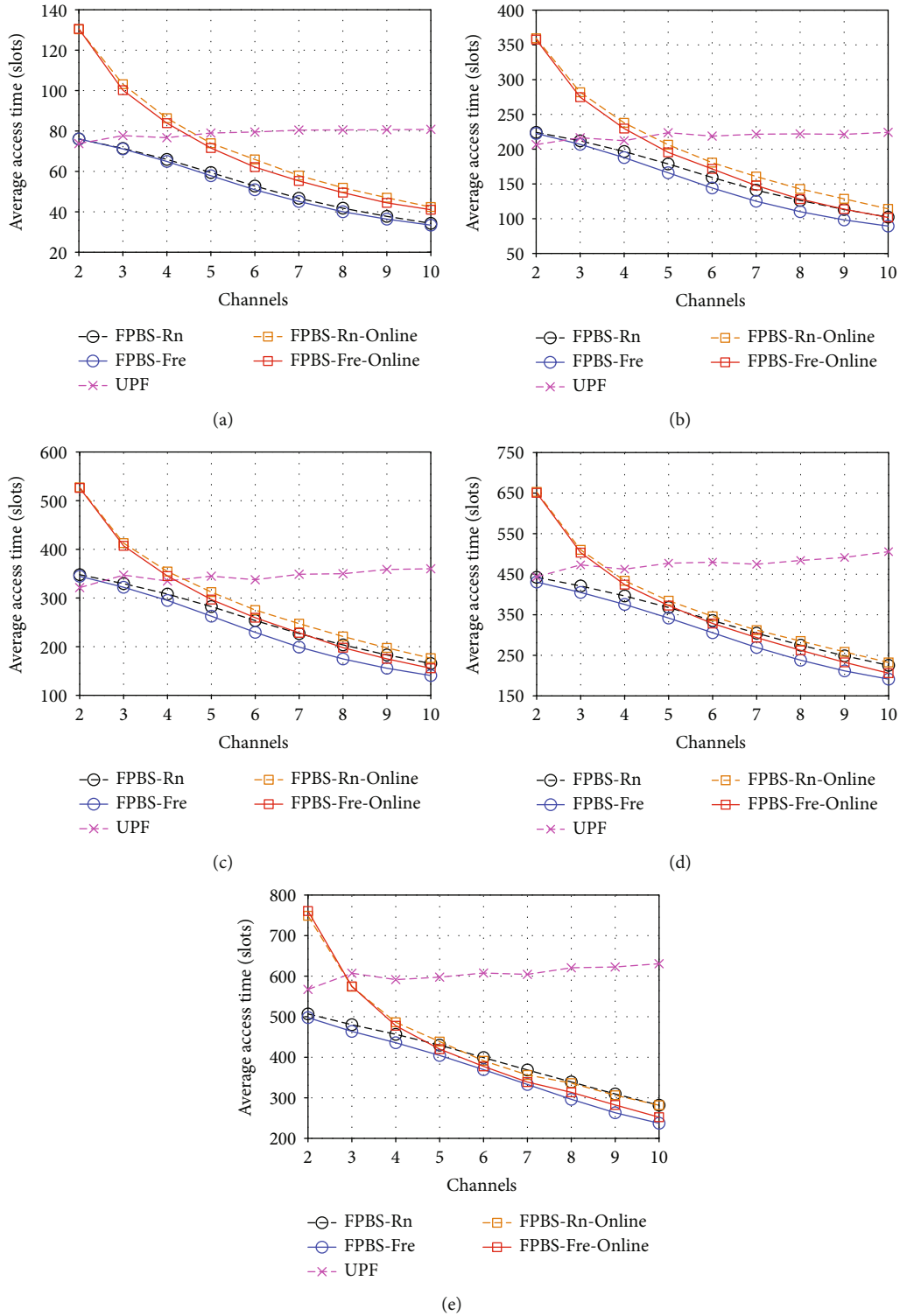


FIGURE 8: Effect of the different number of channels with different sizes of dataset: (a) $|D| = 100$, (b) $|D| = 300$, (c) $|D| = 500$, (d) $|D| = 700$, and (e) $|D| = 900$.

number of channels $|C| = 3$, $|C| = 6$, and $|C| = 9$, respectively. In the $|C| = 3$ channels environment, as shown in Figure 7(a), UPF can outperform the online FPBS approaches, FPBS-Fre-Online and FPBS-Rn-Online, if the size of dataset, $|D|$, is smaller than 800. The offline FPBS,

FPBS-Fre and FPBS-Rn, can always have a better performance than UPF does in all different sizes of dataset.

The results depicted from Figures 7(a)–7(c) show that UPF has similar performances in different number of channels environments and the trends of UPF’s average access

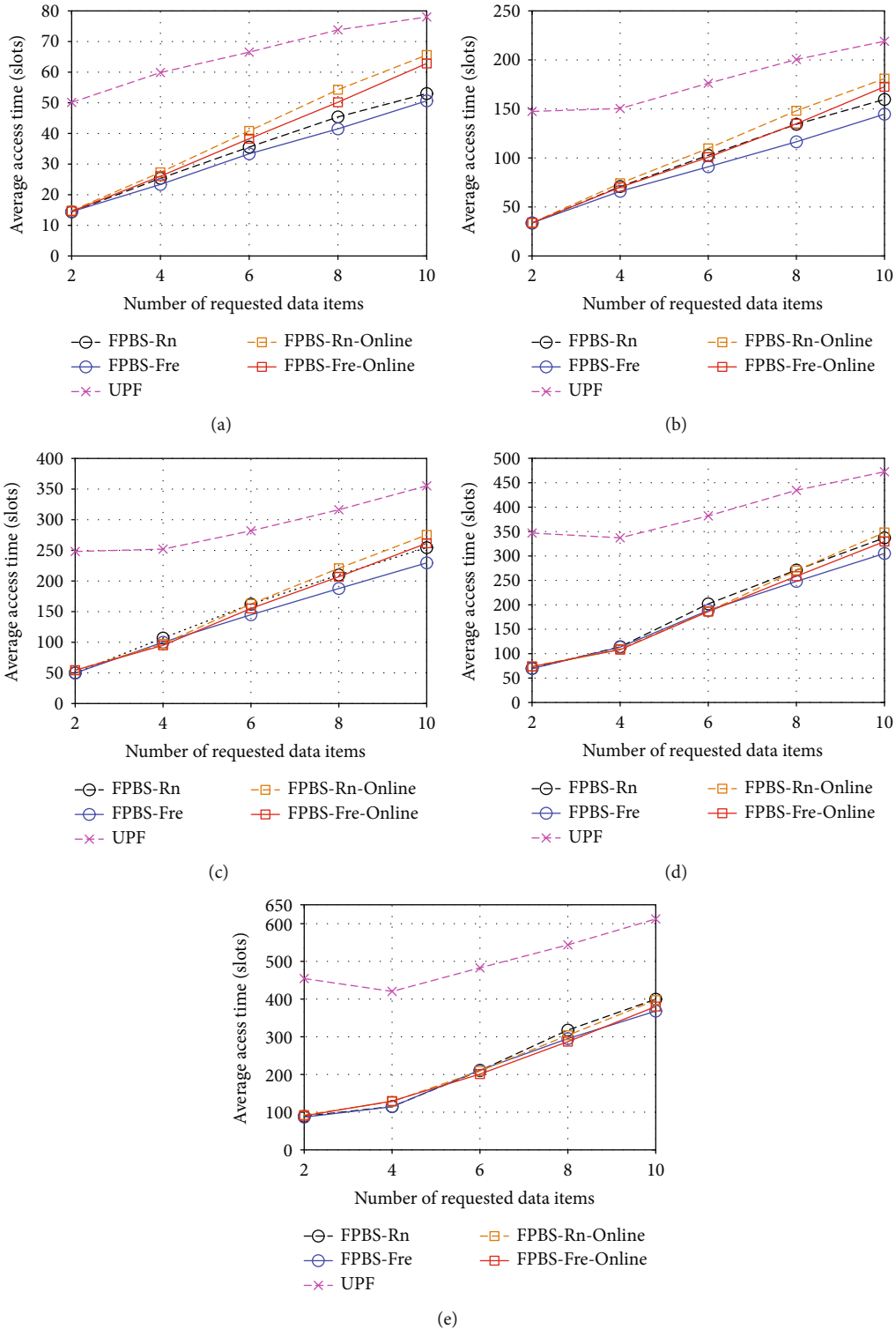


FIGURE 9: Effect of the different number of requested data items with different sizes of dataset: (a) $|D| = 100$, (b) $|D| = 300$, (c) $|D| = 500$, (d) $|D| = 700$, and (e) $|D| = 900$.

time are always linear increasing. According to the results in Figures 7(b) and 7(c), we can know that both of online and offline FPBS approaches can outperform UPF in different sizes of datasets when $|C| \geq 6$. Additionally, the frequency-first strategy, FPBS-Fre, always has the best performance in different scenarios.

6.2. Number of Channels. In this part, we discuss the performance of FPBS in different scenarios that the number of broadcasting channels is set from 2 to 20 and the results are shown in Figure 8. The results indicate the existing method, UPF, is not suitable to multiple channel ($C \geq 4$) broadcasting environments and UPF cannot dynamically

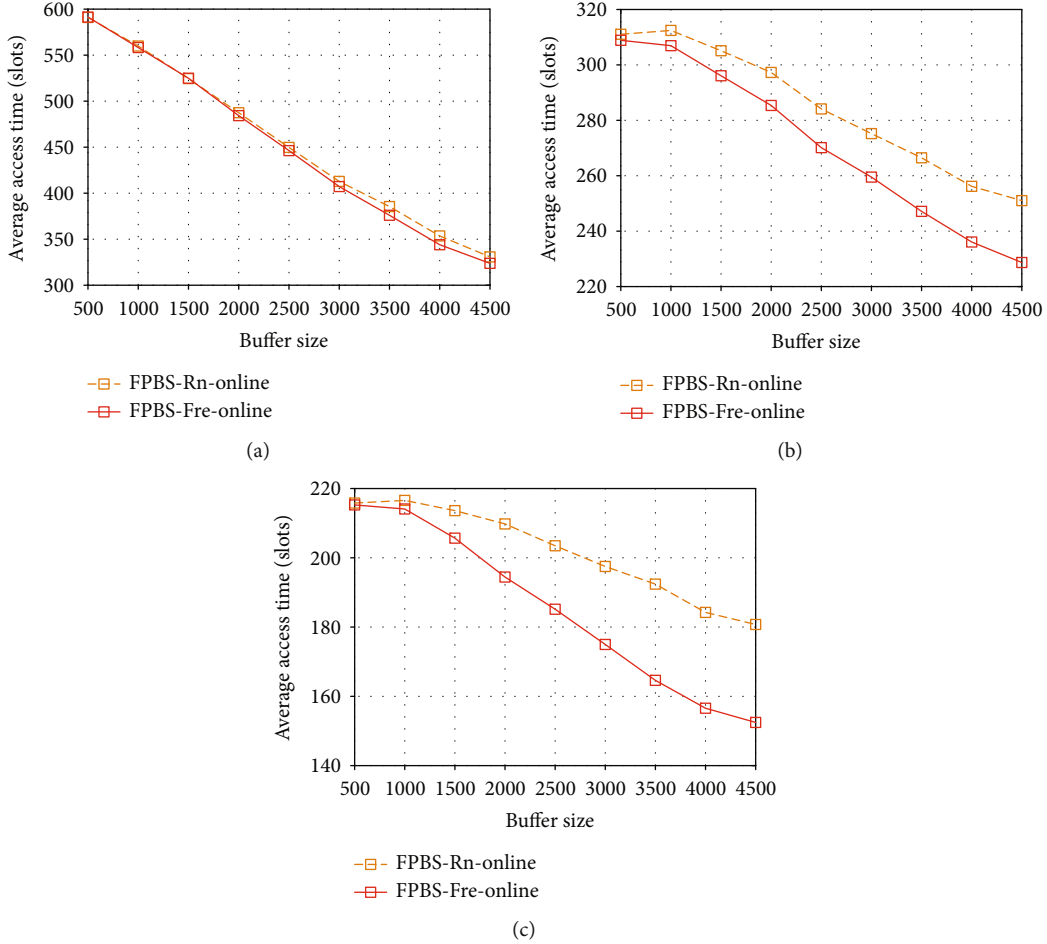


FIGURE 10: Effect of the different size of buffer with different number of channels: (a) $|C| = 3$, (b) $|C| = 6$, and (c) $|C| = 9$.

schedule data items with the consideration of each user's requests. That is to say, in comparison with the proposed approach, UPF can not utilize these channels if $C \geq 4$. Figures 8(a) and 8(b) show that UPF has a stable performance in the broadcasting environments with different number of channels when the size of dataset is small ($|D| \leq 300$). Conversely, the results from Figures 8(c)–8(e) show that the average access time of UPF is unstable and becomes a slightly increasing trend when the size of dataset becomes large ($|D| \geq 500$). The possible reason for this result is that UPF aims to minimize the request miss rate, not the average access time. There may be a trade-off between minimizing the request miss rate and the average access time.

Figures 8(a) and 8(b) shows the results of each approach in small dataset. FPBS in the offline mode, FPBS-Fre and FPBS-Rn, can have a better performance since the system consider all the requests while constructing the FP*-tree. According to the results in Figures 8(c)–8(e), the frequency-first strategies, FPBS-Fre and FPBS-Fre-Online, have better performances than the request-number-first strategies, FPBS-Rn and FPBS-Rn-Online, when the size of dataset becomes large ($|D| \geq 500$).

6.3. Number of Requested Data Items. If the number of requested data items becomes larger, the possibility of data

dependency between each query becomes higher. In this subsection, we consider the effect of the different number of requested data items on the average access time. As shown in Figure 9, one can observe that all the FPBS-based approaches can outperform UPF when the maximum number of requested data items q_{\max} is smaller than 11. When q_{\max} is 2, all the FPBS-based approaches have similar performances on the average access time. As the value of q_{\max} increases, the average access time in all the FPBS-based approaches also increases linearly.

According to the result in Figure 9, we can know that the frequency-first strategies are better than the request-number-first strategies since the performances of FPBS-Fre and FPBS-Fre-Online are more smoothly increasing than the performances of FPBS-Rn and FPBS-Rn-Online. In addition, FPBS-Fre can has the best performance and its trend is almost parallel to the trend of UPS's performance.

6.4. Buffer Size. In the last simulation, we discuss the effect of the different size of buffer on the average access time for comparing two proposed online approaches, FPBS-Rn-Online and FPBS-Fre-Online. We also consider the trend of performance in some scenarios that the number of channel is, respectively, set to 3, 6, and 9.

The result in Figure 10 indicates that both FPBS-Rn-Online and FPBS-Fre-Online can have shorter average access time as the size of buffer increases. In an environment providing small number ($C = 3$) of channels, as shown as Figure 10(a), FPBS-Fre-Online can have a slightly better performance than FPBS-Rn-Online does when the buffer can store more than 2500 data items. The results in Figures 10(b) and 10(c) show that FPBS-Fre-Online is much better than FPBS-Rn-Online with different size of buffer when the number of channels increases ($C \geq 6$).

6.5. Open Issues. In this subsection, we summarize some remaining issues (or potential challenges) in on-demand multi-channel data dissemination systems as follows:

- (i) Hardware constraint: although the minimum cost \hat{t} for channel switching is normalized as one time slot in FPBS, it is difficult to implement a broadcasting system that meets this condition due to hardware limitations
- (ii) Cross-layer system design: in this paper, we design a server-side data scheduling for serving the multi-item requests. For wireless networks, the time-varying and uncertain nature of wireless channels can be considered in the scheduling. Thus, the server needs a new cross-layer system design to simultaneously access the request information in the application layer and channel information in the physical layer and then schedule data items more efficiently

7. Conclusion

In this paper, we investigate and formulate an emerging problem, DBCA, in multichannel wireless data dissemination environments. We also prove that the DBCA problem is \mathcal{NP} -complete. Then, we present a heuristic scheduling approach, FPBS, to avoid data conflicts on multiple broadcasting channels. In FPBS, we use frequent patterns of requested data items to build a FP*-tree for extracting the correlation between each received request. Thus, data conflicts can be avoided. During the construction of FP*-tree's accelerating branch, adding empty nodes at appropriate positions makes the user client have sufficient time to switch the channel for obtaining the required data. We not only analyze that FPBS can be done in polynomial time but also present the upperbound of access time of a request which is related to size of dataset. According to the simulation results, FPBS is much better than the existing work, UPF, in most of cases.

Data Availability

The stock dataset used to support this study is available online and is cited as a reference [31] in relevant places in the text. The program 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 research was supported by the Ministry of Science and Technology, Taiwan under Grant Nos. MOST 107-2221-E-027-099-MY2, MOST 109-2221-E-027-095-MY3, and MOST 110-2222-E-035-004-MY2.

References

- [1] A. Ghorbel, M. Kobayashi, and S. Yang, "Content delivery in erasure broadcast channels with cache and feedback," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 6407–6422, 2016.
- [2] R. Martinez Alonso, D. Plets, E. Fontes Pupo et al., "IoT-based management platform for real-time spectrum and energy optimization of broadcasting networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7287641, 14 pages, 2018.
- [3] S.-J. Ra, M.-S. Baek, J.-H. Song, D.-J. Choi, J.-Y. Jung, and C.-S. Kim, "Implementation and field trials of OFDM-based digital video broadcasting system in commercial broadcasting network for multichannel UHD service," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 1649413, 9 pages, 2019.
- [4] S. Tong and C. Yang, "Improvement of data sharing efficacy of p2p streaming mobile networks for news-broadcast-on-demand services," in *Seventh International Conference on Innovative Computing Technology (INTECH)*, Porto, Portugal, 2017.
- [5] C.-L. Hu and M.-S. Chen, "Adaptive multichannel data dissemination: support of dynamic traffic awareness and push-pull time balance," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 2, pp. 673–686, 2005.
- [6] R. Sotelo, J. Joskowicz, and N. Rondan, "An integrated broadcast-broadband system that merges isdb-t with hbbtv 2.0," *IEEE Transactions on Broadcasting*, vol. 64, no. 3, pp. 709–720, 2018.
- [7] X. Gao, A. Song, L. Hao, J. Zou, G. Chen, and S. Tang, "Towards efficient multi-channel data broadcast for multimedia streams," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 10, pp. 2370–2383, 2019.
- [8] I. Viswanathan, T. Imielinski, and S. Viswanathan, "Adaptive wireless information systems," in *SIGDBS Conference*, Tokyo, 1994.
- [9] S. B. Zdonik, M. J. Franklin, R. Alonso, and S. Acharya, "Are 'disks in the air' just pie in the sky?," in *1994 First Workshop on Mobile Computing Systems and Applications*, Santa Cruz, CA, 1994.
- [10] S. Acharya, R. Alonso, M. Franklin, and S. Zdonik, "Broadcast disks: data management for asymmetric communication environments," in *ACM SIGMOD Conference*, San Jose, CA, May 1995.
- [11] D. Aksoy and M. Franklin, "R×W: a scheduling approach for large-scale on-demand data broadcast," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 846–860, 1999.
- [12] C.-M. Liu and T.-C. Su, "Broadcasting on-demand data with time constraints using multiple channels in wireless broadcast

- environments,” *Information Sciences*, vol. 242, pp. 76–91, 2013.
- [13] P. He, H. Shen, and H. Tian, “On-demand data broadcast with deadlines for avoiding conflicts in wireless networks,” *Journal of Systems and Software*, vol. 103, pp. 118–127, 2015.
- [14] Z. Lu, W. Wu, W. W. Li, and M. Pan, “Efficient scheduling algorithms for on-demand wireless data broadcast,” in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, San Francisco, CA, USA, Apr. 2016.
- [15] Jianliang Xu, Xueyan Tang, and Wang-Chien Lee, “Time-critical on-demand data broadcast: algorithms, analysis, and performance evaluation,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 17, no. 1, pp. 3–14, 2006.
- [16] X. Wu and V. C. Lee, “Wireless real-time on-demand data broadcast scheduling with dual deadlines,” *Journal of Parallel and Distributed Computing*, vol. 65, no. 6, pp. 714–728, 2005.
- [17] Wai Gen Yee, S. B. Navathe, E. Omiecinski, and C. Jermaine, “Efficient data allocation over multiple channels at broadcast servers,” *IEEE Transactions on Computers*, vol. 51, no. 10, pp. 1231–1236, 2002.
- [18] B. Zheng, X. Wu, X. Jin, and D. L. Lee, “TOSA: a near-optimal scheduling algorithm for multi-channel data broadcast,” in *The 6th International Conference on Mobile Data Management*, Ayia Napa, Cyprus, 2005.
- [19] S.-Y. Yi, S. Nam, and S. Jung, “Effective generation of data broadcast schedules with different allocation numbers for multiple wireless channels,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 668–677, 2008.
- [20] Z. Lu, Y. Shi, W. Wu, and B. Fu, “Efficient data retrieval scheduling for multi-channel wireless data broadcast,” in *2012 Proceedings IEEE INFOCOM*, pp. 891–899, Orlando, FL, USA, Mar. 2012.
- [21] Z. Lu, W. Wu, and B. Fu, “Optimal data retrieval scheduling in the multichannel wireless broadcast environments,” *IEEE Transactions on Computers*, vol. 62, no. 12, pp. 2427–2439, 2013.
- [22] Z. Lu, Y. Shi, W. Wu, and B. Fu, “Data retrieval scheduling for multi-item requests in multi-channel wireless broadcast environments,” *IEEE Transactions on Mobile Computing*, vol. 13, no. 4, pp. 752–765, 2014.
- [23] J. Lv, V. C. Lee, M. Li, and E. Chen, “Profit-based scheduling and channel allocation for multi-item requests in real-time on-demand data broadcast systems,” *Data & Knowledge Engineering*, vol. 73, pp. 23–42, 2012.
- [24] K.-F. Lin and C.-M. Liu, “Broadcasting dependent data with minimized access latency in a multi-channel environment,” in *The 2006 International Conference on Wireless Communications and Mobile Computing*, Vancouver, British Columbia, Canada, Jul. 2006.
- [25] Z. Qiu, W. Hu, and B. Du, “RPPM: a request pre-processing method for real-time on-demand data broadcast scheduling,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 11, pp. 2619–2631, 2018.
- [26] K. Liu, V. C. S. Lee, and K. R. P. H. Leung, “Data scheduling for multi-item requests in multi-channel on-demand broadcast environments,” in *The Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, Vancouver, Canada, Jun. 2008.
- [27] R. D. J. Van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*, Artech House, Inc., USA, 1st ed. edition, 2000.
- [28] J. Juran, A. R. Hurson, N. Vijaykrishnan, and S. Kim, “Data organization and retrieval on parallel air channels: performance and energy issues,” *Wireless Networks*, vol. 10, no. 2, pp. 183–195, 2004.
- [29] A. R. Hurson, A. M. Muñoz-Avila, N. Orchowski, B. Shirazi, and Y. Jiao, “Power-aware data retrieval protocols for indexed broadcast parallel channels,” *Pervasive and Mobile Computing*, vol. 2, no. 1, pp. 85–107, 2006.
- [30] T. Gonzalez, “Unit execution time shop problems,” *Mathematics of Operations Research*, vol. 7, no. 1, pp. 57–66, 1982.
- [31] Quandl, “(2016) WIKI various end-of-day data,” 2020, <https://data.nasdaq.com/data/WIKI>.

Research Article

Stock Trading System Based on Machine Learning and Kelly Criterion in Internet of Things

Lili Chen ¹, Lingyun Sun ¹, Chien-Ming Chen ¹, Mu-En Wu ²,
and Jimmy Ming-Tai Wu ¹

¹College of Computer Science and Engineering, Shandong University of Science and Technology, China

²Department of Information and Finance Management, National Taipei University of Technology, Taiwan

Correspondence should be addressed to Jimmy Ming-Tai Wu; wmt@wmt35.idv.tw

Received 6 September 2021; Revised 8 October 2021; Accepted 10 November 2021; Published 3 December 2021

Academic Editor: Chao-Yang Lee

Copyright © 2021 Lili 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 evolution of the Internet of Things (IoT) has promoted the prevalence of the financial industry as a variety of stock prediction models have been able to accurately predict various IoT-based financial services. In practice, it is crucial to obtain relatively accurate stock trading signals. Considering various factors, finding profitable stock trading signals is very attractive to investors, but it is also not easy. In the past, researchers have been devoted to the study of trading signals. A genetic algorithm (GA) is often used to find the optimal solution. In this study, a long short-term (LSTM) memory neural network is used to study stock price fluctuations, and then, genetic algorithms are used to obtain appropriate trading signals. A genetic algorithm is a search algorithm that solves optimization. In this paper, the optimal threshold is found to determine the trading signal. In addition to trading signals, a suitable trading strategy is also crucial. In addition, this research uses the Kelly criterion for fund management; that is, the Kelly criterion is used to calculate the optimal investment score. Effective capital management can not only help investors increase their returns but also help investors reduce their losses.

1. Introduction

In recent years, the development of IoT is facilitating the flourishing of various industries, especially in the economic market. Investment, insurance, and forecasting are among the services related to the Internet of Things in the financial aspect. The financial market is very complicated, especially the various transactions in the stock market. If you want to get high returns through investment, the stock market is a very good choice. But often most investors are affected by emotions and are prone to make irrational judgments, leading to capital losses or low returns. Advances in financial theory and computer technology have made quantitative trading [1, 2] possible. Quantitative trading can be realized automatically through technologies such as computers. Different from general trading methods, the behavior of quantitative trading is to use computer technology to mine relevant information from the historical data of the stock market to increase the profit of the transaction [3, 4]. Therefore, to develop a quantitative

trading strategy that can increase returns is what researchers hope. Whether a stock transaction is profitable has many factors, such as political factors and economic factors [5]. Therefore, stock trading will not only bring higher additional income to people but also bring certain risks. For investors in the stock market, the main purpose is to obtain higher returns, that is, to find a stock trading system that can avoid certain risks and obtain maximum profits.

Researchers mainly focus on two aspects when studying stock trading: stock trading strategy and fund management [6, 7]. A good trading strategy is of great help to investors in the stock market. The Kelly criterion is usually used to determine the optimal investment ratio. In previous work, researchers and investors usually use Kelly's criterion to manage funds [8, 9]. Wu and Chung [10] devised a strategy to reduce losses and increase profits by holding different options. In this strategy, the Kelly criterion is used to calculate the optimal investment ratio to determine the appropriate option portfolio.

Artificial intelligence technology is very popular in the financial field. In recent years, researchers have proposed many prediction methods for the stock market [11, 12]. Many researchers have applied machine learning to financial transactions. Cohen et al. [13] used neural networks to study the unknown rules in price changes in the stock market. Chiang et al. [14] used artificial neural networks to study the fund's year-end net asset value. Experimental results show that the predictive performance of neural networks is better than traditional econometric techniques. Yu and Yan [15] proposed a framework for predicting stock prices based on deep neural networks (DNN). This framework combines phase-space reconstruction (PSR), and its prediction performance is better than some other prediction frameworks. Deep learning is also used in other areas [16, 17]. The stock sequence array convolutional neural network (SSACNN) [18] is proposed on the basis of the convolutional neural network, which combines advanced indicators such as futures and options. The accuracy of this framework for stock price prediction is higher than that of several other frameworks. The price of the financial market is a time series, so some researchers use long short-term memory neural networks in financial transactions [19]. In an independent repeated gambling game with a positive pure income, the Kelly formula assists investors to maximize the growth rate of the principal. In this study, a stock trading system is designed to reduce trading risk by using the Kelly criterion for money management when trading.

The paper mainly includes the following contents. The second part talks about related work. The third part mainly talks about the methods used in this paper, including long- and short-term memory neural networks based on leading indicators, genetic algorithms, trading strategies, and Kelly's criterion. The fourth part elaborates on the stock trading system in detail. The fifth section is the result of the experiment with the proposed stock trading system. The last section is a summary of the whole paper.

2. Related Work

In recent years, machine learning technology has continued to advance, and researchers have applied machine learning in various fields. In the financial field, machine learning highlights its powerful learning capabilities. Lu et al. [20] proposed a framework based on the convolutional neural network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) to predict stock prices within 1,000 trading days. Compared with the other seven methods, the proposed framework has better performance and is more suitable for investors to make correct decisions in the stock market. Nelson et al. [21] used a long short-term memory neural network to predict future stock price fluctuations based on technical indicators and historical prices. Compared with other machine learning methods, this framework has good performance in predicting future stock price fluctuations. Bao et al. [22] used LSTM to predict stock prices based on wavelet transform (WT) and stacked autoencoders (SAEs). Through a series of experiments, the prediction accuracy of the newly proposed framework is higher than that of other

similar models. Budiharto [23] used data from the Indonesian Stock Exchange to predict the price of stocks based on R language and LSTM. Experimental results show that the framework has better prediction performance on short-term data. Based on historical prices and leading indicators in the two stock markets of the United States and Taiwan, Wu et al. [18] proposed a framework composed of CNN and LSTM to predict stock prices. Experimental results show that the framework has better prediction performance.

In 1975, Holland proposed the genetic algorithm for the first time and carried out the practical application of the genetic algorithm [24]. The genetic algorithm is used to find the optimal solution in a limited time [25]. Fu et al. [26] used genetic algorithms to find optimal solutions in portfolio management, including parameter optimization and combination weights. Different algorithms are used for evaluation, and the experimental results reflect the better optimization ability of the genetic algorithm. Chen and Zhou [27] combined LSTM and GA to get a stock prediction system. On the basis of the CSI 300 and China Construction Bank data sets, GA is used to obtain the most effective factors. The experimental results show that the proposed framework is superior to all benchmark frameworks. Chen and Shih [28] used GA to set the best parameters and forecast stock price fluctuations on the basis of technical indicators and Chinese news. Based on real data, a number of experiments have been carried out, and the experiments show that the method has relatively good performance. Therefore, a genetic algorithm is added to our research, and GA is used to determine the optimal threshold of trading signals.

3. Methodology

In this section, the methods used in this research will be discussed. First of all, the first section mainly talks about the framework used in this research, that is, long short-term memory based on leading indicators (LSTMLI), which uses LSTMLI to classify the rise and fall of stock prices. The second section mainly talks about how to use a genetic algorithm to find the best threshold to determine the trading signals in the stock market. The third section uses the Kelly criterion to manage funds.

3.1. Long Short-Term Memory Based on Leading Indicators. In this study, a framework of long short-term memory based on leading indicators is used. The long short-term memory neural network is improved on the basis of the recurrent neural network, which solves many unsolvable problems of the recurrent neural network [29]. The recurrent neural network is a neural network with memory function, which adds a feedback mechanism of timing. This determines that the input of the recurrent neural network has not only the input at the current moment but also the input at the previous moment. For example, when watching a movie, if you want to know the next story, you need to make predictions based on the previous storyline and what is happening now. In this forecast, the issue of time series is involved. Based on such problems, recurrent neural networks are proposed. However, recurrent neural networks still face two problems:

gradient explosion and gradient disappearance [30]. Subsequently, people continued to improve the cyclic neural network and obtained the long short-term memory neural network. The internal structure of LSTM is shown in Figure 1. Three gates are added to this network, namely, input gate, forget gate, and output gate, through which information is controlled.

In this paper, the inputs of LSTM are historical prices and leading indicators. Leading indicators can indicate turning points in the process of economic development and are one of the important indicators for predicting future economic development trends. Therefore, in this study, in addition to the historical stock price, the two leading indicators of futures and options are also used, that is, a total of three data sets. LSTM is used to classify the changes in stock prices. The changes in stock prices are divided into three categories: rising, falling, and unchanged. Therefore, the output of LSTM is also three types, namely, (1,0,0), (0,0,1), and (0,1,0). In other words, when the stock price change after preprocessing is greater than +0.01, the output result of LSTM should be (1,0,0); when the stock price change after preprocessing is less than -0.01, the output result of LSTM should be (0,0,1); when the stock price changes between +0.01 and -0.01, the output result of LSTM should be (0,1,0). According to these changes, a long short-term memory framework based on leading indicators is proposed, referred to as LSTMLI. The structure of LSTMLI is shown in Figure 2.

3.2. Genetic Algorithm. Buying and selling signals are a core issue of stock trading. Finding the right buying and selling signals can help investors obtain high returns. However, appropriate buying and selling signals are difficult to capture. Genetic algorithms are generally used to solve optimization problems. In this research, a genetic algorithm is used to find the threshold of buying signal and selling signal when trading stocks. Therefore, the chromosome is set up in two parts, as shown in Figure 3.

3.2.1. Initial Population. Initialize the chromosome first. Because the sum of the three nodes obtained by the LSTMLI framework is approximately 1, two floating-point numbers between 0 and 1 are randomly generated as the threshold of the buying and selling signal, as shown in Figure 4. Then, use the thresholds of the two buying and selling signals to determine a trading strategy. Next, introduce the trading strategy of this research in detail. Assume that the result obtained by LSTMLI is shown in Figure 5, where “1, 2, 3, ...” on the left represents the day. The first row of Figure 2 is the result of the first day obtained from the prediction. The value of its first node is 0.85551. It can be seen that $0.85551 > 0.76848$, so a buy signal appeared on the first day. A buying operation must be followed by a selling operation, and the trading strategy set in this study does not allow simultaneous buying and selling transactions on the same day. Therefore, sell signals should be found from the next day. To find a buy signal, you need to observe the value of the third node. The value of the third node on the second day is 0.79545, which is less than 0.82843, so there is no sell

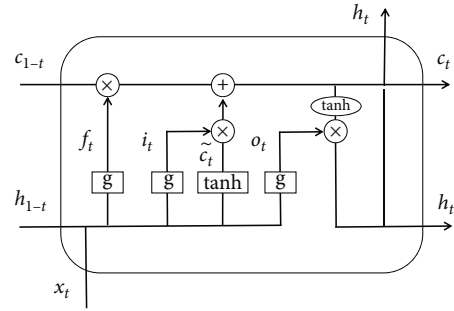


FIGURE 1: Schematic diagram of the internal structure of LSTM.

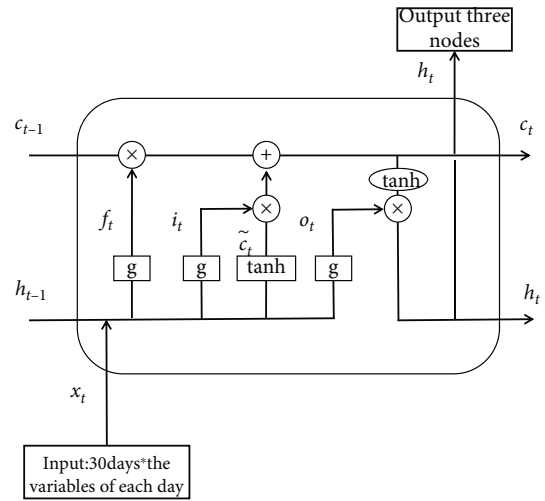


FIGURE 2: A long short-term memory framework based on leading indicators.

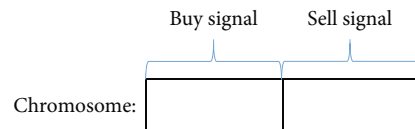


FIGURE 3: Representation of chromosome.

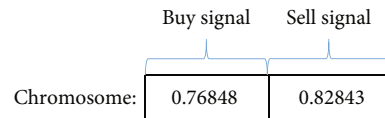


FIGURE 4: An example of the threshold of randomly generated buying and selling signals.

signal on the second day. The value of the third node on the third day is 0.83419, and it can be known that $0.83419 > 0.82843$, so a sell signal appeared on the third day, according to this trading strategy to buy and sell stocks.

3.2.2. Fitness Evaluation. The fitness function is shown in Formulas (1), (2), and (3). Use fitness function to evaluate chromosomes. The result of this fitness function is the final funds, and the more funds, the better the chromosomes. B

$Close_i$ represents the closing price at the time of the buy signal, and $SClose_i$ represents the closing price at the time of the sell signal:

$$share_i = \frac{\text{Total amount}_i}{B\text{Close}_i}, \quad (1)$$

$$\text{balance}_i = \text{Total amount}_i - \text{share}_i * B\text{Close}_i, \quad (2)$$

$$\text{Final amount}_i = \text{balance}_i + \text{share}_i * S\text{Close}_i. \quad (3)$$

3.2.3. Crossover. The process of crossing two chromosomes is shown in Figure 6. Among them, one is the paternal chromosome, and the other is the maternal chromosome. A new chromosome can be obtained by crossing and splicing the two chromosomes. The crossover rate is set to 0.8.

3.2.4. Mutation. The process of mutation of the two chromosomes is shown in Figure 7. Randomly select a chromosome for mutation, and set the mutation rate to 0.2.

3.2.5. Iteration. After reaching a certain number of iterations, the iteration stops.

3.3. Trading Strategy. A trading strategy is the core of a stock trading system, which is related to whether the stock trading can obtain higher returns. The trading strategy includes two key points: the time to buy stocks and the time to sell stocks. In this research, a genetic algorithm is used to find the threshold of buying and selling signals to determine the time point of stock trading. The process of finding buying and selling signals has been introduced in detail in Section 3.2. Figure 8 is an example of looking for buying and selling signals. In Figure 8, there are multiple sets of buying and selling signals. When a buy signal appears, the stock is bought; when a sell signal appears, the stock is sold. Buy and sell signals appear in pairs. When a buy signal appears, the next signal that appears will only be a sell signal.

3.4. Kelly Criterion. After formulating a trading strategy, you can trade stocks. However, how much money should be invested when buying each stock is still a question worth studying. The Kelly criterion is a powerful tool for optimizing asset investment scores. It enables researchers or investors to maximize returns under repeated games. For example, there is now a binary outcome game; that is, there are only two situations of winning and losing. Assuming that the gambler puts \$1 as a bet and the winning rate is p , if the gambler wins, it will get all the principal plus b_1 times the principal, that is, $1 + b_1 * 1$, which means that the net profit is b_1 . If the gambler loses, he will be deducted b_2 times the principal, that is, $1 - b_2 * 1$. Assuming that the gambler's initial capital is A_0 , f times the total capital is invested as a bet in each game, where $0\% \leq f \leq 100\%$, and the total capital to step t is A_t . Therefore, if the gambler wins the t th game, the total funds for the t th game are $A_t = A_{t-1} + b_1 * f * A_{t-1}$. If the gambler loses in the t th game, the total funds for the t th game are $A_t = A_{t-1} - b_2 * f * A_{t-1}$. After the T round of the game, assuming that the gambler wins W times

	Buy signal	Sell signal
1	0.85551	0.02341
2	0.14344	0.79545
3	0.17535	0.83419
.....
.....	0.71714	0.15575

FIGURE 5: An example of the values of three nodes output by SSACNN.

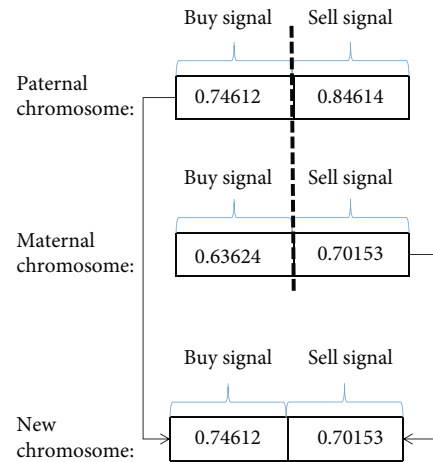


FIGURE 6: An example of when a buy or sell signal appears.

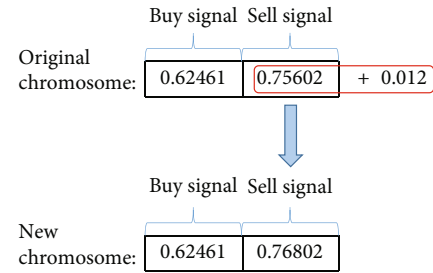


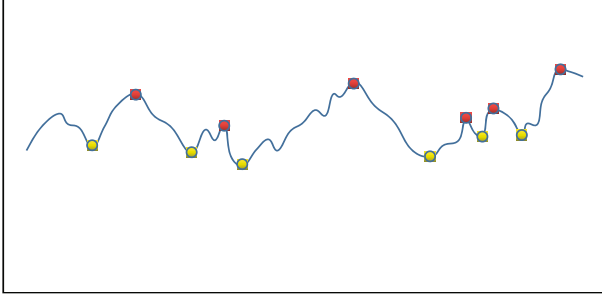
FIGURE 7: An example of when a buy or sell signal appears.

and loses L times, that is, $W + L = T$, the total funds for the t th time are as follows:

$$A_T = A_0(1 + b_1f)^W(1 - b_2f)^L; \quad (4)$$

divide both sides of Equation (4) by A_0 , then take the logarithm at the same time and divide by T to get

$$\frac{1}{T} \log \left(\frac{A_T}{A_0} \right) = \frac{W}{T} \log (1 + b_1f) + \frac{L}{T} \log (1 - b_2f); \quad (5)$$



- Buy signal
- Sell signal

FIGURE 8: An example of when a buy or sell signal appears.

assuming that the game is played countless times, that is, T tends to infinity, and the winning rate is p , then Equation (6) can be obtained:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \left(\frac{A_T}{A_0} \right) = p \log(1 + b_1 f) + (1 - p) \log(1 - b_2 f) \quad (6)$$

derivatives on both sides of Equation (6) can get the optimal investment score f , as shown in

$$f = \frac{pb_1 - (1-p)b_2}{b_1 b_2}. \quad (7)$$

Therefore, if you want to know the investment score, you first require the winning rate p , the reward b_1 , and the penalty b_2 . Take an example to illustrate the process of solving f . Assuming that 0.5 times of the total assets are invested in each transaction, according to the trading strategy, the profit and loss of each transaction can be obtained. For example, a total of ten transactions have been made, and the profit and loss of each transaction are +5, -2, -3, +7, -4, +4, +8, +6, -3, and +6. In order to find rewards and punishments, it is necessary to separate profit and loss. Then, the profit is +5, +7, +4, +8, +6, and +6; the loss is -2, -3, -4, and -3. Therefore, the winning rate is $6/10 = 0.6$; the reward is $(5 + 7 + 4 + 8 + 6 + 6)/6 = 6$, and the penalty is $-(2 + 3 + 4 + 3)/4 = -3$. Substituting the data into the Kelly formula can get f .

4. A Stock Trading System Based on Trading Signals and Fund Management

This section mainly includes two parts: the input of the framework and the structure of the stock trading system.

4.1. The Input of the LSTMLI Framework. The data sets used in this experiment are historical prices and leading indicators. Leading indicators can indicate turning points in the process of economic development and are one of the important indicators for predicting future economic development trends. Therefore, in this study, in addition to the historical

TABLE 1: Example of input matrix.

	1	2	3	...	30
Open	266	262	262	...	230
High	266	264	262.5	...	234.5
Low	260	261	257	...	230
Close	262.5	263	257.5	...	231.5
Change	-2.5	0.5	-0.5	...	0.5

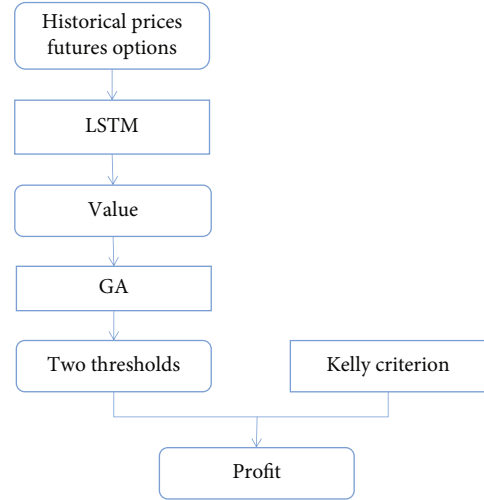


FIGURE 9: The flow chart of the stock trading system.

stock price, the two leading indicators of futures and options are also used, that is, a total of three data sets. In the experiment, we used data from China's Taiwan stock market and data from the US stock market. However, there are only historical prices and options in the data of the US stock market. Therefore, when researching US stocks, only these two data sets are used. When researching Taiwan stocks, three data sets are used.

In the experiment, the 30-day data was integrated into a matrix as the input of the framework. For example, use historical data as input to the framework, as shown in Table 1. Among them, "open, high, ..." are the attributes of the data, and "1, 2, ..." means the day.

4.2. Stock Trading System. This paper is mainly to study the issue of stock returns. First, use the framework LSTMLI to classify stock price fluctuations. The classification method is shown in

$$\text{label} = \begin{cases} \text{class} + 1, & \text{change} > 0.01, \\ \text{class} 0, & \text{others}, \\ \text{class} - 1, & \text{change} < -0.01. \end{cases} \quad (8)$$

Among them, change represents the fluctuation value of the stock price. Then, use GA to study the output value of LSTMLI to obtain the optimal threshold for buying and

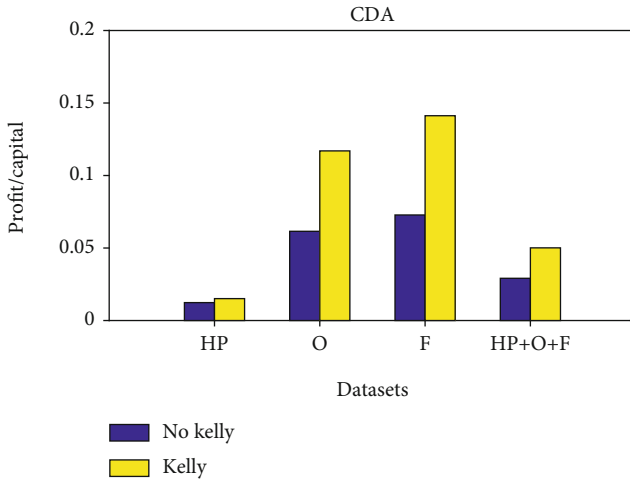


FIGURE 10: The ratio of CDA stock's profit to capital.

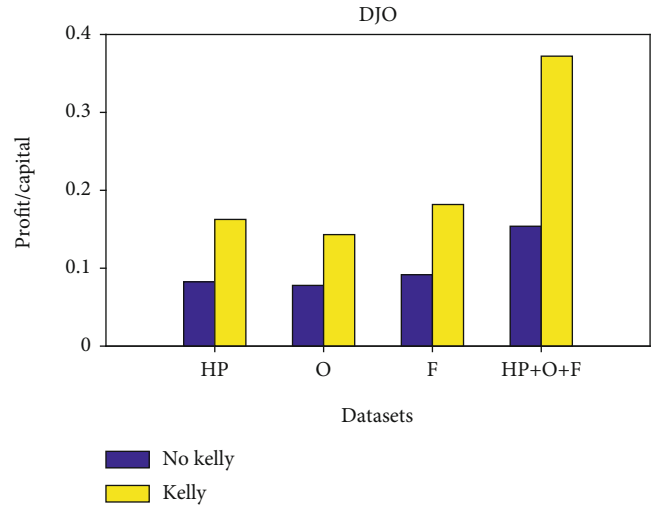


FIGURE 12: The ratio of DJO stock's profit to capital.

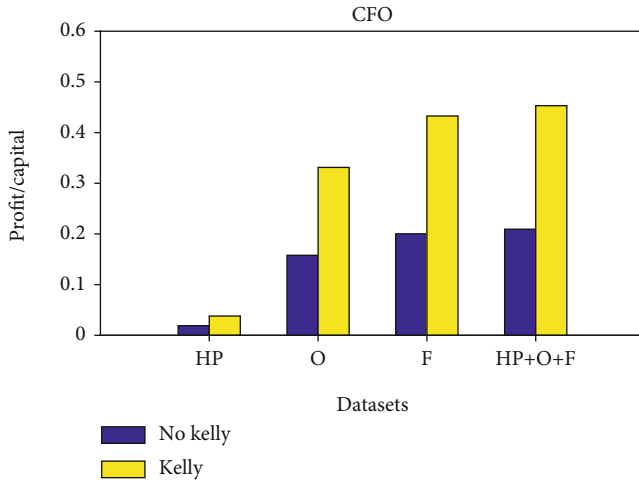


FIGURE 11: The ratio of CFO stock's profit to capital.

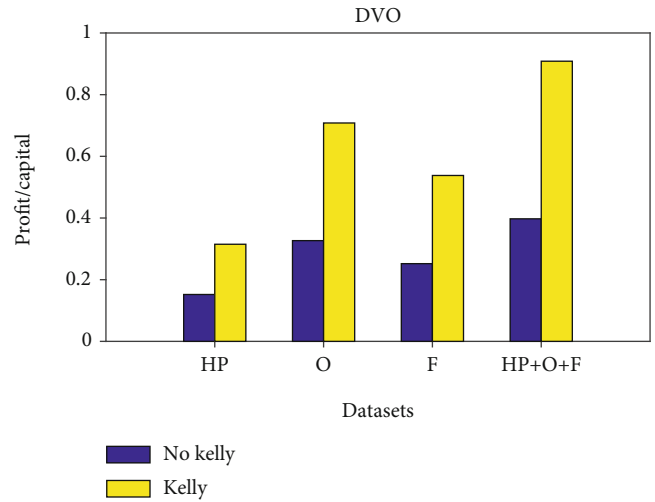


FIGURE 13: The ratio of DVO stock's profit to capital.

selling signals. Thresholds can help investors find trading signals to determine profitable trading strategies. As we all know, stock trading has a certain degree of risk. If you invest incorrectly, you may face bankruptcy. Therefore, the amount of investment needs to be managed. Use the Kelly criterion to determine the proportion of the investment amount and then trade according to the trading strategy to calculate the return on the investment. The flow chart of the stock trading system is shown in Figure 9. The main process regarding the genetic algorithm is shown in Subsection 3.2. The appropriate trading signals are crucial to make high profits in stock trading. Trading signals include buy signals and sell signals, which are employed to allow investors to identify the right time to trade. In addition, the Kelly criterion is applied to determine the proportion of money invested in stock to decrease the risk of trades. The thresholds and the Kelly criterion simultaneously constrain the trading system in

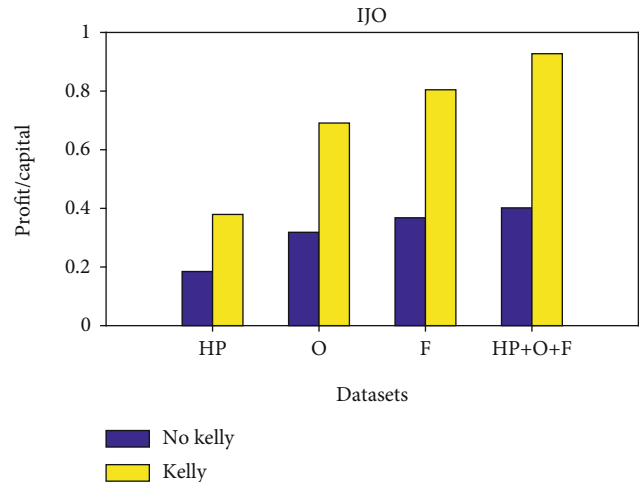


FIGURE 14: The ratio of IJO stock's profit to capital.

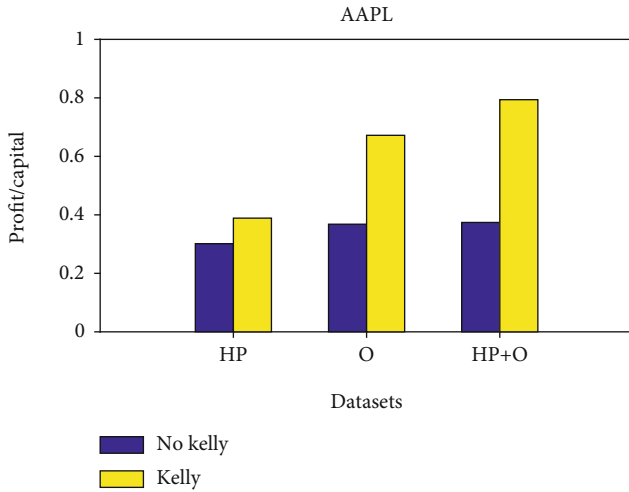


FIGURE 15: The ratio of AAPL stock’s profit to capital.

order to increase the profits of the trade, which is in the final rounded rectangle.

5. Experimental Results

This section mainly introduces the results of the experiment. In the experiment, 10 stocks were selected (including 5 stocks in the Taiwan market and 5 stocks in the US market). Taking stock historical price data, futures data, and option data as data sets, the proposed stock trading system is used to conduct experiments on 10 stocks. “No Kelly” means that the Kelly criterion is not used in the process of seeking profit, and it is set to use 0.2 times of the total funds for transactions each time. “Kelly” means that when trading stocks, the Kelly criterion is used to calculate the share of funds for each transaction.

First, the experimental results of 5 stocks in the Taiwan market are shown in Figures 10–14. The abscissa represents the data set, where “HP” is historical price, “O” is options, “F” is futures, and “HP+O+F” is a data set that combines historical prices, futures, and options. The ordinate represents the ratio of profit to capital. It can be seen from the figure that no matter which data set is used, the ratio of profit to capital obtained by using the Kelly criterion for fund management in stock trading is much higher than that obtained without the Kelly criterion.

Then, experiment on the data set of the US stock market. The experimental results are shown in Figures 15–19. The abscissa represents the data set, where “HP” is historical price, “O” is options, and “HP+O” is a data set that combines historical prices and options. The ordinate represents the ratio of profit to capital. Similarly, when using Kelly criteria for fund management, the benefits obtained are higher. In general, the benefits of using a data set based on historical prices and leading indicators are higher than the benefits of using a single data set.

In terms of stock price volatility forecasting, leading indicators do facilitate accuracy. This experiment utilizes the Kelly criterion for money management of stock trading.

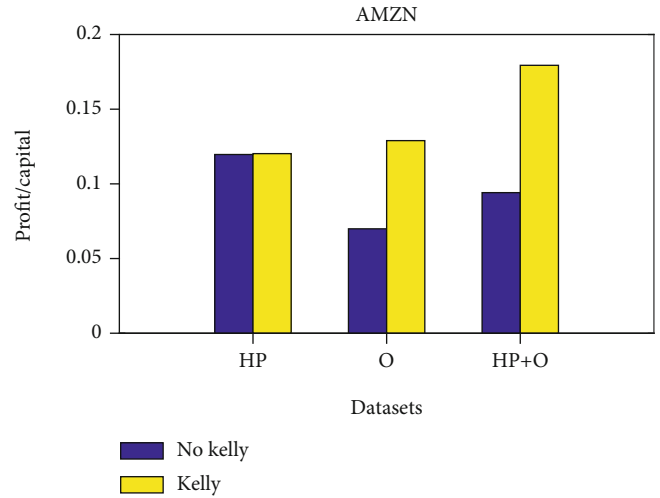


FIGURE 16: The ratio of AMZN stock’s profit to capital.

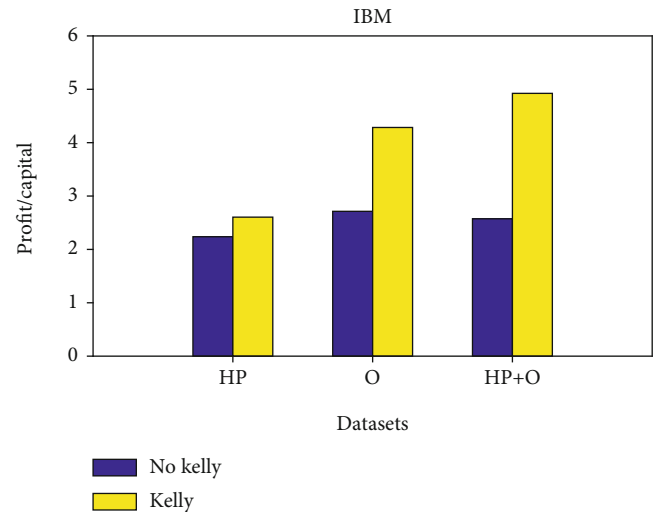


FIGURE 17: The ratio of IBM stock’s profit to capital.

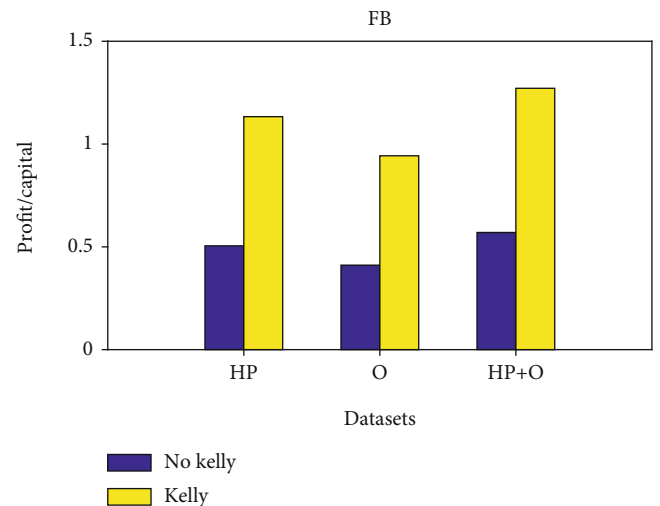


FIGURE 18: The ratio of FB stock’s profit to capital.

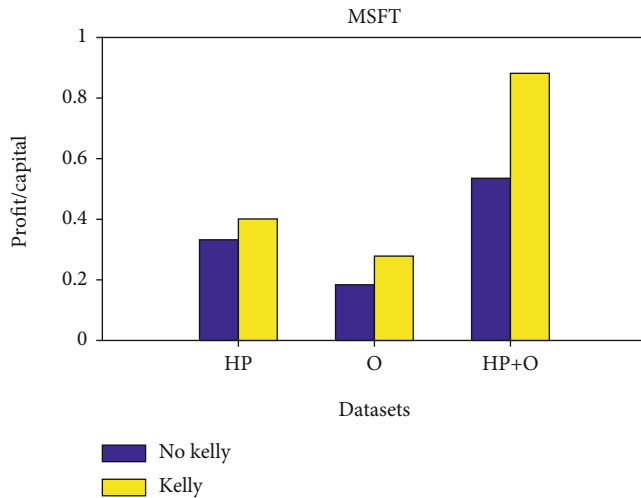


FIGURE 19: The ratio of MSFT stock's profit to capital.

As shown in the above figures, the trading strategy incorporating the Kelly criterion is more profitable in stock trading.

6. Conclusion

Stock trading has always attracted people's attention because of its high return characteristics, but it also comes with great risks. The Kelly Code is a powerful tool for fund management. In this research, a stock trading system is proposed. First, a long- and short-term memory neural network based on leading indicators is used to classify stock price fluctuations. Among them, the stock price fluctuations are divided into three categories: +1, 0, and -1. Second, use genetic algorithms to find the optimal threshold, which is the threshold of trading signals (buy signals and sell signals). Finally, the Kelly criterion is used to determine the best investment score to control the risk of the transaction. A series of experiments show that the use of Kelly's criterion for fund management does reduce the risk of trading, and the return is higher. Although a simple stock trading system has been implemented in this work, there are several aspects requiring further improvement: first, the prediction accuracy of stock price fluctuation trends should be enhanced; second, it is expected that more accurate trading signals can be obtained by other intelligent algorithms.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the Shandong Provincial Natural Science Foundation (ZR201911150391).

References

- [1] Y. Liu, Q. Liu, H. Zhao, Z. Pan, and C. Liu, "Adaptive quantitative trading: an imitative deep reinforcement learning approach," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 2, pp. 2128–2135, 2020.
- [2] A. Patel, D. Patel, and S. Yadav, "Prediction of stock market using artificial intelligence," *SSRN Electronic Journal*, 2021.
- [3] X. Wu, H. Chen, J. Wang, L. Troiano, V. Loia, and H. Fujita, "Adaptive stock trading strategies with deep reinforcement learning methods," *Information Sciences*, vol. 538, pp. 142–158, 2020.
- [4] J. M.-T. Wu, L. Sun, G. Srivastava, and J. C.-W. Lin, "A novel synergetic lstm-ga stock trading suggestion system in internet of things," *Mobile Information Systems*, vol. 2021, Article ID 6706345, 15 pages, 2021.
- [5] S.-Y. Kuo, C. Kuo, and Y.-H. Chou, "Dynamic stock trading system based on quantum-inspired tabu search algorithm," in *2013 IEEE Congress on Evolutionary Computation*, pp. 1029–1036, Cancun, Mexico, June 2013.
- [6] T.-L. Luo, M.-E. Wu, and C.-M. Chen, "A framework of deep reinforcement learning for stock evaluation functions," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 5639–5649, 2020.
- [7] S. Lototsky and A. Pollok, "Kelly criterion: from a simple random walk to lévy processes," *SIAM Journal on Financial Mathematics*, vol. 12, no. 1, pp. 342–368, 2021.
- [8] W. Mu-En, J.-H. Syu, G. Srivastava, and J. C.-W. Lin, Eds., *Informative Index for Investment Based on Kelly Criterion*, Enterprise Information Systems, 2021.
- [9] M. Nygrén, *Optimizing Technical Indicators with Kelly Criterion*, Aalto University School of Business, 2021.
- [10] M.-E. Wu and W. H. Chung, "A novel approach of option portfolio construction using the Kelly criterion," *IEEE Access*, vol. 6, pp. 53044–53052, 2018.
- [11] S. Mokhtari, K. K. Yen, and J. Liu, "Effectiveness of artificial intelligence in stock market prediction based on machine learning," 2021, <https://arxiv.org/abs/2107.01031>.
- [12] B. N. Mohapatra, B. Nagargoje, P. Zurunge, and S. More, "Artificial Intelligence In Stock Market Investment," *Journal of Engineering*, vol. 28, no. 3, pp. 96–100, 2021.
- [13] K. J. Cohen, S. F. Maier, R. A. Schwartz, and D. K. Whitcomb, *On the Existence of Serial Correlation in an Efficient Securities Market*, Salomon Brothers Center for the Study of Financial Institutions, 1977.
- [14] W.-C. Chiang, T. L. Urban, and G. W. Baldrige, "A neural network approach to mutual fund net asset value forecasting," *Omega*, vol. 24, no. 2, pp. 205–215, 1996.
- [15] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Computing and Applications*, vol. 32, no. 6, pp. 1609–1628, 2020.
- [16] Z. Zhao, H. Zhou, L. Qi, L. Chang, and M. C. Zhou, "Inductive representation learning via cnn for partially-unseen attributed networks," *IEEE Transactions on Network Science & Engineering*, vol. 8, no. 1, pp. 695–706, 2021.
- [17] Z. Zhao, H. Zhou, C. Li, J. Tang, and Q. Zeng, "DeepEmLAN: deep embedding learning for attributed networks," *Information Sciences*, vol. 543, pp. 382–397, 2021.
- [18] J. M.-T. Wu, Z. Li, N. Herencsar, B. Vo, and J. C.-W. Lin, "A graph-based cnn-lstm stock price prediction algorithm with leading indicators," *Multimedia Systems*, pp. 1–20, 2021.

- [19] M. Roondiwala, H. Patel, and S. Varma, "Predicting stock prices using lstm," *International Journal of Science and Research (IJSR)*, vol. 6, no. 4, pp. 1754–1756, 2017.
- [20] W. Lu, J. Li, J. Wang, and L. Qin, "A cnn-bilstm-am method for stock price prediction," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4741–4753, 2021.
- [21] D. M. Q. Nelson, A. C. M. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with lstm neural networks," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1419–1426, Anchorage, AK, USA, May 2017.
- [22] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS One*, vol. 12, no. 7, article e0180944, 2017.
- [23] W. Budiharto, "Data science approach to stock prices forecasting in Indonesia during covid-19 using long short-term memory (lstm)," *Journal of Big Data*, vol. 8, no. 1, pp. 1–9, 2021.
- [24] J. R. Sampson, *Adaptation in Natural and Artificial Systems*, J. H. Holland, Ed., Society for Industrial and Applied Mathematics, 1976.
- [25] D. E. Goldberg and J. H. Holland, *Genetic Algorithms and Machine Learning*, Kluwer Academic Publishers, Norwell, MA, USA, 1988.
- [26] T. C. Fu, C. P. Chung, and F. L. Chung, "Adopting genetic algorithms for technical analysis and portfolio management," *Computers & Mathematics with Applications*, vol. 66, no. 10, pp. 1743–1757, 2013.
- [27] S. Chen and C. Zhou, "Stock prediction based on genetic algorithm feature selection and long short-term memory neural network," *IEEE Access*, vol. 9, pp. 9066–9072, 2020.
- [28] C.-H. Chen and P. Shih, "A stock trend prediction approach based on Chinese news and technical indicator using genetic algorithms," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1468–1472, Wellington, New Zealand, June 2019.
- [29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [30] F. A. Gers, J. Schmidhuber, and F. Cummins, *Learning to Forget: Continual Prediction with lstm.*, MIT Press, 1999.

Research Article

Research on a Power Grid Cascading Failure Prevention and Control Method considering WSN

Huiqiong Deng ^{1,2}, Junyuan Wu ^{1,2}, Jie Luo ^{1,2}, Renwu Yan,^{1,2} Cheng Zhang,^{1,2} Peiqiang Li,^{1,2} Kuo-Chi Chang,^{3,4} and Rongjin Zheng^{1,2}

¹School of Electronic, Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350108, China

²Fujian Provincial University, Engineering Research Center of Smart Grid Simulation Analysis and Integrated Control, Fujian 350118, China

³Department of Applied Intelligent Mechanical and Electrical Engineering, Yu Da University of Science and Technology, Miaoli County, Taiwan

⁴Department of Business Administration, North Borneo University College, Sabah, Malaysia

Correspondence should be addressed to Huiqiong Deng; 1123233466@qq.com

Received 18 August 2021; Revised 25 October 2021; Accepted 29 October 2021; Published 22 November 2021

Academic Editor: Daniel G. Costa

Copyright © 2021 Huiqiong Deng 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 practical application of wireless sensor networks (WSNs) in hot fields is summarized. It is found that compared with traditional monitoring methods, it has better adaptability to complex environments and low cost. It is suitable for monitoring power grid operation parameters. Therefore, this paper combines the above network and cascading failures, analyzes its 24-hour continuous and dynamic monitoring of the operation parameters of the power grid, and considers how to use the obtained parameters to analyze the disturbance of the remaining lines after the initial fault of the power grid. To prevent cascading failures in the power grid, a preventive control model considering safety and economy is proposed, and the model is solved by nondominated sorting genetic algorithm II and particle swarm optimization (NSGA2-PSO). Finally, the rationality of this method is verified in the IEEE39 node system.

1. Introduction

In recent years, WSNs have been applied in military, environmental, medical, power systems, and other fields [1]. The application in the military field mainly takes advantage of its rapid deployment, strong concealment, and good environmental adaptability to quickly collect battlefield information and facilitate leaders to formulate the next battle plan, to effectively monitor the ecological environment and effectively predict possible natural disasters and to reduce economic losses and social impact. The application in the medical field mainly focuses on the remote monitoring and care of patients' physical conditions, which can provide telemedicine services for patients in time [2].

In the field of power grid parameter monitoring, its wide application will promote the rapid development of the smart grid. Reference [3] proposed a power quality control system model based on the wireless network, which is used to realize efficient information transmission between sensors and facilitate the comprehensive analysis of power quality in the distribution area. Learn from ZigBee technology to establish a remote monitoring system, which is convenient for operators to grasp the current operation state of the power grid in time. If a failure occurs in the power grid, an alarm message will be sent to the monitoring center to remind the operators to take corresponding measures to prevent the further spread of the failure and even lead to a blackout, which is conducive to ensuring the stable

operation of the power grid. Literature [4] analyzes the performance of the linear network model from the aspects of delay and load, introduces cellular communication technology, and proposes a reconfigurable network model, which reduces the delay of data transmission and improves the capacity of the network.

At present, cascading large-scale power outages occur all over the world, such as the large-scale power outage in Texas caused by the cold weather in 2021.2. Because the blackout caused by cascading failures in the power grid will bring significant economic losses and adverse social impact, many scholars pay extensive attention to the mechanism of cascading failures and control measures to prevent cascading failures and gradually get some research results that can be used for reference [5, 6]. For example, in reference [7], considering the utilization degree of relevant transmission lines in the process of transmitting power from generator nodes to load nodes in the power grid and considering the influence of generation capacity and load level, the power flow intermediate index of transmission lines is proposed as the basis for judging key lines. In reference [8], aiming at the disadvantages of the previous active power safety correction control methods in power systems, a safe distance sensitivity index is proposed, which can provide a criterion for the adjustment of generators in the power grid.

According to the selection of control time, the current control for a large-scale power outage caused by cascading failures can be roughly divided into two categories: preventive control before the failures and blocking control in an accident [9]. If the power grid is disturbed by the expected accident set during normal operation so that the power grid is in an abnormal operation state, it is called preaccident preventive control by taking control measures to return the system to a safe operating state; If a failure has occurred in the power grid, taking blocking control to prevent the further spread of cascading failures is called blocking control in the accident [10].

For example, literature [11] proposes a curve fitting method for the cascading failures of transmission line under freezing disaster, monitors the icing thickness of transmission line in real time, and gives early warning before the ice thickness reaches the critical value of line breaking, to achieve the purpose of prevention and control. These studies do not consider the relationship between the security margin of the power grid and the cost of preventive control or blocking control, which may lead to the situation that the security margin of the power grid is still low under the condition of high cost.

According to the above analysis, we find that although the monitoring network has been adopted in the power system, it has not been comprehensively considered with cascading failures. Therefore, this paper will use it to analyze the prevention and control of cascading failures based on 24-hour continuous dynamic monitoring of power grid operation parameters. Firstly, the overall structure of the monitoring network and the architecture of power grid data acquisition are analyzed to analyze the disturbance of the remaining lines after the initial failure of the power grid. Then, the security of the power grid uses the node injection

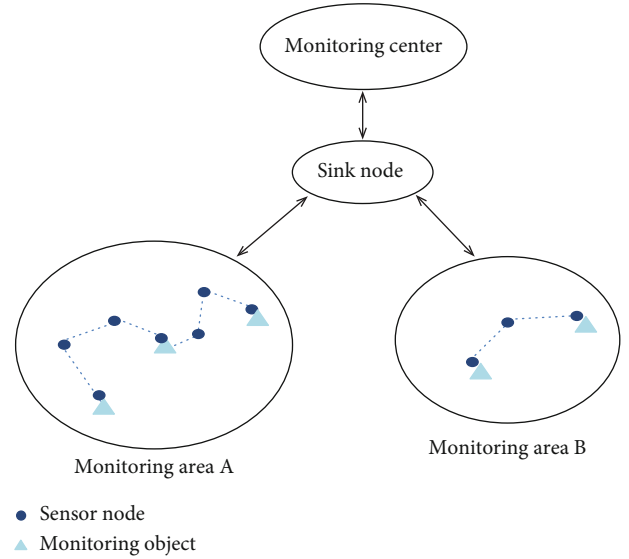


FIGURE 1: Basic structure of WSN system.

power to characterize the current operation state. To prevent cascading failures in the power grid, a preventive control model is proposed. The preventive control model belongs to a two-level optimization model, which is solved by the NSGA2-PSO algorithm. The effectiveness of this method is verified in the IEEE39 node system.

2. Prevention and Control of Cascading Failures in Power Grid Based on WSN

2.1. WSN and Power Grid Monitoring. In the smart grid, to ensure the efficient and reliable transmission of the power grid, it is necessary to monitor the operation of the power grid in real time. The real-time monitoring, diagnosis, and protection of power system operation information can greatly avoid damage to the whole power grid due to power equipment failures or other natural disasters. Therefore, the application of it to the monitoring and data acquisition of power grid can help the operators to carry out online safety evaluation of the state of power equipment and give an alarm in time before the fault, to take necessary measures to prevent the occurrence of cascading failures [12, 13].

2.2. Overall Structure of Monitoring Network. Typical wireless sensor networks are composed of sensor nodes, sink nodes, and monitoring centers. Figure 1 shows the basic structure of the wireless sensor network system. In the monitoring area, according to the monitoring requirements, the wireless sensor node is deployed on the object to be monitored, such as the transmission line. The wireless sensor node cooperates to collect various operating state parameters of the transmission line, such as current and voltage, reaching the sink node in the form of multihop routing, and the sensor node directly measures the voltage value and current value. With the help of the processor of the gateway device, other parameters of the power grid, such as active power and apparent power, can be obtained. Through the ZigBee communication mode and optical fiber, the data

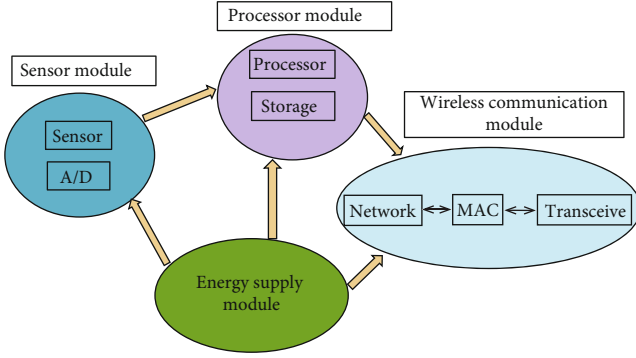


FIGURE 2: Sensor node structure.

is transmitted to the monitoring center to realize the monitoring and management of the system [14].

When transmitting data through the wireless communication channel, it will be affected by electromagnetic noise, which will lead to the loss of collected information. To solve this problem, we can learn from the data reconstruction algorithm of structural noise matrix completion and establish the packet index mechanism, to realize the data recovery [15]. In addition, how to reduce the energy consumption of sensor nodes and prolong the service life of nodes is an important content to be considered in deploying the system shown in Figure 1. Reference [16] proposes a reliable data transmission method for low-energy code distribution, which can effectively reduce energy loss and ensure the reliability of data transmission.

2.3. Sensor Node Structure. As can be seen from Figure 2, the sensor node is composed of four parts: sensor module, a processor module, wireless communication module, and energy supply module [16]. The sensor module mainly realizes the data acquisition of the monitoring object, then A/D conversion of the data and transmission to the processor module, which is responsible for the operation control, data storage, and processing of the whole sensor node. The wireless communication module is responsible for information exchange and communication. The energy supply module is responsible for providing energy sources to ensure the normal operation of each module [17].

2.4. Collection of Power Grid Parameters. As an accurate index to evaluate and feedback the operation state of the power grid, power grid parameters must be monitored dynamically in real time. Therefore, this paper focuses on the sensor module in the sensor node, and its hardware system design is shown in Figure 3.

As shown in Figure 3, in the sensor module, the wireless sensor acquisition node deployed in the power network is responsible for the acquisition, processing, and wireless transmission of voltage and current. It mainly includes a data acquisition circuit, preprocessing circuit, and power module. The power grid parameter acquisition node collects the current signal and voltage signal through the high-precision current transformer and voltage transformer, which are processed by the analog conditioning circuit and input to the sampling unit of CC2530 to complete the A/D

conversion of the signal. Finally, wireless communication is realized by the RF transceiver.

Herein, the measured sampling value is converted into a frequency-domain signal to obtain the voltage phasor U_i of the node and the current phasor I_{ij} of all lines connected to it, which can be expressed as

$$\begin{cases} U_i = \frac{2}{N} \sum_{k=0}^{N-1} u_k e^{-j(2\pi/N)k}, \\ I_{ij} = \frac{2}{N} \sum_{k=0}^{N-1} i_k e^{-j(2\pi/N)k}, \end{cases} \quad (1)$$

where u_k and i_k represent the k th voltage sampling value and current sampling value, respectively, and N is the total number of samples.

According to the voltage phasor and current phasor obtained above, the node injection power can be calculated by further using these data, and the calculation formula is as follows:

$$\begin{cases} S_{ij} = (\dot{I}_{ij})^* \times \dot{U}_i, \\ P_{ij} = \text{re}(S_{ij}), \\ S_i = \sum_{j \in i} S_{ij}, \end{cases} \quad (2)$$

where \dot{U}_i represents the voltage phasor of node i and \dot{I}_{ij} represents the current phasor flowing through the line; S_{ij} is the complex power on the i side of the line; P_{ij} is the real part of S_{ij} , expressed as the active power of the line; and S_i is expressed as the injection power of node i .

2.5. Prevention and Control Model of Cascading Failures in Power Grid. This paper mainly establishes the chain failure prevention and control model from the perspective of safety and economy. At first, WSN is used to obtain the injected power of each node under the current operation state of the power grid and combined with the safety margin index given below to judge the current operation state of the power grid, to measure the safety level of power grid operation at this time. If the safety margin of the power grid is low, the power grid shall be adjusted according to the preventive control model to ensure the normal operation of the power system.

2.5.1. Safety Margin of Power Cascading Failures. Power equipment data uses the monitoring network analyzed above to obtain, and its system structure is shown in Figure 4. The system is composed of sensor nodes and sink nodes. Deployed on the transmission line, it can continuously and dynamically obtain the real-time operation parameters of the power grid for 24 hours. It is necessary and important for the future transmission system and can also greatly improve the stability of the transmission system. The wireless communication network is constructed by the ZigBee wireless communication mode, in which the data is

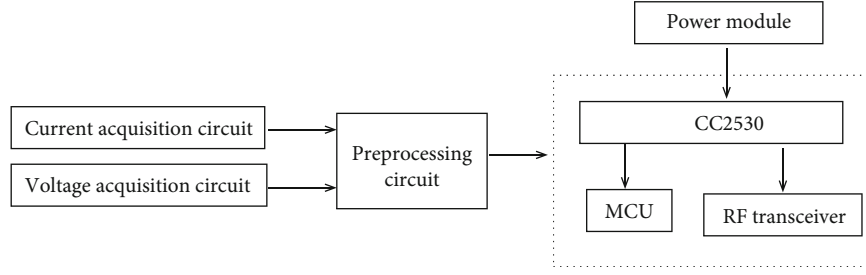


FIGURE 3: Hardware diagram of sensor module.

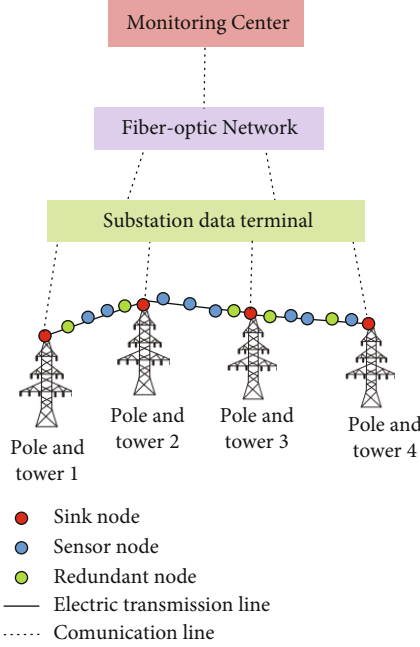


FIGURE 4: Architecture of WSN data collection.

reliably transmitted by a low-energy code distribution algorithm [16]. Firstly, the sensor nodes on the transmission line measure the physical parameters of the conductor, such as voltage and current. Then, these parameters are transmitted to the sink node of the tower by multihop routing, and the sink node forwards them to the substation data terminal. Finally, with the help of an optical fiber network, the data is sent to the remote data monitoring center. Since the delay time of ZigBee technology for data transmission is usually between 15 ms and 30 ms, it meets the requirements of real-time monitoring [18].

In addition, in the process of monitoring information transmission, information loss may occur due to the influence of obstacles. Given this phenomenon, the redundant mechanism is adopted to eliminate the impact of data loss; that is, the system adds redundant nodes, and redundant data are used to improve the accuracy of target identification [19]. If the data is lost due to node failure, take the failed node as the center of the circle and the coverage hole area as the radius to find the nearest redundant node in the area to join the network as a substitute node. Moreover, when redundant nodes are deployed, all nodes in the network do not need to be activated at all times to continuously perform

communication tasks [20]. Some nodes in the network can be activated while other nodes are dormant or energy-saving, to prolong the life cycle of the whole network monitoring task on the premise of ensuring the normal connection of the network [21].

There is no doubt that the application of WSN still faces some urgent problems to be solved, such as how to reduce the energy loss of sensor nodes, prolong the whole network life cycle, and design network communication mechanisms to meet the communication requirements of sensor networks.

Generally, whether the remaining lines in the power grid have interlocking tripping is closely related to the backup protection of the lines. If the electrical quantity of each branch enters the action range of backup protection after power flow redistribution, the backup protection will be started to remove the failure, resulting in secondary failure of the power grid and further aggravation of power flow transfer. When a new line trips due to failure, it is called Level 3 failure, and so on, which will be accompanied by a complex dynamic process.

Assuming that each line of the power grid is equipped with current type backup protection, and the power grid has m branches, when the i th branch L_i has an initial failure and is removed, the operator can use the system in Figure 4 to obtain the current value of the remaining line, compare it with the set value, and judge whether there are cascade failures in the remaining line L_k in the power grid, which is judged by equation (3) [22]:

$$I_{k,\text{dist}} = |I_{k,\text{set}}| - |I_k|, \quad (3)$$

where $I_{k,\text{set}}$ is the backup protection setting value of the k th line, I_k is the current measurement value of the k th line, and $I_{k,\text{dist}}$ is used to measure the interlocking disturbance degree of the remaining branches affected by the initial failure branch. If it is greater, it will be more obviously affected by the initial failures. It is the measured value of the k th line, which is used as the index of measurement and physical quantity. When $I_{k,\text{dist}} > 0$, the branch does not have interlocking tripping; when $I_{k,\text{dist}} = 0$, branch L_k is at the boundary of the interlocking trip. When $I_{k,\text{dist}} < 0$, the branch will have interlocking tripping.

Each transmission line in the power grid is a closely connected whole. When a line fails due to impact or serious disturbance, it will be removed from the power grid to ensure the normal operation of the power grid, which will inevitably

cause the power flow of the failure line to transfer to other lines. The instantaneous tide transfer will not make the operation state of the power grid transient. That is, the injection power of the node remains approximately unchanged. It can be seen that the main factor determining whether cascading failures occur in the power grid is the node injection power of the power grid before the initial failures are removed [23]. Furthermore, the form of safety margin after the initial failures can be given according to the node injection power:

$$D'(S) = \sqrt{(S - S')^2}, \quad (4)$$

where S is the node injection power vector in the current operating state and S' is the corresponding node injection power vector in the critical state of cascading failures. When S changes, it will lead to different values of $D'(S)$, so $D'(S)$ is approximately regarded as a function of S .

2.5.2. WSN and Power Grid Monitoring. To prevent and control cascading failures of the power grid, we should effectively control the operation state of the power grid. Besides, we also pay attention to the economic cost of system operation. In this paper, the prevention and control cost C defined in the literature [24] can be expressed as

$$C = \sum_{i=1}^n \alpha_i (P_{Gi} - P_{G0})^2, \quad (5)$$

where n is the set of generators participating in the adjustment in the system; P_{Gi} and P_{G0} are, respectively, the active output of the generator on the node i before preventive control and the active output of the generator on the node i after preventive control; and α_i is the control cost coefficient.

2.5.3. WSN-Based Power Grid Cascading Failure Prevention and Control Model. According to the safety margin index, it can be judged that the power grid is likely to be in abnormal operation under the disturbance of expected initial failures. Therefore, preventive measures must be taken to improve the safety of power grid operation. In this paper, a double-layer optimization model of chain failure prevention and control is established from the perspective of safety and economy.

(1) Inner Layer Model. The inner layer considers the shortest distance between the current power grid operation state and the critical state of cascading failure, that is, optimize the minimum safety margin of the power grid, and its objective function can be expressed as [23]

$$D(S) = \min (D'(S)), \quad (6)$$

where the smaller the $D(S)$, the closer the system is to the critical point of the interlocking trip; that is, the power grid has a high probability of interlocking trips.

Constraints:

(1) Power flow constraints

Equation (6) is an optimization problem, and the electrical constraints that must be satisfied include equality constraints and inequality constraints. On the one hand, the equality constraints mainly include the power flow constraints of the power grid before the initial failure and the power flow constraints of the power grid after the initial failures. Among them, the power flow constraints of the power grid before the initial failures can be abbreviated as

$$h_0(x) = 0, \quad (7)$$

where h_0 is the mapping relationship corresponding to the power flow before the initial failure occurs and x is the power grid state variable. Using equation (8) can record the power flow constraint of the power grid after the initial failure:

$$h_1(x) = 0, \quad (8)$$

where h_1 is the mapping relationship corresponding to the power flow after the initial failure occurs and x is the power grid state variable.

On the other hand, the inequality constraints to be satisfied during the normal operation of the power grid are shown in

$$\begin{cases} P_{Gi,\min} \leq P_{Gi} \leq P_{Gi,\max}, & i = 1, \dots, N_1, \\ Q_{Gi,\min} \leq Q_{Gi} \leq Q_{Gi,\max}, & i = 1, \dots, N_1, \\ P_m \leq P_{m,\max}, & m = 1, 2, \dots, l, \\ U_{k,\min} \leq U_k \leq U_{k,\max}, & k = 1, 2, \dots, N_2, \end{cases} \quad (9)$$

where P_{Gi} and Q_{Gi} , respectively, represent the active output and reactive output of the i th generator in the system; $P_{Gi,\min}$ and $P_{Gi,\max}$ are the lower limit and upper limit of the active output of the i th generator, respectively; $Q_{Gi,\min}$ and $Q_{Gi,\max}$ are the lower limit and upper limit of reactive output of the i th generator, respectively, and N_1 is the total number of generators in the power grid; P_m is the active power transmitted by branch L_m , and $P_{m,\max}$ is the active power limit transmitted by branch L_m ; and U_k is the voltage of node k , $U_{k,\max}$ and $U_{k,\min}$ are the upper and lower limits of the voltage of node k , respectively, N_2 is the total number of nodes of the power grid, and equation (9) is written in abbreviated form, which can be expressed as

$$g_0(x) \geq 0. \quad (10)$$

Similarly, the inequality constraints that need to be satisfied after power grid failure can be abbreviated as

$$g_1(x) \geq 0. \quad (11)$$

(2) Constraints during critical operation of power grid

Under the disturbance of initial failures, combined with the analysis of equation (3), it can be obtained that after the initial failure line is removed, the constraints of the critical state of cascading failures in the power grid are as follows:

$$\begin{cases} |J| = 0, \\ J = \text{diag}(I_{1,\text{dist}}, I_{2,\text{dist}}, \dots, I_{i,\text{dist}}, \dots, I_{m-1,\text{dist}}), \\ I_{i,\text{dist}} \geq 0, \end{cases} \quad (12)$$

where $I_{i,\text{dist}}$ have the same meaning as $I_{k,\text{dist}}$ in equation (3).

(2) *Outer Layer Model.* From the perspective of complex network theory, considering the prevention and control of power grid cascading failures, on the premise of meeting the load requirements and power grid operation characteristics, if the active output of generator units in the power grid is changed to ensure that $D(\mathbf{S})$ in equation (4) is greater than zero and as large as possible, the current operation state of the power grid is far away from the critical state, to achieve the effect of preventing cascading tripping. Moreover, in the prevention and control, the economic cost of power grid operation has also been paid more attention to. That is, when taking control measures, we should not only ensure the smooth and safe operation of the power grid but also consider the economy of different control methods. Therefore, its mathematical expression is

$$\begin{cases} F_1 = \max(D(\mathbf{S})), \\ F_2 = \min(C), \end{cases} \quad (13)$$

where F_1 represents the optimization effect on power grid safety margin, F_1 and F_2 are the objective functions of the outer model, \mathbf{S} represents the node injection power vector, and C is the cost control.

Constraint condition:

To meet the system power flow equality constraints and inequality constraints is necessary when the power grid is in normal operation, such as generator active power, reactive power output constraints, node voltage constraints, and line transmission power constraints. Its mathematical expression is

$$\begin{cases} \text{s.t. } p_0(x) = 0 \\ p_1(x) = 0 \\ m_0(x) = 0 \\ m_1(x) = 0, \end{cases} \quad (14)$$

where $\mathbf{p}_0(\mathbf{x})$ and $\mathbf{p}_1(\mathbf{x})$ are the mapping relations corresponding to the power flow before and after the initial failures, $\mathbf{m}_0(\mathbf{x})$ and $\mathbf{m}_1(\mathbf{x})$ correspond to inequality constraints, and \mathbf{x} is the power grid state variable.

When the power grid is in a safe operation state, equation (15) must be satisfied:

$$I_{i,\text{dist}} > 0 \quad (i = 1, \dots, m-1), \quad (15)$$

where $I_{i,\text{dist}}$ have the same meaning as $I_{k,\text{dist}}$ in equation (3).

3. Solution of Prevention and Control Model

3.1. *Calculate Fitness Function Value.* In this paper, the objective function of the required solution is combined with the constraints to construct an augmented objective function, that is, the fitness function:

$$\begin{cases} f_1 = D(\mathbf{S}) + \sum_k \frac{1}{\alpha_k} [\min(0, -h_k(x))]^2 + \frac{1}{\delta_k} [\min(0, -g_k(x))]^2 + \sum_m \frac{1}{\zeta_m} [\min(0, -d_m(x))]^2, \\ f_2 = F_1 + \sum_k \frac{1}{\alpha_k} [\min(0, -p_k(x))]^2 + \sum_k \frac{1}{\delta_k} [\min(0, -m_k(x))]^2 + \sum_m \frac{1}{\zeta_m} [\min(0, -e_m(x))]^2, \\ f_3 = F_2 + \sum_k \frac{1}{\alpha_k} [\min(0, -p_k(x))]^2 + \sum_k \frac{1}{\delta_k} [\min(0, -m_k(x))]^2 + \sum_m \frac{1}{\zeta_m} [\min(0, -e_m(x))]^2, \end{cases} \quad (16)$$

where equations (12) and (15) are abbreviated as $\mathbf{d}(\mathbf{x}) > 0$ and $\mathbf{e}(\mathbf{x}) > 0$, respectively, and $\mathbf{h}_k(\mathbf{x})$ corresponds to the k th component in $\mathbf{h}_0(\mathbf{x})$ or $\mathbf{h}_1(\mathbf{x})$; $g_k(\mathbf{x})$, $p_k(\mathbf{x})$, $m_k(\mathbf{x})$, and $h_k(\mathbf{x})$ have similar meanings; α , β , and ζ is the penalty factor, which is given according to the actual situation and is positive.

3.2. *The NSGA2-PSO Algorithm.* Literature [25] compares multiobjective particle swarm optimization (MOPSO), non-dominated sorting genetic algorithm (NSGA2), and archived multiobjective simulated annealing algorithm (AMOS), respectively. It is found that the NSGA2 algorithm has the advantages of low complexity and high precision. Therefore,

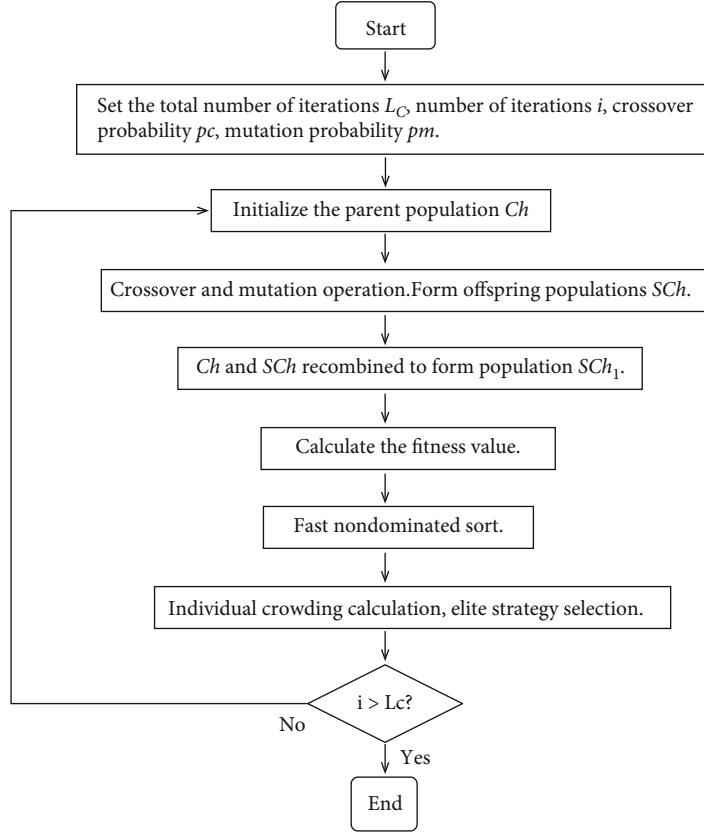


FIGURE 5: NSGA2 algorithm flow chart.

in this paper, the outer layer of the double-layer optimization model uses the NSGA2 algorithm; the PSO algorithm solves the inner model. Among them, the inner model transmits the result of equation (6) to the outer model. The exterior model adjusts the generator's active power output according to the information of the inner model and transmits the current power grid operation state to the inner model for further optimization of the inner model. The population size of the NSGA2-PSO algorithm is usually between 20 and 50. If the total population exceeds 50, the optimization efficiency of the algorithm will not be improved, but the calculation time will be increased [26].

3.2.1. NSGA2 Algorithm. In the traditional multiobjective optimization algorithm, multiple objective functions are usually converted into a single objective function for the solution. However, in the conversion process, it is easy to be affected by the subjective factors of decision-makers, which will cause deviation to the results. Moreover, in the multiobjective problem, each objective restricts and affects the other. On the contrary, the NSGA2 algorithm adopts crowding degree calculation and elite strategy. In the process of evolution, it retains some excellent individuals in the population and reduces the impact of human factors on solution set distribution [27].

The NSGA2 algorithm is a multiobjective intelligent optimization algorithm based on Pareto optimal solution theory. The specific process is shown in Figure 5.

(1) Initialization

Define the number of iterations L_c , crossover probability p_c , and mutation probability p_m . Since the number of variables to be optimized mentioned above is m , the length of each individual is set to m . Assuming that the population size is pop_1 , the initial parent population Ch is randomly generated according to

$$Ch = \text{rand}(pop_1, m) * (\max v - \min v) + \min v, \quad (17)$$

where $\max v$ is the upper limit of the variable and $\min v$ is the lower limit of the variable and $\text{rand}(pop_1, m)$ is a random matrix that generates a pop_1 row and m column with elements within 1.

(2) Crossover, mutation, and recombination

The parent population Ch was crossed and mutated to generate the offspring population SCh . The parent population and the child population were recombined to form the population SCh_1 . The convergence of normal power flow and breaking power flow of ground-state data of each particle is verified. All particles can enter the next step only after passing the verification. If the power flow of a particle does not converge, the original ground state data is given to the particle.

- (3) Calculate fitness value
- (4) Fast nondominated sorting

According to the individual's nondominated ranking value i_{rank} , the whole population is nondominated stratified. Individuals with the same nondominated ranking value i_{rank} belong to the same nondominated layer.

- (5) Individual congestion distance calculation

The purpose is to selectively sort the individuals with the same nondominated order value and select the excellent individuals. Set the crowding distance of the i th individual in the same layer to 0, and record it as $L_d[i] = 0$. The individuals in the same layer are arranged in descending order according to the fitness value. Next, let $L_d[1] = L_d[\text{end}] = M$, where M is a large number. Finally, calculate the congestion distance for the individuals in the middle of the ranking according to equation (18) [28]:

$$L_d[i] = \frac{L_d[i] + (L_{d,m}[i+1] - L_{d,m}[i-1])}{f_{m,\max}, f_{m,\min}}, \quad (18)$$

where $L_{d,m}[i+1]$ represents the m th fitness value of the $i+1$ st individual, $f_{m,\max}$ represents the maximum fitness value corresponding to the m th objective function, and $f_{m,\min}$ represents the minimum fitness value corresponding to the m th objective function.

- (6) Elite strategy selection calculation

The population SCh_1 was screened to obtain a new parent population Ch . Therefore, according to the ranking of i_{rank} from low to high, particles with the same i_{rank} value are put into Ch in turn until they exceed the population size pop_1 when they are put into a certain layer of F_i ; then, they are sorted from large to small according to the crowding distance of individuals in F_i and continue to fill Ch until the population number is pop_1 . Enter the next cycle with the newly obtained parent population Ch .

3.2.2. PSO Algorithm. Particle swarm optimization (PSO) simulates the predation process of birds and continuously approaches the best advantage of search according to each particle's learning and group experience. Each particle's speed and position can reflect its motion state, and the speed and position of each particle can be updated according to equation (19) [29]:

$$\begin{cases} v_i^{k+1} = \omega v_i^k + c_1 r_1 (P_{\text{best},i} - x_i^k) + c_2 r_2 (g_{\text{best}} - x_i^k), \\ x_i^{k+1} = v_i^k + x_i^{k+1}, \end{cases} \quad (19)$$

where v_i^{k+1} and x_i^{k+1} are the iterative speed and position of particle i at $k+1$ st time, respectively; v_i^{k+1} meets $v \min \leq v_i^{k+1} \leq v \max$; $P_{\text{best},i}$ is the optimal solution experienced by particle i itself; G_{best} is the population optimal solution; W

is the inertia coefficient, which generally decreases linearly from 0.9 to 0.1; c_1 and c_2 are acceleration factors, and their values are generally taken as 2; and r_1 and r_2 are random numbers on $[0, 1]$.

To search the critical operation state of interlocking tripping, the line most affected by the initial failure and the node-set that plays a pivotal role in the branch can be analyzed, which will help to improve the calculation speed. Therefore, when solving the inner model, this paper uses the power sensitivity matrix to find the key node set [30].

This paper assumes that the reactance value on the line is much greater than the resistance value. From the relevant knowledge, the relationship between branch current and node current can be obtained:

$$I_B = Y_B A^T Y_N^{-1} I_N, \quad (20)$$

where I_B is the branch current vector, Y_B is the branch susceptance matrix, A is the node branch incidence matrix, Y_N^{-1} is the inverse of node susceptance matrix, and I_N is the node current vector. For the convenience of the following description, let $D(\mu) = Y_B A^T Y_N^{-1}$ be the coefficient matrix representing the correlation degree between the branch current vector and the node injection current vector.

Assuming that there are n nodes and m lines in the power network, the expression of line i current can be obtained according to equation (20):

$$I_{i,B} = \mu_{i,1} I_{1,N} + \mu_{i,2} I_{2,N} + \cdots + \mu_{i,n} I_{n,N}. \quad (21)$$

Equation (21) is processed, converted into the relationship between branch power and node injection power, and sorted to obtain

$$P_{i,B} + jQ_{i,B} = \sum_{l=1}^n \frac{\mu_{i,l} (P_{l,N} + jQ_{l,N}) U_{i,B} (\cos \phi_{i,B} + \sin \phi_{i,B})}{U_{l,N} (\cos \phi_{l,N} + j \sin \phi_{l,N})}, \quad (22)$$

where $P_{i,B}$ and $Q_{i,B}$ represent the active power and reactive power flowing through the line i , respectively; $\mu_{i,N}$ is the element in row i and column n of $D(\mu)$; $U_{i,B}$ and $\phi_{i,B}$ are the voltage amplitude and phase angle of the first node of the i branch, respectively; and $U_{l,N}$ and $\phi_{l,N}$ are the voltage amplitude and phase angle of the l th node.

In this paper, only the active power of nodes is considered for the prevention and control of cascading failures, so equation (22) is simplified to equation (23). That is, the relationship between the active power of transmission line i and the active power injected by all nodes in the grid can be obtained as follows:

$$\begin{cases} P_{i,B} = \sum_{l=1}^n \lambda_{i,l} P_l (\cos \phi_{i,B} \cos \phi_{l,N} + \sin \phi_{i,B} \sin \phi_{l,N}), \\ \lambda_{i,l} = \frac{\mu_{i,l} U_i}{U_l}, \end{cases} \quad (23)$$

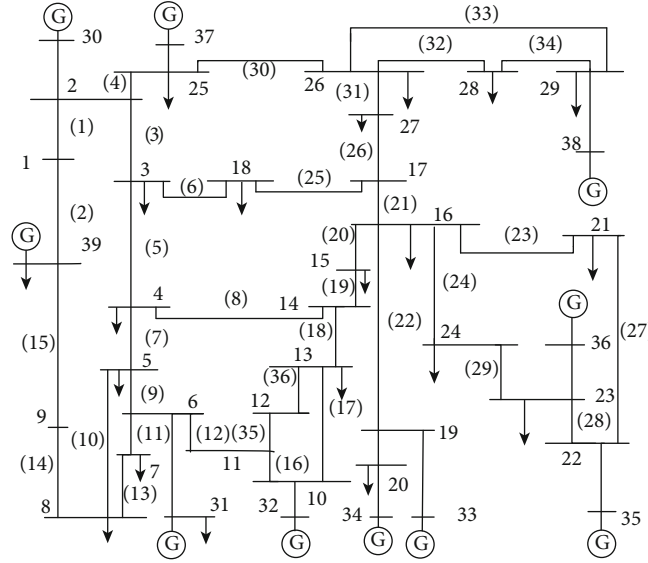


FIGURE 6: Diagram of IEEE39 bus system.

where P_i is the active power passing through the transmission line i , P_l is the active power injected by the l th node, N is the total number of nodes in the system, and $\mu_{i,l}$ is the coefficient matrix $D(\mu)$ of the element that intersects row i and column l .

By further simplifying equation (23), the sensitivity coefficients of branch active power P_i and node injection active power P_l can be obtained, which can be expressed as

$$\beta_{i,l} = \frac{\partial P_i}{\partial P_l} = \lambda_{i,l} P_l (\cos \phi_i \cos \phi_l + \sin \phi_i \sin \phi_l), \quad (24)$$

where $\beta_{i,j}$ is the sensitivity coefficient, and the meaning of other parameters is consistent with equation (23).

Define incidence matrix $\mathbf{L}(\beta)$, $\beta_{i,j}$ is the element in the matrix \mathbf{L} where the i th row and the l th column intersect. If $\beta_{i,j} < 0$ is obtained, it can be multiplied by -1 to make it positive, and the sensitivity after treatment is still recorded as $\beta_{i,j}$. It can be seen from equation (24) that the greater the value of $\beta_{i,j}$, the more obvious the impact on the power of the i th branch when the node l power changes.

When searching for the key nodes in the critical state of the nearest cascading failures, the node injection power vector can be assigned to each particle. The specific methods are as follows: in the first step, calculate the current value $I_{k,\text{dist}}$ of the remaining lines in the power grid after the initial failure shutdown according to equation (3) from the expected initial failures, and calculate the minimum value of $I_{k,\text{dist}}$ and the corresponding branch number and record it as lm in the second step, solve the sensitivity matrix of the most seriously disturbed branch lm according to equations (23) and (24). According to the actual needs, an appropriate threshold E_{value} is defined to judge the relationship between elements $\beta_{i,j}$ and E_{value} of the sensitivity matrix. If $\beta_{i,j} > E_{\text{value}}$, the node number corresponding to element $\beta_{i,j}$ is classified into set \mathbf{A} . Finally, the variable with the solution is set

to the node injection power in set \mathbf{A} , and the penalty function is defined and solved in combination with the particle swarm optimization algorithm.

4. Example Analysis

According to the previous ideas in this paper, the power grid security assessment using WSN is mainly to analyze the power grid based on the input power of the power grid nodes. In order to focus on the verification of the evaluation method, this example assumes that the IEEE39 node system has been configured with a parameter monitoring device according to the structure shown in Figure 6.

Use the system in Figure 4 with the equipment in Figure 7 to collect node voltage and line current data, as shown in Tables 1 and 2, and their values are expressed as per unit value. The study found that in the process of data transmission using a ZigBee routing protocol, part of data packets need to pass through multiple paths, which is easy to cause the drawbacks of uneven energy distribution and low routing efficiency. Therefore, there is a certain error when comparing the power grid operation parameters collected through the above network with the data of the IEEE39 node system [31]. According to the obtained voltage and current, it is easy to calculate the node injection power according to equation (2). Limited to space, the node injection power data will not be listed here. In the calculation, considering the actual characteristics of power flow, when considering the node injection power state, this paper mainly combines the active power and reactive power of PQ node and the active power of PV node to form the vector \mathbf{S} for injection power combination [32]. The control coefficient in equation (5) can be used in the literature [9].

On the IEEE39 node system, the initial fault lines $L_{(17)}$ and $L_{(20)}$ are analyzed as the initial fault lines. Set the population number to 30; the maximum number of iterations of the NSGA2 algorithm is 40; the maximum number of



FIGURE 7: Experimental equipment.

TABLE 1: Voltage collected by WSN.

Node number	Voltage phasor	Node number	Voltage phasor	Node number	Voltage phasor
1	1.051	14	1.022	27	1.041
2	1.055	15	1.026	28	1.053
3	1.042	16	1.035	29	1.051
4	1.024	17	1.035	30	1.052
5	1.025	18	1.033	31	0.983
6	1.027	19	1.049	32	0.984
7	0.998	20	0.993	33	0.998
8	0.998	21	1.009	34	1.023
9	1.029	29	1.045	35	1.048
10	1.026	30	1.044	36	1.064
11	1.023	31	0.991	37	1.031
12	1	32	0.994	38	1.032
13	1.023	33	0.997	39	1.028

TABLE 2: Current collected by WSN.

Branch head node number	Branch end node number	Current phasor	Branch head node number	Branch end node number	Current phasor
1	2	1.161	16	24	0.903
1	39	1.232	17	18	1.819
2	3	3.671	17	27	0.234
2	25	2.461	21	22	5.954
3	4	1.389	22	23	0.653
3	18	0.422	23	24	3.434
4	5	1.651	25	26	0.752
4	14	2.639	26	27	2.693
5	6	4.769	26	28	1.352
5	8	3.209	26	29	1.843
6	7	4.312	28	29	3.395
6	11	3.483	11	12	0.282
7	8	1.897	13	12	0.332
8	9	0.891	31	6	5.782
9	39	0.171	32	10	6.712
10	11	3.482	33	19	7.132
10	13	2.961	34	20	5.131
13	14	2.932	35	22	6.123
14	15	0.339	36	23	5.311
15	16	3.213	37	25	5.272
16	17	2.041	30	2	2.366
16	19	4.372	38	29	8.052
16	21	3.197	20	19	4.363

iterations of the PSO algorithm is 40. The simulation diagram is as follows.

As can be seen from Figures 8 and 9, with the improvement of the power grid safety margin, the control cost of the generator set is also increasing. When adjusting, if only safety margin or control cost is considered, the following situations will occur. The first one is by only considering the improvement of power grid safety margin and ignoring the control cost, the increase of safety margin may not be obvious, but more control funds are invested. The second is only considering reducing control costs and pursuing economic benefits may lead to abnormal operation of the power grid. Therefore, in the actual operation of the power grid, it is

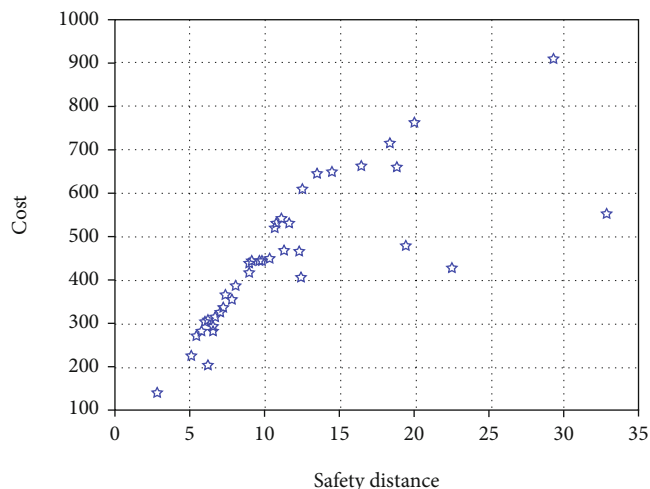


FIGURE 8: Pareto obtained from the initial failure line $L_{(17)}$.

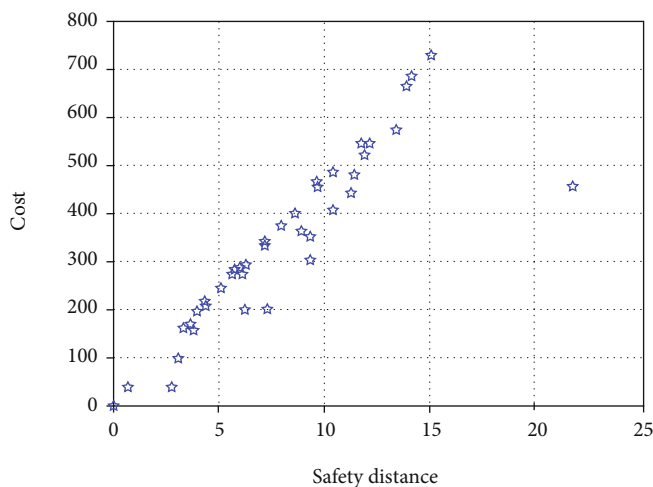


FIGURE 9: Pareto obtained from the initial failure line $L_{(20)}$.

necessary to consider the security margin of the power grid together with the control cost. For example, if the limited control cost (unit: USD) ranges from 400 to 600, from the above figure, the safety margin is between 10 and 15, which can meet the normal operation of the power grid. Moreover, within a given control cost range, it can be flexibly adjusted according to needs, which can ensure the reliable operation of the power grid.

In this paper, the proposed system is verified not only by simulation in MATLAB but also by using the existing experimental equipment. The experimental equipment is shown in Figure 7.

5. Conclusion

This paper combines WSN technology with cascading failure prevention and control and establishes a prevention and control model considering economy and safety. The main conclusions are as follows:

- (1) Using WSN to monitor the power grid parameters can timely feedback the injected power information of the current power grid nodes to the monitoring center. Combined with the safety margin index in this paper, the safety level of the power grid can be quickly judged
- (2) This paper establishes a double-layer optimal preventive control model, and the NSGA2-PSO algorithm is used to solve the model. By adjusting the generator active power output strategy, the power generation cost is minimized based on improving the safety margin, so that the power grid can take into account the requirements of safety and economy at the same time

In short, the method proposed in this paper can provide a reference for further research on cascading failures of the power grid.

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 research was financially supported by the Scientific Research Development Foundation of Fujian University of Technology under the grant GY-Z17149, Scientific and Technical Research Project of Fuzhou under the grant GY-Z18058, and Open Fund of Fujian Provincial University Engineering Research Center under the grant KF-X19016 and KF-D21009.

References

- [1] L. I. Jing, M. Wen, and M. Xue, "Research on key technologies of wireless sensor networks for power grid monitoring," *Journal of Shanghai University of Electric Power*, vol. 33, no. 4, pp. 367–371+384, 2017.
- [2] R. Zhiling, Z. Guangquan, D. Lin, Z. Zhongbao, and Z. Xing, "Overview of wireless sensor network applications," *Sensors and Microsystems*, vol. 37, no. 3, pp. 1-2+10, 2018.
- [3] T. Zhijun, W. Zhongbao, L. Guoqiang, and Q. Yinlong, "Power quality monitoring and analysis system based on ZigBee," *Electrical measurement and instrumentation*, vol. 49, no. 2, pp. 68–71, 2012.
- [4] Y. C. Wu, L. F. Cheung, K. S. Lui, and P. W. T. Pong, "Efficient communication of sensors monitoring overhead transmission lines," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1130–1136, 2012.
- [5] Z. Yue, X. Guanglong, Z. Quan, H. Xinyang, Z. Rui, and Z. Jun, Eds., "Analysis of the February 15 blackout in Texas and its enlightenment to China's power development," *China power*, vol. 54, no. 4, pp. 192–198 + 206, 2021.

- [6] C. Zhongjiao and L. Yun, "Analysis of power grid restoration after "3.21" blackout in Brazil," *Power grid technology*, vol. 45, no. 3, pp. 1078–1088, 2021.
- [7] L. Cai, L. Wenying, X. Dan Yangqing, and Z. X. Peng, "Power flow medium of transmission line and its application in key line identification," *Power system automation*, vol. 38, no. 8, pp. 35–40, 2014.
- [8] G. Xueping, Z. Shuo, L. Haiping, J. Jinghua, and W. Yong, "Risk assessment of power grid cascading faults considering system operation," *Power system protection and control*, vol. 38, no. 24, pp. 124–130, 2010.
- [9] D. Ming, Q. Yucheng, Z. Jingjing, H. Jian, and Y. Jun, "Coordinated control model of power system cascading faults based on risk assessment," *Power system automation*, vol. 40, no. 7, pp. 1–8, 2016.
- [10] W. Jishun, *Static Security Analysis of Power System*, Shanghai Jiaotong University Press, Shanghai, 1985.
- [11] Y. Jianghua, T. Huan, L. Ming, and Z. Lunfang, "Prevention and control of power grid cascading failure under ice disaster," *Journal of power system and automation*, vol. 27, no. 5, pp. 67–74, 2015.
- [12] Z. Yong, "Research on wireless sensor network system based on Zigbee technology for short distance transmission," *Journal of Physics: Conference Series*, vol. 1802, article 022008, no. 2, 2021.
- [13] B. Fateh, M. Govindarasu, and V. Ajarapu, "Wireless network design for transmission line monitoring in smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 1076–1086, 2013.
- [14] L. Miao, C. Xiaobo, J. Xinchun, and S. Jiawen, "Monitoring for overhead transmission lines based on WSN in smart grid," *Shaanxi Electric Power*, vol. 44, no. 10, pp. 1–5, 2016.
- [15] Z. Shuang, *Research on Wireless Link Data Format and Network Topology of Low Energy Consumption Internet of Things*, Jilin University, 2020.
- [16] W. Haiyong, *Research on Key Technologies of Reliable Data Transmission in Wireless Sensor Networks*, Nanjing University of Posts and telecommunications, 2016.
- [17] Z. Yang, *Power Transformer Temperature Monitoring System Based on Wireless Sensor Network*, Nanjing University of technology, 2012.
- [18] L. Hong, Y. Hainian, and W. Guoxing, "Research on power tower monitoring system based on wireless sensor network," *Electrical measurement and instrumentation*, vol. 49, no. 8, pp. 51–54, 2012.
- [19] Q. Yang, S. He, J. Li, J. Chen, and Y. Sun, "Energy-efficient probabilistic area coverage in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 367–377, 2015.
- [20] Y. Jian, *Research on Key Technologies of Fault Tolerance in Wireless Sensor Networks*, Nanjing University of Posts and telecommunications, 2017.
- [21] D. Tao and T. Y. Wu, "A survey on barrier coverage problem in directional sensor networks," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 876–885, 2015.
- [22] F. Weidong and D. Huiqiong, "Research on key nodes in power grid interlocking trip events," *Journal of Fujian Institute of engineering*, vol. 13, no. 6, pp. 578–583, 2015.
- [23] D. Huiqiong, L. Jie, L. Qinbin, Z. Rongjin, and L. Peiqiang, "Grid critical state search considering interlocking tripping," *Journal of Fujian Institute of engineering*, vol. 18, no. 4, pp. 343–348, 2020.
- [24] Z. Jingjing, Y. Yang, L. Xiaoyan, L. Tingting, and L. Xiaojing, "Power system cascading fault coordinated control model considering safety and economy," *Chinese Journal of electrical engineering*, vol. 38, no. 16, pp. 4784–4791 + 4983, 2018.
- [25] H. Liang, *Research on Urban Public Transport Scheduling Optimization Method Based on Multi-Objective*, Jiangxi University of technology, 2020.
- [26] W. Weibo, L. Chuan, and Z. Yongkang, "Experiment and analysis of parameters in particle swarm optimization," *Journal of Xihua University (NATURAL SCIENCE EDITION)*, vol. 1, p. 76–80+105–106, 2008.
- [27] W. Jing, *Multi Objective Optimization Design of Water Supply Network Based on Improved NSGA2 Algorithm*, Beijing University of technology, 2016.
- [28] W. Xiuli, L. Shuhui, C. Haoyong, W. Xifan, and M. Yao, "Multi-objective and multi area power grid planning based on non dominated genetic algorithm and co evolutionary algorithm," *Chinese Journal of electrical engineering*, vol. 12, pp. 11–15, 2006.
- [29] B. Ying, *Research on Multi-Objective Optimization Method Based on Particle Swarm Optimization*, Harbin Engineering University, 2019.
- [30] X. Yan and Z. Jing, "Line overload zoning emergency control strategy based on power sensitivity," *Journal of electrotechnics*, vol. 30, no. 15, pp. 60–72, 2015.
- [31] M. Conti, G. Maselli, G. Turi, and S. Giordano, "Cross-layering in mobile ad hoc network design," *Computer*, vol. 37, no. 2, pp. 48–51, 2004.
- [32] W. Xifan, *Analysis of Modern Power System*, Science Press, Beijing, 2003.

Research Article

The Hybrid Traffic Offloading Mode for Disaster-Resilient Communication Networks Based on User Mobility

Ang-Hsun Tsai¹ and Chung-Hsien Tsai²

¹National Yang Ming Chiao Tung University, Taiwan

²Chung Cheng Institute of Technology, National Defense University, Taiwan

Correspondence should be addressed to Ang-Hsun Tsai; anghsun@gmail.com

Received 8 July 2021; Revised 21 August 2021; Accepted 30 August 2021; Published 20 October 2021

Academic Editor: Chao-Yang Lee

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

Emergency communication systems play a major role in disaster-relief environments. In terms of the public safety research, the emergency relief communication system can provide a high system capacity for networks based on the development of Long-Term Evolution. However, in the event of a disaster, mass traffic information can cause congestion in the core network, and communications between relief workers may be limited. Consequently, spectrum efficiency can be very weak. This paper provides a hybrid traffic offloading mechanism combining Device-to-Device (D2D) and Local IP Access (LIPA) modes for the disaster-resilient network. With receiving power, the distance between relief workers and the distance between relief workers and the vehicular eNodeB (VeNB) as the LIPA/D2D switching criteria, the network can select an appropriate mode to prevent core network congestion. This paper also considers the effects of the mobility models (i.e., random walk and random direction) on the spectrum efficiency of the disaster-resilient communication system. The proposed hybrid LIPA/D2D traffic offloading mechanism can prevent the local communication traffic from flowing into the core network and significantly improve the system spectrum efficiency when the core network is under congestion. Therefore, the proposed mechanism can effectively improve the quality of the communication between relief workers served by the same VeNB for performing rescue operations. Moreover, the hybrid LIPA/D2D traffic offloading mechanism can be applied to the smart city and smart home in the future.

1. Introduction

1.1. Background and Motivation. The evolution of mobile communication systems from the first generation (1G) to the third generation (3G) is still essentially based on the philosophy of conventional telecommunication systems. These mobile communication systems are stable and reliable for voice services but lack sufficient bandwidth and operational resilience for multimedia data services. These deficiencies are well corrected in the fourth generation (4G) mobile communication system. The 4G mobile communication system effectively enhances system capabilities and resolves channel effects and interference to bring a new high-speed Internet experience to users. In addition to the improved network architecture, the 4G mobile communication system is equipped with advanced signal processing technology that enhances the effectiveness of the receiver. In the framework of the IP-based 4G communi-

cation network architecture [1], when a user equipment (UE) communicates with an adjacent receiver within an evolved node B (eNodeB or eNB) service area, data packets shall be transmitted through a radio access network (RAN) to the evolved packet core (EPC) and should then be transferred to the receiver along the original path. The disadvantage of such a transmission link is that all data needs to be transmitted by many network devices and Internet service providers (ISPs) during transmission. In this case, an increase in transmission traffic may lead to high latency and congestion of the core network.

With the dramatic global climate change, different types of natural disasters occur more frequently, resulting in serious social and economic losses and even endangering the lives of the general population. However, where there are major natural disasters (e.g., windstorms, floods, earthquakes, tsunamis, and landslides), buildings, roads, and infrastructure related

to the livelihoods of the general public may be destroyed, complicating rescue operations. In the Great East Japan Earthquake, for example, most eNBs were destroyed, and only a few eNBs could normally provide communication services. Thirty minutes after the earthquake, the total number of calls increased by a factor of eight [2]. The significant increase in data packets is due to the safety confirmation of the general public and their families, instant media reports on the network media, and messages during search and rescue (SAR) operations. Within a short time, a large amount of communication traffic was generated and transmitted to the core network, causing congestion of the core network. Consequently, the delay of the packets or even the abandonment took place, reducing the efficiency of the spectrum in the disaster area.

For a disaster-resilient communication network in the disaster area, an efficient traffic offloading mechanism is necessary. With a rescue operation, the relief workers in the disaster area may randomly move to places where there may be people in help. The relief workers may move by walking or by car, and the mobilities of the relief workers are different. The distance between relief workers are changing all the time. Some messages are passed between workers, and some messages need to be reported back to the command center. Not all the traffic needs to go through the core network. The disaster-resilient communication network needs the traffic offloading mechanism to differentiate between the internal messages and the external messages. In addition, the traffic offloading mechanism should avoid a large amount of disaster relief communication traffic passing through the core network to result in the congestion problem.

This paper considers the use of two traffic offloading mechanisms [3] to resolve the congestion problem faced by the emergency communication network. Using the traffic offloading technology, the communication traffic generated in the disaster area can be transmitted directly or indirectly to the receiver, thereby reducing the communication traffic back to the core network. This can effectively prevent core network congestion from causing a low spectrum efficiency in the disaster area and thereby prevent the interruption of internal communication in the disaster area. One mechanism is Local IP Access (LIPA) [4], in which UE data is directly distributed (or accessed) from eNBs without being transmitted to the core networks of ISPs, as shown in Figure 1. In the case of data transmission between UEs and other nodes in a near end network, local packets can be directly routed via the eNB to the receiver (UE), without being transmitted to the core network. In this way, the transmission delay can be reduced, the traffic can be offloaded, and the load of the core network and the transmission costs can be reduced. The other mechanism is the direct Device-to-Device (D2D) communication [5], in which the devices communicate directly with each other, without being routed through the eNB. In D2D mode [6, 7], spectral resources of the network can be utilized repeatedly to enhance the spectrum efficiency and reduce the load of eNB. The D2D mode is used for communication between adjacent devices. It enables direct communication between devices in order to reduce the backhaul traffic and thereby achieve the offloading effect of traffic.

1.2. Related Works. In recent years, owing to the requirements to improve the system data traffic and reduce the core network load, interest in D2D communication technology has increased. To respond to the demand for more data and less delay in next-generation communication systems, D2D is applied in the existing cellular system, which has become the trend of future communication systems [8, 9]. However, most studies have not examined the potential effects of UE mobility models in actual situations on the transmission rate of communication in the disaster area.

Some studies did not consider the mobility of users in the disaster-resilient communication network, as shown in Table 1. In actual situations, relief workers have different mobility models depending on their operational needs. For example, when relief workers are informed of the exact location of a victim, they perform an SAR operation within the key area only; when relief workers perform casualty transport or fixed-point movement operation in the disaster area, they move at a fixed rate and in a fixed direction. However, in the studies of References [8, 10–14], the simulation scenarios do not account for the mobility of the relief workers in the disaster area. To make the simulation scenario better accord with the actual situation, this paper considered the possible impacts of UE mobility on the communication traffic offloading.

Some papers considered the mobility of users for the D2D communication network, such as [15, 16] and [17, 18], as shown in Table 1. The authors in [15, 16] designed a suitable group-based data synchronization method and proposed a mobility-aware data update mechanism to improve the overhead of maintaining the consistency of shared data items in an MP2P network. A theoretical analysis of transmission rate on D2D mode selection with user mobility has been investigated in [18]. However, the papers [15–18] only considered the D2D mode and did not consider the other traffic offloading mechanism.

1.3. Objective. This paper proposes a hybrid LIPA/D2D traffic offloading mechanism for a disaster-resilient communication network. When a disaster occurs that results in a lack of communication resources in the disaster area, a vehicular eNodeB (VeNB) enters the disaster area, and each relief worker is equipped with a UE that can instantly report his or her global positioning system (GPS) location coordinates. In this case, the hybrid mechanism can switch to a better mode of traffic offloading from D2D or LIPA depending on the distance and received power between the VeNB and relief workers. In addition, this paper investigates the impact of different mobility models such as random walk and random direction during rescue operations on the overall transmission rate, in order to evaluate the efficiency of the hybrid LIPA/D2D traffic offloading mechanism.

1.4. Our Contributions. Under the disaster-resilient communication network proposed in this paper, the hybrid LIPA/D2D traffic offloading mechanism can switch the suitable traffic offloading mode when the core network is congested. Therefore, the hybrid LIPA/D2D traffic offloading mechanism can avoid a large amount of disaster relief communication traffic passing

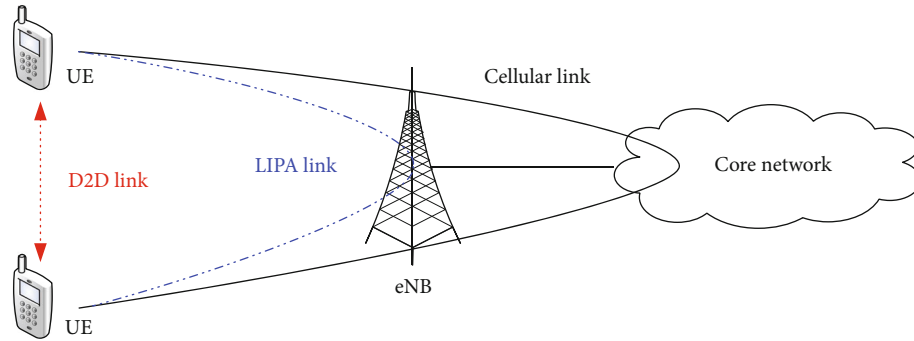


FIGURE 1: Schematic of the communication model.

TABLE 1: Application of synthesis between LIPA/D2D and cellular network.

Related works	Method	Mobility	Criterion
Doppler et al. [8]	D2D	—	Throughput
Frlan [11]	D2D	—	Received power
Tang and Ding [14]	D2D	—	Local IP
Sankaran [13]	LIPA	—	—
Khan et al. [12]	LIPA	—	—
Chang et al. [10]	LIPA	—	—
Orsino et al. [17]	D2D	Random walk	—
Lai et al. [15, 16]	D2D	Random way point	Time
Xu et al. [18]	D2D	Random way point	Received power
Our work	Hybrid LIPA/D2D	Random walk; random direction	Received power; distance

through the core network and causing the delay or interruption of the internal communication network in the disaster area. In addition, the mobilities of the relief workers may be different according to the disaster environment. The single traffic offloading mechanism cannot support the rescue operation. The hybrid LIPA/D2D traffic offloading mechanism can provide the proper link for the relief workers for improving the spectrum efficiency. The simulation results show that the hybrid D2D and LIPA traffic offloading mechanism can further improve the spectrum efficiency of the disaster-resilient communication system under the core network congestion. Therefore, the operation of the disaster-resilient communication network can be maintained, and the transmission of important information can be achieved in the disaster area.

The hybrid LIPA/D2D traffic offloading mechanism also can be applied to other scenes, such as smart city and smart home. When in a crowded area, the traffic for some users may be within the area. At this time, the network can first distinguish the traffic within the area or the traffic that needs to be sent to the core network. If the traffic belongs to the same coverage of the eNB, our traffic offloading mechanism can provide the location-based services (LBS) based on the users' locations. When the two users are close enough, the network can switch to the D2D mode for the users, while the LIPA mode can be switched when the two users are too far apart. This is a very useful mechanism for smart city and smart home. In the future, if the traffic can be actually measured, the traffic offloading model will be trained by

the machine learning and other artificial intelligence techniques for satisfying the demands of various smart cities and smart homes.

The remainder of this paper is organized as follows. Section 2 introduces a disaster relief scenario, channel model, mobility models, and performance metrics. The proposed hybrid LIPA/D2D mode is detailed in Section 3. Section 4 shows the simulation results, and the concluding remarks are given in Section 5.

2. System Model

2.1. System Architecture. A disaster-resilient communication network is considered as shown in Figure 2. In the event of a disaster, the terrestrial communication infrastructure in the disaster area may be destroyed, and the communication services may be interrupted or unstable. The governmental organizations (e.g., Executive Yuan in Taiwan) set up the Central Emergency Operation Center (CEOC) to support the rescue operations. At this time, a disaster relief team and a vehicular eNodeB (VeNB) enter the disaster area to establish a Forward Command Post (FCP) and set up an emergency communication network as soon as possible so that the relief workers in the disaster area can perform the rescue, communication/coordination, and disaster information notification. Relief workers can be regarded as mobile user equipments (MUEs) because they must constantly perform SAR and provide disaster information. The FCP

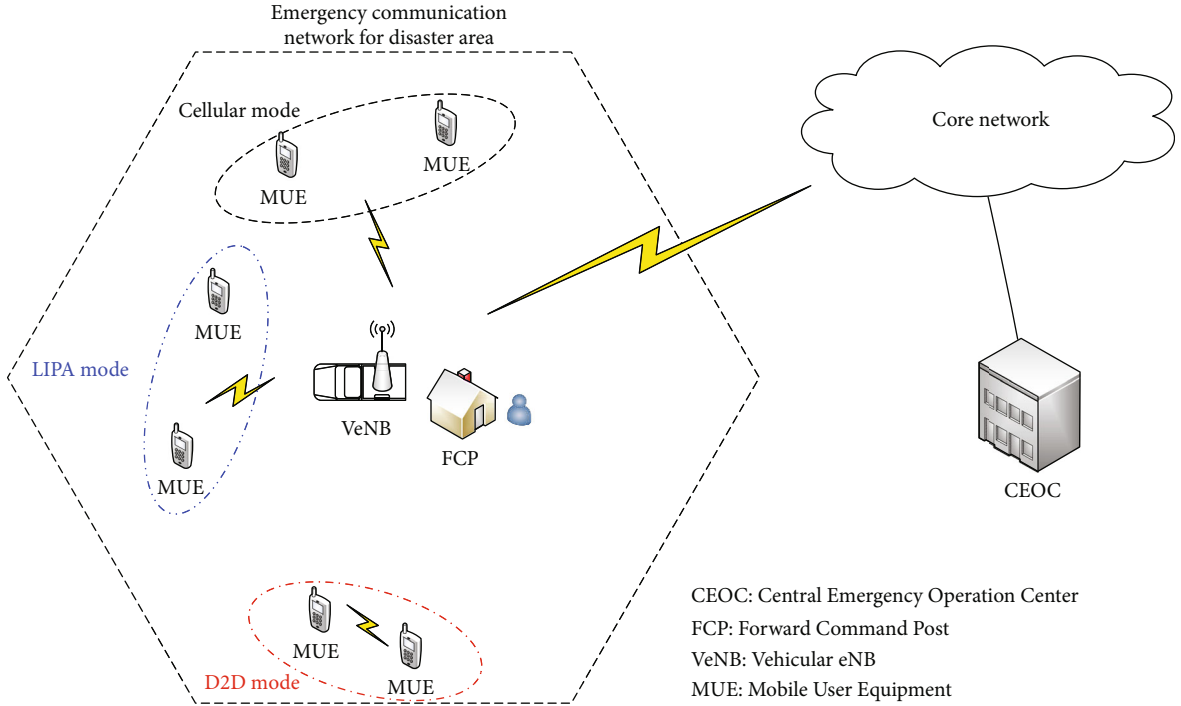


FIGURE 2: A disaster-resilient communication network.

commander will collect all types of disaster information to develop a common operation picture (COP) and report to the CEOC in real time, so that the CEOC can effectively manage the overall situation of the disaster area. In view of the fact that the wired communication infrastructure has been destroyed, the emergency communication network in this context connects to the core network through the wireless transmission.

As shown in Figure 1, when relief workers perform communication/coordination and disaster information reporting in rescue operations, there are mainly three communication modes: cellular mode, Local IP Access (LIPA) mode, and Device-to-Device (D2D) mode. In the cellular mode, the information for intra-area communication/coordination between two devices must be transmitted to the eNodeB (eNB) and the core network before being sent back to the destination device. In the LIPA mode, the information for intra-area communication/coordination between two devices can be transmitted by the eNB directly to the destination device, without being transmitted to the core network. In the D2D mode, the information for intra-area communication/coordination between two devices can be transmitted directly, without being routed via the eNB.

2.2. Channel Model. This paper considers the large-scale fading phenomenon, including two components: (1) path loss caused by the distance between the transmitter and receiver and (2) shadowing caused by the obstacles in the channel. In the simulation scenario of this study, different path losses are set for the LIPA and D2D modes according to References [19–24], as shown in Equation (1). There must be path losses, regardless of the distance between the devices. According to

the technical reports of the Third-Generation Partnership Project (3GPP), the path loss can be calculated as a function of the interdevice distance (in km) and can be expressed as

$$L(d)_{dB} = \begin{cases} 128.1 + 37.6 \log_{10}(d_{\text{cellular}}) \\ 148 + 40 \log_{10}(d_{\text{D2D}}) \end{cases}, \quad (1)$$

where d_{cellular} is the distance from the VeNB to the device and d_{D2D} is the distance between one pair of devices. Generally, the shadowing effect is represented by a log-normal random variable (ξ), which ranges from 6 dB to 10 dB. Because the disaster area selected in this paper is subject to the shadowing effect of the complex terrain, the shadowing effect in LIPA and D2D modes is set to the maximum value (i.e., 10 dB) [19]. In addition to the large-scale fading phenomenon, this paper also considers the small-scale fading phenomenon in the channel model. The small-scale fading is considered as the Rayleigh fading channel.

2.3. Mobility Module. When relief workers are performing search and rescue (SAR) operations in the disaster area, they must adapt their mobility models to different terrain constraints or operational needs due to the complex terrain of the disaster area. Therefore, in this paper, a mobility module has been designed based on the mobility characteristics of relief workers to simulate the mobility patterns during their SAR operations in the disaster area. This module supports two mobility models: random walk model (RWM) and random direction model (RDM).

The RWM can be used to establish a path along for a relief worker to move at a fixed rate in a random direction.

In the beginning, the RWM randomly generates an azimuth angle from $[0, 2\pi]$ as the initial moving direction for a relief worker. The relief worker moves to the next search location at a fixed rate. Then, the RWM randomly regenerates an azimuth angle again from $[0, 2\pi]$ as the next moving direction for the relief worker. This paper assumes that the RWM updates the azimuth angle every second. The RWM can simulate a scenario where a relief worker may keep changing the moving direction when the relief worker is unsure of the key SAR area.

The RDM can be used to simulate that the relief worker has to change a path owing to the constraint of communication coverage. Similar to the RWM, the RDM randomly selects an azimuth angle from $[0, 2\pi]$ as the initial moving direction for a relief worker at first. However, the relief worker keeps moving to the boundary of the communication coverage at a fixed rate. Then, the RDM randomly regenerates an azimuth angle of reflection within the communication coverage from $[0, 2\pi]$ as the following moving direction for the relief worker. Different from the RWM, the RDM updates the azimuth angle of reflection whenever the relief worker moves to the boundary of the communication coverage. The RDM can simulate a scenario in which a relief worker needs to return from the edge of the disaster area.

Figure 3 illustrates the results of the mobility module (i.e., RWM and RDM) simulating two relief workers moving to perform the SAR operation in the disaster area at velocities of 3 km/h in Figure 3(a), 10 km/h in Figure 3(b), and 30 km/h in Figure 3(c), respectively. The simulation scenario is that the VeNB (marked with a diamond) enters the disaster area after the occurrence of the disaster. The area inside the hexagon represents the service area of the VeNB, and two relief workers are combined into a team to enter the disaster area for SAR. The distance between the initial locations of the two relief workers is set as 10 meters, and they constantly move to perform SAR operations within one hour. The movement of the relief workers has been simulated at different speeds of 3 km/h, 10 km/h, and 30 km/h, whether by walking or by a vehicle.

2.4. Efficiency Indicators. The signal-to-noise ratio (SNR) is commonly used to evaluate the performance of systems. As shown in Equation (2), if the signal power is large, then the system efficiency and transmission quality will be high. The SNR has a direct effect on the bit error rate and also influences the information capacity. The ratio of the signal power (S) to the noise power (N_0) is shown as

$$\text{SNR} = \frac{S}{N_0}. \quad (2)$$

For channel characteristics, the large-scale fading (i.e., path loss and shadowing effect) and small-scale fading (i.e., multipath fading) are considered. In this paper, we consider only one VeNB and the UEs of two relief workers in the disaster area. For the LIPA mode, the VeNB has no interference in the uplink, and the UEs have no interference in the downlink. Therefore, the signal-to-noise ratio (SNR) of the VeNB in the uplink and that of the UE in the downlink can be expressed, respectively, as

$$\gamma_t^{\text{UL}} = \frac{P_a \cdot 10(\xi_{\text{UL},t}/10) \cdot |h_{\text{UL},t}|^2 / 10^{(L(d_{\text{UL},t})/10)}}{N_0}, \quad (3)$$

$$\gamma_t^{\text{DL}} = \frac{P_b \cdot 10(\xi_{\text{DL},t}/10) \cdot |h_{\text{DL},t}|^2 / 10^{(L(d_{\text{DL},t})/10)}}{N_0}, \quad (4)$$

where P_a and P_b , respectively, represents the transmit power of a UE and the VeNB, and both of their units are watts. t represents the time duration of SAR of relief workers, i.e., the total simulation time. N_0 represents the noise power. d_{UL} represents the distance between a relief worker and the VeNB in the uplink, and d_{DL} represents the distance between the VeNB and a relief worker in the downlink. L represents the path loss calculated according to Equation (1). $\xi_{\text{UL},t}$ and $\xi_{\text{DL},t}$ are a log-normal random variable to calculate the shadowing effect in the uplink and downlink, respectively. $h_{\text{UL},t}$ and $h_{\text{DL},t}$ represent the Rayleigh fading channel in the uplink and downlink, respectively.

Similarly, there is no interference in the D2D mode, and we assume that the D2D mode is only available in the uplink phase. Therefore, the SNR for D2D communication can be expressed as

$$\gamma_t^{\text{D2D}} = \frac{P_a \cdot 10(\xi_{D,t}/10) \cdot |h_{D,t}|^2 / 10^{(L(d_{D,t})/10)}}{N_0}, \quad (5)$$

where $d_{D,t}$ represents the distance between the devices, $\xi_{D,t}$ is a log-normal random variable to calculate the shadowing effect, and $h_{D,t}$ represents the Rayleigh fading channel.

According to Equations (3) to (5), the theoretical spectrum efficiency of LIPA, D2D, and hybrid LIPA/D2D modes within the time T is expressed as

$$\text{SE}_{\text{LIPA}} = \frac{1}{T} \sum_{t=1}^T \log_2(1 + \min\{\gamma_t^{\text{UL}}, \gamma_t^{\text{DL}}\}), \quad (6)$$

$$\text{SE}_{\text{D2D}} = \frac{1}{T} \sum_{t=1}^T \log_2(1 + \gamma_t^{\text{D2D}}), \quad (7)$$

$$\text{SE}_{\text{D2D}} = \frac{1}{T} \sum_{t=1}^T \log_2(1 + \max\{\min\{\gamma_t^{\text{UL}}, \gamma_t^{\text{DL}}\}, \gamma_t^{\text{D2D}}\}). \quad (8)$$

In Equations (6) to (8), SE_{LIPA} , SE_{D2D} , and $\text{SE}_{\text{Hybrid}}$ represent the theoretical spectrum efficiency in the LIPA, D2D, and hybrid LIPA/D2D modes, respectively. t represents the start time of SAR, and T represents the end time of SAR. In the LIPA mode, assuming that the timeslots are divided into uplink and downlink timeslots, and the total transmission rate is limited by the minimum transmission rate of the uplink and downlink. Therefore, the average effective spectrum efficiency is the lower one between the uplink SNR and downlink SNR. However, the theoretical spectrum efficiency in the D2D mode is calculated directly using the SNR of the D2D link. In the hybrid LIPA/D2D mode, the higher one between the SNR (LIPA) and SNR (D2D) in each

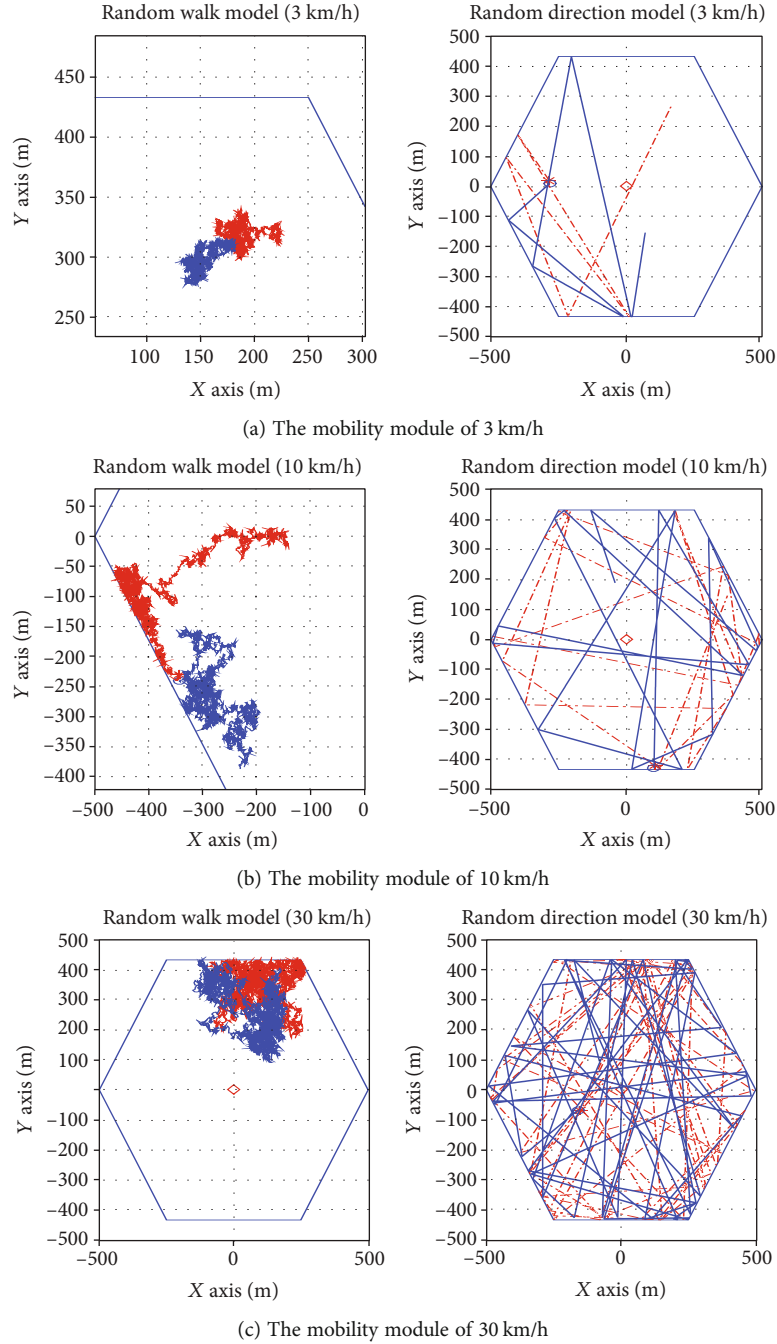


FIGURE 3: Simulation trajectory of the mobility module.

timeslot is used as the theoretical spectrum efficiency of the hybrid LIPA/D2D traffic offloading mechanism.

In this paper, we consider the modulation and coding schemes (MCSs) to mapping the effective spectrum efficiency instead of Shannon's Law in the LTE communication system. That is because the spectrum efficiency calculated by Shannon's Law is the theoretical value. The MCSs have three modulation modes, namely, QPSK, 16-QAM, and 64-QAM with different code rates. Table 2 lists the considered MCSs, the corresponding received SNR, and the effective spectrum efficiency Γ . With the information of received SNR, the MCS

and the corresponding effective spectrum efficiency can be determined according to Table 2. This paper uses the average effective spectrum efficiency as an indicator for performance evaluation.

3. Hybrid LIPA/D2D Mode

As shown in Figure 4, in normal cases, the cellular connection is performed between UEs through the general cellular communication architecture. When a disaster occurs, a VeNB enters the disaster area, and an external communication

TABLE 2: Modulation and coding schemes.

Modulation Scheme	Code Rate	Received SNR (dB)	Effective spectrum Efficiency Γ (b/s/Hz)
QPSK	1/2	-2.5	0.25
QPSK	1/2	0.5	0.5
QPSK	1/2	3.5	1
QPSK	3/4	6.5	1.5
16-QAM	1/2	9	2
16-QAM	3/4	12.5	3
64-QAM	1/2	14.5	3
64-QAM	2/3	16.5	4
64-QAM	3/4	18.5	4.5

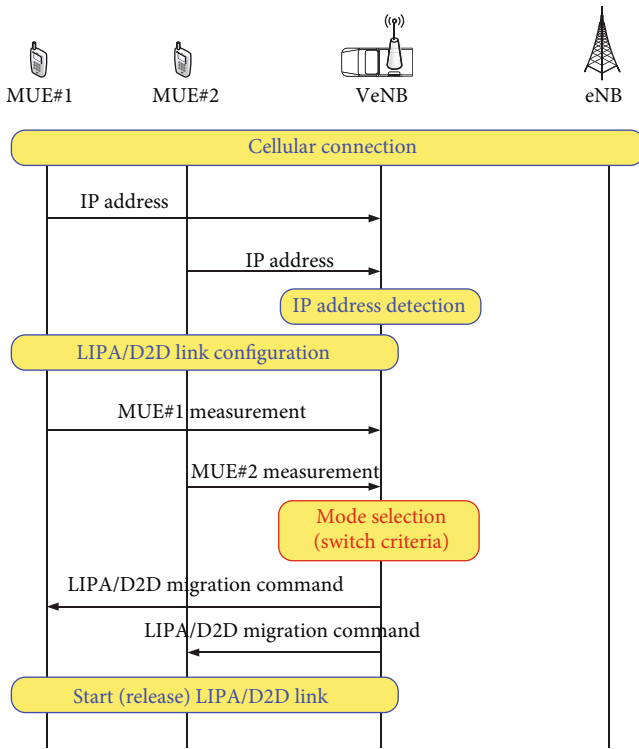


FIGURE 4: A flowchart of the hybrid LIPA/D2D traffic offloading mechanism in the cellular communication mode.

network is constructed. Through the local IP address detection, the VeNB can take over from the destroyed eNB to provide communication services. Since the communication and coordination of relief workers occur primarily within the inside of the disaster area, this paper mainly investigates the switching of intra-area communication modes. After taking over the communication services, the VeNB can measure the reference signal intensity and the GPS location information of MUEs. Then, the VeNB rapidly evaluates the LIPA/D2D downlink transmission in every time slot and notifies the MUEs of the mode switching to obtain the maximum average spectrum efficiency.

The distance and the received power are typically used as criteria for mode switching to investigate the differences in the spectrum efficiency, as shown in Figure 5. After the VeNB enters the disaster area and an external communication network is built, the distance between relief workers (d_3) and the distances between the VeNB and MUEs (d_1 and d_2) are calculated according to the location information that MUE#1 and MUE#2 send back to the VeNB. Then, the distance information can be used to estimate the received power of the VeNB and MUEs for LIPA and D2D links. On the other hand, the received power of the VeNB and MUEs for LIPA and D2D links can be measured by the VeNB and MUEs. Therefore, the received power with estimation and measurement between the VeNB and MUEs (LIPA link) and between MUEs (D2D link) can be used as criteria for switching the transmission mode. If the received power of the direct communication between two relief workers (in D2D mode) is stronger than that in the LIPA mode, the D2D link is selected for communication; otherwise, the LIPA link is chosen.

In the case of using the received power as a criterion for LIPA/D2D switching, the received power of the D2D link (P_r^{D2D}) is calculated according to the MUE transmit power and the path loss of the transmission distance. In the LIPA mode, the uplink received power ($P_r^{LIPA,UL}$) and the downlink received power ($P_r^{LIPA,DL}$) can be estimated based on the large-scale and small-scale fading channels and the transmission power of the VeNB and MUEs. Therefore, the received power of the actual receiver in the LIPA mode (P_r^{LIPA}) is the smaller received power between the uplink and downlink. Following the calculation of the received power in the D2D and LIPA modes, a judgement is made regarding which received power is greater. If the received power in the LIPA mode is higher than that in the D2D mode, the LIPA mode is selected; otherwise, the D2D mode is selected, as shown in Algorithm 1.

In the case of using the distance as a criterion for LIPA/D2D switching, the distance in the D2D mode is the distance between MUEs (d_3). In the LIPA mode, the uplink distance is the distance between MUE#1 and the VeNB (d_1), which is calculated according to the location coordinate sent by MUE#1 back to the VeNB. The downlink distance is the distance between VeNB and MUE#2 (d_2), which is calculated according to the location coordinate sent by MUE#2 back to the VeNB. The uplink and downlink distances are calculated in parallel to obtain the reference values of the reception distance (d^*) in the LIPA mode. After the reference values of the distances (d^*) in the D2D and LIPA modes are calculated, a judgement is made regarding which is greater. If the distance in the LIPA mode is greater than that in the D2D mode, i.e., the distance between MUE#1 and MUE#2 is shorter; the D2D mode is selected; otherwise, the LIPA mode is selected, as shown in Algorithm 2.

4. Simulation Results and Analysis

In this paper, we use MATLAB to simulate the architecture of the disaster-resilient communication networks. The simulation scenario consists of one VeNB and two MUEs. Table 3 presents the simulation parameter settings in the

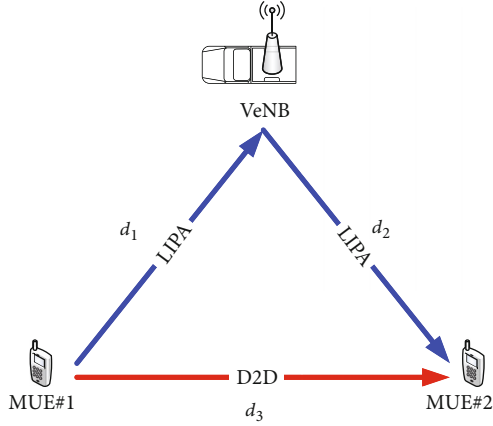


FIGURE 5: Schematic diagram of the traffic offloading mode switching scenario.

hybrid LIPA/D2D, LIPA Only, and D2D Only modes. The service area of the VeNB is set as 500 meters. The SAR team entered the disaster area in pairs. Therefore, there are two MUEs in the disaster area, and the initial distance between the pair of relief workers is set to 10 meters. The mobility rate ranges from 3 km/h to 30 km/h. The transmit power of the MUEs and the VeNB is set as 23 dBm and 36 dBm, respectively. The white Gaussian noise power is set as -174 dBm/Hz.

According to the mode switching criterion, four traffic offloading modes are simulated and analyzed within the SAR time, including (1) the hybrid mode based on the received power, (2) the hybrid mode based on the reception distance, (3) the D2D Only mode, and (4) the LIPA Only mode. The user mobility is also considered as shown in Section 2.3. According to the simulation results at different velocities and the SAR time for RWM and RDM schemes, the hybrid traffic offloading mechanism proposed in this paper can effectively improve the average effective spectrum efficiency at different velocities and a constant SAR time.

4.1. The Effect of the Mobility. Figure 6 shows the average effective spectrum efficiency against the velocity of MUEs under the RWM scheme for the four traffic offloading modes. From the figure, we have the following observations:

- (1) For the D2D Only mode in RWM, the spectrum efficiency decreases as the velocity of MUEs increases. This is because the variance of the distance is smaller when the MUEs move at low speeds. As the speed increases, the variance of distance increases. The two MUEs at a low velocity can maintain a shorter distance than those at a high velocity. Consequently, the D2D Only mode with low speed has a better spectrum efficiency than that with a high speed. However, the spectrum efficiency of the LIPA Only mode keeps almost the same value, irrespective of the velocity variation of MUEs. Since the VeNB can provide the stable signal strength to MUEs with the LIPA Only mode in the coverage of disaster area, the average effective spectrum efficiency can stay a

near constant. In this example, the D2D Only mode can achieve 41.58% more spectrum efficiency than the LIPA Only mode when the velocity of MUEs is at 3 km/h. However, as the velocity of MUEs increases to 30 km/h, the LIPA Only mode can achieve 196.92% more spectrum efficiency than the D2D Only mode. Therefore, when the velocity of MUEs is below 8 km/h, the D2D Only mode can provide better spectrum efficiency than the LIPA Only mode. On the other hand, the LIPA Only mode can provide better spectrum efficiency than the D2D Only mode when the velocity of MUEs is above 8 km/h.

- (2) For the hybrid mode, the spectrum efficiency also decreases as the velocity of MUEs increases due to the increasing variance of distance. Nevertheless, the power criterion scheme can achieve higher spectrum efficiency than the distance criterion scheme. This is because the power criterion scheme is based on the received power measured by the VeNB and MUEs while the distance criterion scheme is based on the reception distance estimated by GPS information. The estimation of the reception distances exists some errors from the measurement of the received power because of the channel effect. However, the error between the power criterion scheme and the distance criterion scheme is insignificant, and the complexity of the power criterion scheme is more than that of the distance criterion scheme. In this example, the power criterion scheme can achieve 0.36% and 2.06% more spectrum efficiency than the distance criterion scheme when the velocity of MUEs is at 3 km/h and 30 km/h, respectively.
- (3) Both schemes of the proposed hybrid mode can achieve higher spectrum efficiency than the D2D Only mode and the LIPA Only mode at an arbitrary velocity, because the proposed hybrid LIPA/D2D traffic offloading mode can switch the appropriate traffic offloading mode based on the current communication environment. In the example, when the velocity of MUEs is 3 km/h, the power criterion scheme of the hybrid mode can achieve 43.03% and 1.03% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively. Moreover, the distance criterion scheme of the hybrid mode can achieve 42.51% and 0.66% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively. As the velocity of MUEs approaches 30 km/h, the power criterion scheme of the hybrid mode can achieve 6.12% and 215.10% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively. Furthermore, the distance criterion scheme of the hybrid mode can achieve 3.98% and 208.75% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively.

Input:
 d_1 //The distance between MUE#1 and VeNB;
 d_2 //The distance between MUE#2 and VeNB;
 d_3 //The distance between MUE#1 and MUE#2;
 P //The transmission power;
 $\{L, \xi, h\}$ //The channel variable.

Output:
LIPA mode or D2D mode.

Pseudocode:

- 1: Calculate $P_r^{LIPA,UL}$ and $P_r^{LIPA,DL}$
- 2: $P_r^{LIPA,UL} = P_a \cdot 10^{\xi_{UL,t}/10} \cdot |h_{UL,t}|^2 / 10^{L(d_{UL,t})/10}$
- 3: $P_r^{LIPA,DL} = P_b \cdot 10^{\xi_{DL,t}/10} \cdot |h_{DL,t}|^2 / 10^{L(d_{DL,t})/10}$
- 4: Determine the received power of the LIPA link (P_r^{LIPA})
- 5: $P_r^{LIPA} = \min \{P_r^{LIPA,UL}, P_r^{LIPA,DL}\}$.
- 6: Calculate P_r^{D2D}
- 7: $P_r^{D2D} = P_a \cdot 10^{\xi_{D,t}/10} \cdot |h_{D,t}|^2 / 10^{L(d_{D,t})/10}$
- 8: **if** $P_r^{D2D} \geq P_r^{LIPA}$ **then**
- 9: Switch to D2D mode
- 10: **else**
- 11: Switch to LIPA mode
- 12: **end if**

ALGORITHM 1: Hybrid LIPA/D2D mode selection based on received power.

Input:
 d_1 // The distance between MUE#1 and VeNB;
 d_2 // The distance between MUE#2 and VeNB;
 d_3 // The distance between MUE#1 and MUE#2;
 d^* // The reference values of the reception distance.

Output:
LIPA mode or D2D mode.

Pseudo code:

- 1: Calculate the reference values of the reception distance (d^*)
- 2: $d^* = d_1 d_2 / d_1 + d_2$
- 3: **if** $d^* \geq d_3$ **then**
- 4: Switch to D2D mode
- 5: **else**
- 6: Switch to LIPA mode
- 7: **end if**

ALGORITHM 2: Hybrid LIPA/D2D mode selection based on reception distance.

TABLE 3: Simulation parameter settings.

Parameters	Values
VeNB radius (R_m)	500 m
Initial distance between two MUEs (R_d)	10 m
Velocity of MUEs	3-30 km/h
Transmission power (MUE/VeNB)	23/36 dBm [24]
Noise power (N_0)	-174 dBm/Hz
Noise figure of VeNB/MUE	7 dB
SAR time	1-24 hours
Shadowing (ξ)	10 dB

Figure 7 shows the average effective spectrum efficiency against the velocity of MUEs under the RDM scheme for the four traffic offloading modes. From the figure, we have the following observations:

- (1) For the four traffic offloading modes in RDM scheme, the average effective spectrum efficiency does not vary too much when the velocity of MUEs increases. That means the average effective spectrum efficiency is irrespective of the velocity of MUEs in the RDM scheme. In this example, when the velocity of MUEs is 3 km/h, the spectrum efficiency of the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, the LIPA Only mode, and the D2D Only mode is 3.05 bits/s/Hz, 3.01 bits/s/Hz, 2.92

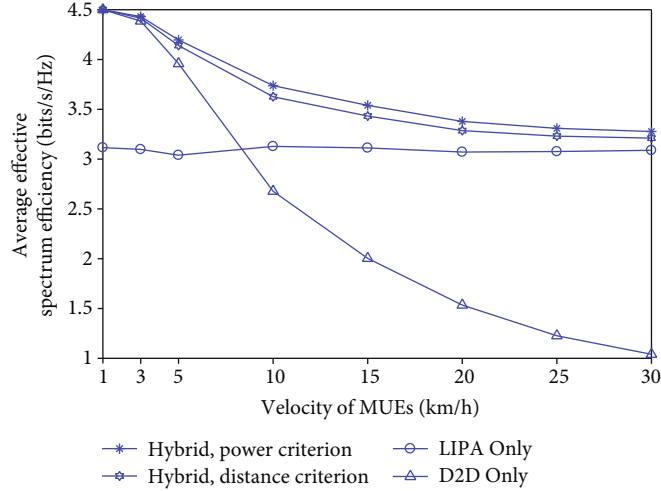


FIGURE 6: Average effective spectrum efficiency versus the velocity of MUEs under the RWM scheme.

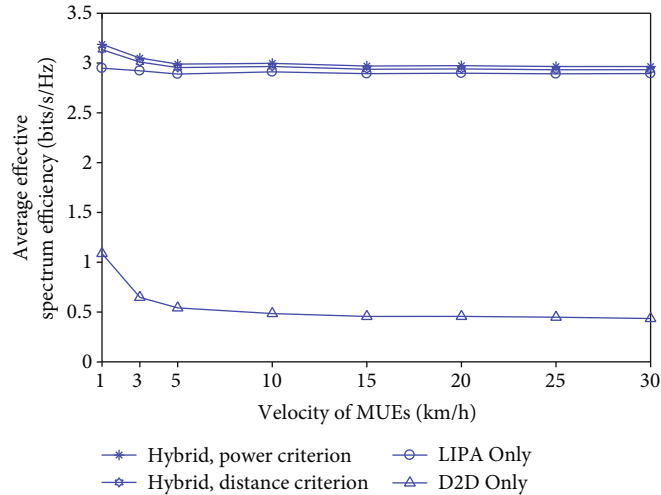


FIGURE 7: Average effective spectrum efficiency versus the velocity of MUEs under the RDM scheme.

bits/s/Hz, and 0.65 bits/s/Hz, respectively. As the velocity of MUEs increases to 30 km/h, the spectrum efficiency of the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, the LIPA Only mode, and the D2D Only mode is 2.97 bits/s/Hz, 2.93 bits/s/Hz, 2.90 bits/s/Hz, and 0.43 bits/s/Hz, respectively. It shows that the impact of the MUE velocity is very small in RDM for the four traffic offloading modes.

(2) The power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, and the LIPA Only mode have the similar average effective spectrum efficiency. However, both schemes of the hybrid mode can perform better spectrum efficiency than the LIPA Only mode. In this example, the power criterion scheme of the hybrid mode and the distance criterion scheme of the hybrid mode can achieve 4.38% and 3.01% more spectrum effi-

ciency than the LIPA Only mode when the velocity of MUEs is at 3 km/h. As the velocity of MUEs increases to 30 km/h, the power criterion scheme of the hybrid mode and the distance criterion scheme of the hybrid mode can achieve 2.42% and 1.31% more spectrum efficiency than the LIPA Only mode. It is shown that the LIPA Only is applicable in the RDM scheme. Nevertheless, the hybrid mode can achieve more performance.

(3) Moreover, the power criterion scheme of the hybrid mode has the best spectrum efficiency, and the D2D Only mode has the worst performance. This is because the RDM scheme makes each MUE search in one random direction and then search in another random direction when the MUE reaches the boundary of the disaster area. Therefore, the separation distance between the two MUEs is usually too far during the rescue operation to benefit from the

D2D Only mode. In this example, the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, and the LIPA Only mode can achieve 371.04%, 364.86%, and 351.27% more spectrum efficiency than the D2D Only mode when the velocity of MUEs is at 3 km/h. As the velocity of MUEs increases to 30 km/h, the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, and the LIPA Only mode can achieve 581.70%, 574.34%, and 565.61% more spectrum efficiency than the D2D Only mode. The result shows that the D2D Only mode is not applicable in the RDM scheme.

4.2. The Effect of the SAR Time. Figure 8 shows the average effective spectrum efficiency against the SAR time of MUEs under the RWM scheme for the four traffic offloading modes. From the figure, we have the following observations:

- (1) For the D2D Only mode in RWM, the spectrum efficiency decreases as the SAR time of MUEs increases. This is because the variance of the distance between MUEs is small in the beginning of the rescue operation for RWM scheme. As the SAR time increases, the region of the rescue becomes more and more broad and so is the distance of MUEs. Accordingly, the D2D Only mode cannot effectively achieve a better spectrum efficiency with a long SAR time. On the other hand, the LIPA Only mode has a nearly constant spectrum efficiency no matter how long the SAR time is. In this example, the D2D Only mode can achieve 25.39% more spectrum efficiency than the LIPA Only mode when the SAR time of MUEs is three hours. However, as the SAR time of MUEs increases to 24 hours, the LIPA Only mode can achieve 53.85% more spectrum efficiency than the D2D Only mode. Therefore, if the SAR time of MUEs is below eight hours, the D2D Only mode can provide better spectrum efficiency than the LIPA Only mode. However, the LIPA Only mode can provide better spectrum efficiency than the D2D Only mode when the SAR time of MUEs is above eight hours.
- (2) Similar to the D2D Only mode, both of the hybrid mode have the decreasing spectrum efficiency when the SAR time of MUEs is increasing. Although the power criterion scheme can achieve more spectrum efficiency than the distance criterion scheme, the distance criterion scheme has less complexity than the power criterion scheme. Furthermore, the difference in spectrum efficiency between the measurement of the received power and the estimation of the reception distances due to the channel effect is slight. In this example, the power criterion scheme can achieve 1.56% and 3.08% more spectrum efficiency than the distance criterion scheme when the SAR time of MUEs is 3 hours and 24 hours, respectively.

- (3) Since the proposed hybrid LIPA/D2D traffic offloading mode can switch the proper traffic offloading mode according to the real communication environment, both schemes of the hybrid mode can achieve higher spectrum efficiency than the D2D Only mode and the LIPA Only mode with a given SAR time. In the example, when the SAR time of MUEs is three hours, the power criterion scheme of the hybrid mode can achieve 35.13% and 7.77% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively. In addition, the distance criterion scheme of the hybrid mode can achieve 33.05% and 6.11% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively. If the SAR time of MUEs is 24 hours, the power criterion scheme of the hybrid mode can achieve 14.18% and 75.77% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively. Moreover, the distance criterion scheme of the hybrid mode can achieve 10.77% and 70.53% higher spectrum efficiency than the LIPA Only mode and the D2D Only mode, respectively.

Figure 9 shows the average effective spectrum efficiency against the SAR time of MUEs under the RDM scheme for the four traffic offloading modes. From the figure, we have the following observations:

- (1) In light of the figure, the influence of the SAR time on the average effective spectrum efficiency is insignificant for the four traffic offloading modes in RDM scheme. In this example, when the SAR time of MUEs is three hours, the spectrum efficiency of the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, the LIPA Only mode, and the D2D Only mode is 2.98 bits/s/Hz, 2.95 bits/s/Hz, 2.90 bits/s/Hz, and 0.49 bits/s/Hz, respectively. As the SAR time of MUEs increases to 24 hours, the spectrum efficiency of the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, the LIPA Only mode, and the D2D Only mode is 2.96 bits/s/Hz, 2.93 bits/s/Hz, 2.89 bits/s/Hz, and 0.43 bits/s/Hz, respectively.
- (2) Both schemes of the hybrid mode and the LIPA Only mode have the similar average effective spectrum efficiency. Nevertheless, the power criterion scheme of the hybrid mode has the best spectrum efficiency, and the distance criterion scheme of the hybrid mode has better spectrum efficiency than the LIPA Only mode. In this example, the power criterion scheme of the hybrid mode and the distance criterion scheme of the hybrid mode can achieve 2.97% and 1.80% more spectrum efficiency than the LIPA Only mode when the SAR time of MUEs is three hours. As the SAR time of MUEs increases to 24 hours, the power criterion scheme of the hybrid mode and the distance criterion scheme of the hybrid mode can achieve 2.35% and

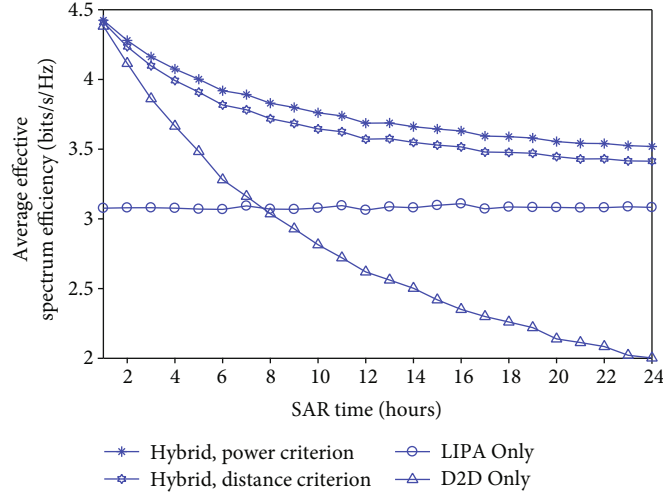


FIGURE 8: Average effective spectrum efficiency versus the SAR time of MUEs under the RWM scheme.

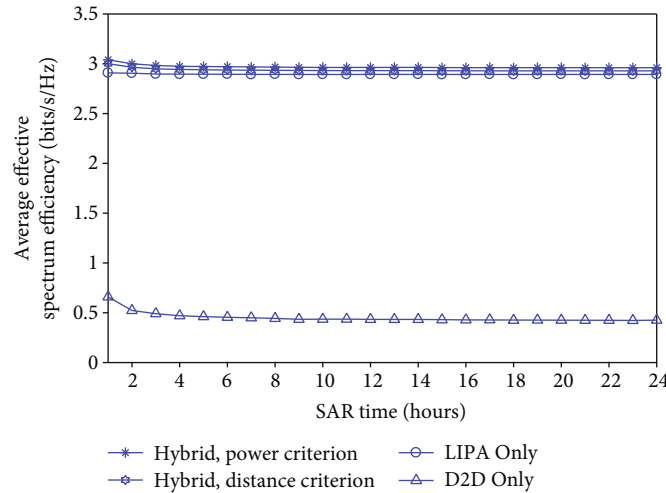


FIGURE 9: Average effective spectrum efficiency versus the SAR time of MUEs under the RDM scheme.

1.24% more spectrum efficiency than the LIPA Only mode. Although the LIPA Only is applicable in the RDM scheme, the hybrid LIPA/D2D traffic offloading mode can achieve more performance.

- (3) The power criterion scheme of the hybrid mode has the best spectrum efficiency, while the D2D Only mode has the worst performance. Because the distance between the two MUEs during the rescue operation in RDM scheme cannot be kept within an acceptable range, the D2D Only mode cannot provide the satisfactory quality of service for the disaster-resilient communication network. In this example, the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, and the LIPA Only mode can achieve 508.03%, 501.10%, and 490.50% more spectrum efficiency than the D2D Only mode when the SAR time of MUEs is 3 hours. As the SAR time of MUEs

increases to 24 hours, the power criterion scheme of the hybrid mode, the distance criterion scheme of the hybrid mode, and the LIPA Only mode can achieve 503.55%, 497.02%, and 489.69% more spectrum efficiency than the D2D Only mode. It is shown that the LIPA Only mode is more applicable than the D2D Only mode in the RDM scheme, and the hybrid LIPA/D2D traffic offloading mode can improve more spectrum efficiency.

4.3. Summary. The RWM scheme can simulate a scenario of user mobility in which the SAR of relief workers is narrowed or expanded within a fixed area, while the RDM scheme can simulate the linear movement of relief workers (fixed-point movement or casualty transport). The distance between relief workers changes with respect to the velocity or SAR time. In the D2D Only mode, the average effective spectrum efficiency depends on the reception distance or received

power between relief workers. In the LIPA Only mode, messages are forwarded by the VeNB to the receiver, therefore the average effective spectrum efficiency is less likely to be affected by the reception distance or received power. Since the initial locations of relief workers are adjacent, the average effective spectrum efficiency in the D2D Only mode is higher than that in the LIPA Only mode. However, the distance between relief workers increases with the user mobility and SAR time, and the average effective spectrum efficiency in the LIPA mode may be higher than that in the D2D mode. In addition, the spectrum efficiency of the RDM scheme is lower than that of the RWM scheme for the D2D Only mode.

According to the characteristics of the two aforementioned mobility models and the mode switching standards, the hybrid traffic offloading mode can select a better mode (D2D or LIPA) for obtaining the highest average effective spectrum efficiency. That is, when the distance between relief workers is short enough to provide adequate signal quality with velocity or SAR time, the average effective spectrum efficiency of the hybrid mode can exhibit a similar trend to that of the D2D Only mode. However, the average effective spectrum efficiency of the hybrid mode can exhibit a similar trend to that of the LIPA Only mode if the distance between relief workers is too long to provide adequate signal quality with velocity or SAR time. Therefore, the average effective spectrum efficiency in the hybrid mode changes depending on whether LIPA mode or D2D mode is selected. In addition, when the received power is used as the criterion for mode switching, the average effective spectrum efficiency is slightly higher than that when the distance is used as the criterion for mode switching. However, the relative complexity is higher. Therefore, by sacrificing a little performance, the distance is a good choice as the criterion for mode switching.

5. Conclusions

This paper investigated the effect of user mobility on the spectrum efficiency with four traffic offloading modes for disaster-resilient communication networks. The LIPA Only mode needs the eNBs surviving in the disaster area or the VeNB entering the disaster area to provide communication services for relief workers in the disaster area. If eNBs in the disaster area are destroyed or the VeNB cannot enter the disaster area owing to terrain constraints, the LIPA link is interrupted. In the D2D Only mode, because the transmit power of communication devices such as mobile phones is small, the communication quality deteriorates owing to the complex terrain or distance between relief workers in the disaster area. The hybrid LIPA/D2D traffic offloading mode can overcome the limitations of both LIPA Only mode and D2D Only mode to provide a more stable emergency communication system. This is because the proposed hybrid LIPA/D2D traffic offloading mechanism can switch LIPA link or D2D link based on the criterion of the received power or reception distance between MUEs of relief workers and the VeNB. In addition, the overall average effective spectrum efficiency is evaluated according to the different mobility

characteristics (i.e., RWM and RDM) of the relief workers. When the hybrid LIPA/D2D traffic offloading mechanism is used, the local communication traffic generated during the rescue operation is not sent back to the core network, reducing the load of the core network. In this manner, the quality of communication from the disaster area to the outside network is ensured, which is beneficial to the implementation of disaster relief.

Data Availability

No datasets were generated or analyzed during the current study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was sponsored by the Ministry of Science and Technology (MOST) of Taiwan under grants MOST 109-2221-E-606-007-, MOST 110-2634-F-009-021-, and MOST 1102221-E-A49-039-MY3. The authors would like to thank Wen-Pin Chen for his assistance in simulating traffic offloading modes in the experiment.

References

- [1] 3GPP, "Technical Specifications and Technical Reports for a UTRAN-based 3GPP system," *3GPP TSG SSA, Tech. Rep. TS 21.101 V14.1.0*, 2017.
- [2] Y. Shibata, N. Uchida, and N. Shiratori, "Analysis of and proposal for a disaster information network from experience of the Great East Japan Earthquake," *IEEE Communications Magazine*, vol. 52, no. 3, pp. 44–50, 2014.
- [3] R. Maallawi, N. Agoulmine, B. Radier, and T. B. Meriem, "A comprehensive survey on offload techniques and management in wireless access and core networks," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1582–1604, 2015.
- [4] M. J. Yang, S. Y. Lim, H. J. Park, and N. H. Park, "Solving the data overload: device-to-device bearer control architecture for cellular data offloading," *IEEE Vehicular Technology Magazine*, vol. 8, no. 1, pp. 31–39, 2013.
- [5] 3GPP, "Report on the Design and Evaluation of the Tuak Algorithm Set: A Second Example Algorithm Set for the 3GPP Authentication and Key Generation Functions," *3GPP TSG SSA, Tech. Rep. TS 35.934 V14.0.0*, 2017.
- [6] D. Jaliha, R. D. Koilpillai, P. Khawas et al., "A rapidly deployable disaster communications system for developing countries," in *2012 IEEE International Conference on Communications (ICC)*, pp. 6339–6343, Ottawa, ON, Canada, June 2012.
- [7] S. Y. Lien, C. C. Chien, F. M. Tseng, and T. C. Ho, "3GPP device-to-device communications for beyond 4G cellular networks," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 28–35, 2016.
- [8] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine*, vol. 7, no. 12, pp. 42–49, 2009.

- [9] K. Doppler, C.-H. Yu, C. B. Ribeiro, and P. Jänis, "Mode selection for device-to-device communication underlying an LTE-advanced network," in *2010 IEEE Wireless Communication and Networking Conference*, pp. 1–6, Sydney, NSW, Australia, April 2010.
- [10] C. H. Chang, H. L. Chao, C. L. Liu, and K. L. Huang, "A local IP access mechanism for VoIP service in LTE home eNodeB systems," in *2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pp. 1–6, Taichung, Taiwan, July 2013.
- [11] E. Frlan, *Direct communication wireless radio system*, vol. 6047178, 2000United States Patent, US, 2000.
- [12] M. F. Khan, M. I. Khan, and K. Raahemifar, "Local IP Access (LIPA) enabled 3G and 4G femtocell architectures," in *2011 24th Canadian Conference on Electrical and Computer Engineering(CCECE)*, pp. 001049–001053, Niagara Falls, ON, Canada, May 2011.
- [13] C. B. Sankaran, "Data offloading techniques in 3GPP Rel-10 networks: a tutorial," *IEEE Communications Magazine*, vol. 50, no. 6, pp. 46–53, 2012.
- [14] H. Tang and Z. Ding, "Mixed mode transmission and resource allocation for D2D communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 162–175, 2016.
- [15] C.-C. Lai and C.-M. Liu, "A mobility-aware approach for distributed data update on unstructured mobile P2P networks," *Journal of Parallel and Distributed Computing*, vol. 123, pp. 168–179, 2019.
- [16] C.-M. Liu and C.-C. Lai, "A group-based data-driven approach for data synchronization in unstructured mobile P2P systems," *Wireless Networks*, vol. 27, no. 7, pp. 2465–2482, 2018.
- [17] A. Orsino, A. Ometov, G. Fodor et al., "Effects of heterogeneous mobility on D2D- and drone-assisted mission-critical MTC in 5G," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 79–87, 2017.
- [18] X. Xu, Y. Zhang, Z. Sun, Y. Hong, and X. Tao, "Analytical modeling of mode selection for moving D2D-enabled cellular networks," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1203–1206, 2016.
- [19] 3GPP, "Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS," *3GPP, Tech. Rep. TR 30.03U V3.2.0*, 1998.
- [20] 3GPP, "Radio Frequency (RF) Requirements for LTE Pico Node B," *3GPP TSG RAN, Tech. Rep. TR 36.931 V10.0.0*, 2011.
- [21] 3GPP, "Study on LTE Device to Device Proximity Services," *3GPP TSG RAN, Tech. Rep. TR 36.843 V12.0.1*, 2014.
- [22] 3GPP, "Study on Channel Model for Frequencies from 0.5 to 100 GHz," *3GPP TSG RAN, Tech. Rep. TR 38.901 V16.1.0*, 2019.
- [23] B. V. R. Gorantla and N. B. Mehta, "Resource and computationally efficient subchannel allocation for D2D in multicell scenarios with partial and asymmetric CSI," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5806–5817, 2019.
- [24] W.-K. Lai, Y.-C. Wang, H.-C. Lin, and J.-W. Li, "Efficient resource allocation and power control for LTE-A D2D communication with pure D2D model," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3202–3216, 2020.

Research Article

Rotated Black Hole: A New Heuristic Optimization for Reducing Localization Error of WSN in 3D Terrain

Qing-Wei Chai¹ and Jerry Wangtao Zheng²

¹Qing-Wei Chai College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China

²Weinberg College of Arts and Sciences, Northwestern University, USA

Correspondence should be addressed to Jerry Wangtao Zheng; wangtaozheng2022@u.northwestern.edu

Received 28 June 2021; Revised 25 August 2021; Accepted 27 September 2021; Published 11 October 2021

Academic Editor: Chao-Yang Lee

Copyright © 2021 Qing-Wei Chai and Jerry Wangtao Zheng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wireless sensor network (WSN) attracts the attention of more and more researchers, and it is applied in more and more environment. The localization information is one of the most important information in WSN. This paper proposed a novel algorithm called the rotated black hole (RBH) algorithm, which introduces a rotated optimal path and greatly improves the global search ability of the original black hole (BH) algorithm. Then, the novel algorithm is applied in reducing the localization error of WSN in 3D terrain. CEC 2013 test suit is used to verify the performance of the novel algorithm, and the simulation results show that the novel algorithm has better search performance than other famous intelligence computing algorithms. The localization simulation experiment results reveal that the novel algorithm also has an excellent performance in solving practical problems. WSN localization 3D terrain intelligence computing rotated the black hole algorithm.

1. Introduction

Due to the maturity of microelectronics and wireless communication technology, a lot of funds and researchers are attracted by wireless sensor networks (WSNs). These technologies have promoted the development of cheap, ultra-small, multifunctional, and smart sensor nodes (SNs), which have the ability to communicate with each other through wireless media [1]. Due to size and cost limitations, SN has limited functions and low computing power. Depending on the application, a collection of hundreds or thousands of nodes can be deployed in the area of interest. These nodes can communicate with each other through wireless media and form a network called a WSN.

More and more fields can achieve better performance with the help of WSN, such as target tracking [2], military affairs, disaster management, environmental monitoring, and so forth. In order to better perceive the target in the above scene, the location information of the sensor node is indispensable; otherwise, the information obtained by the sensor node will have a low correlation with the target to be monitored.

In order to obtain the accurate location of the sensor node, a global positioning system (GPS) module needs to be installed to provide accurate location information. However, due to the high cost and high energy consumption, it is impossible to install a GPS module on any sensor node. Without a GPS module, how to know the location of a node has become a problem that plagues researchers. In recent years, many algorithms have been proposed to calculate the positions of all sensor nodes in WSN [3]. According to the different of localization mechanisms, these algorithms can be divided into two categories: range-based localization algorithms and range-free localization algorithms. In some localization algorithms, such as received signal strength indication (RSSI) [4], the time of arrival (ToA) [5, 6], and the angle of arrival (AoA) [7] of the signal, these algorithms use specific physical data to estimate the location of unknown nodes (not equipped with GPS modules), called distance-based positioning algorithm. Although the range-based positioning algorithm provides more accurate position information than the range-free positioning algorithm, it needs to be equipped with a specific module, which greatly

increases the cost and energy consumption. WiFi distance measurement is used to improve accurate of indoor localization [8].

On the other hand, some algorithms only need the connective information of WSN and can estimate the position of unknown nodes. These algorithms are called range-free localization algorithms, such as centroid algorithm [9], distance vector-hop (DV-Hop) algorithm [10, 11], and approximate point-in-triangulation test (APIT) algorithm [12].

In [13], anchor nodes are set at the border land of monitoring regions and improve the estimation of per hop distance. The enhanced PSO algorithm is proposed and used to enhance the localization accurate of WSN in [14]. The hop size of each anchor node is replaced by an average hop size of all anchor nodes in [15]. Chen et al. proposed a method which adds a reference anchor node on hop size that calculates formulate to reduce the error of hop size of each anchor node [16]. In [17], the weighted least square algorithm is introduced to reduce the localization error of DV-Hop. In [18], a model is introduced to analysis localization error of mobile Lidar.

The traditional 2D WSN has gradually been unable to meet the needs of the current society, such as smart city [19–22]. Some algorithms applied to 2D WSN must be extended to 3D WSN to adapt to new social needs. [23] proposes a method to optimize the deployment strategy of WSN in a 3D environment. Pan et al. proposed a new algorithm to optimize the coverage rate of WSN in 3D terrain [24], and a DV-Hop localization method applied in 3D terrain is proposed [25]. In [26], the authors utilize the GA algorithm to improve localization accurate of WSN in 3D terrain.

In order to find the optimal solution of engineering problems quickly and accurately, many optimization algorithms have been established, some of them are inspired by natural phenomena, and [27] has briefly introduced the swarm intelligence algorithm. For example, the most popular evolution-inspired technique is genetic algorithms (GA) [28, 29] that simulates the Darwinian evolution. Differential evolution (DE) is an enhanced GA algorithm with better performance than GA [30–32]. Other popular algorithms are evolution strategy (ES) [33], probability-based incremental learning (PBIL) [34], genetic programming (GP) [35], Phasmatodea population evolution (PPE) algorithm [36], and biogeography-based optimizer (BBO) [37]. The particle swarm optimization (PSO) algorithm was developed based on the swarm behavior, such as fish and bird schooling in nature [38, 39]. Whale optimization algorithm (WOA) [40, 41] is a metaheuristic optimization algorithm by mimicking the hunting behavior of humpback whales. Grey wolf optimization (GWO) algorithm can be regarded as an enhanced PSO and achieve better performance [42, 43]. The ant colony optimization (ACO) is inspired by social behavior of ants in an ant colony [44]. Bat algorithm (BA) was inspired by the echolocation behavior of bats [45]. Black hole (BH) algorithms are inspired by the black hole phenomenon [46]. The gravitational search algorithm (GSA) was constructed based on the law of gravity and the notion of mass interactions [47]. The novel algorithm called moth search algorithm (MSA) is proposed which combine Levy fly and

moth search [48]. In addition, some methods are proposed to improve the search ability of algorithms, [49–51] introduce the concept of compact, and significantly reduce the memory usage. The surrogate method utilizes the Kriging model to enhance the algorithm run speed [52–56].

There are three contributions in this paper: Firstly, this study improves the search path of the black hole algorithm to improve the speed of obtaining the optimal solution. Secondly, the novel algorithm proposed in this article gets a better balance between local search ability and global search ability. Thirdly, this paper further enhances the localization accurate of WSN in 3D terrain.

The rest of the paper is organized as follows: in Section 2, the knowledge related to DV-Hop and BH algorithm are briefly introduced. The novel algorithm is shown in Section 3, and Section 4 utilizes the novel algorithm to reduce the localization error of WSN in 3D terrain. CEC 2013 test suit and simulation experiment of sensor nodes deploring are used to test the performance of novel algorithm in Section 5. Conclusion is the content of Section 6.

2. Related Work

2.1. DV-Hop Localization Algorithm in 3D Terrain

2.1.1. Original DV-Hop. This section briefly introduces the original DV-Hop localization algorithm. In this localization algorithm, every node is stored in a table with three attributes which are X , Y , Hop. X and Y are the horizontal and vertical coordinates of the node, and Hop is the hop count of the node connect with other sensor nodes. In the first stage after the installation of the WSN, the hop count of all sensor nodes is initial to zero, and each anchor node (the node with the GPS module installed) broadcasts a message packet containing its localization information in the network. If a sensor node receives a message from another sensor node, the hop count between them would increase one. There may be multiple path links between any two nodes; so, the number of hops between them may have different values, but only the minimum number of hops between any two sensor nodes is reserved in DV-Hop. Secondly, the anchor node collects the number of hops and distance information between itself and other connected anchor nodes, thereby calculating the average length per hop of this anchor node. The calculating equation is presented in the following:

$$\text{Hopsize}_i = \frac{\sum_{j=1}^m \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{j=1}^m \text{Hop}_{i,j}}, \quad (1)$$

where the position of i th anchor node is (x_i, y_i) , and m is the number of all anchor nodes. The hop count between i th anchor node and j th anchor node is represented as $\text{Hop}_{i,j}$, and the average length of per hop of i th anchor node is Hopsize_i .

Each anchor node broadcasts its own Hopsize_i to the network by controlling flooding. The distance between

unknown node u and anchor node A can be calculated by Eq (2). At the same time, the unknown node u forwards the information to surrounding nodes, and the corresponding hop count is increased by 1.

$$\text{Distance}_{A,u} = \text{hop}_{A,u} \cdot \text{Hopsize}_A \quad (2)$$

Then, the location of the unknown node can be estimated by the trilateration localization method, and the localization error can be reduced by the least-squares method [10].

2.1.2. DV-Hop in 3D Terrain. The wireless signal is transmitted from the transmitter to the receiver through electromagnetic radiation. Different from deploying WSN in a 2D environment, in a 3D environment, electromagnetic signals will be absorbed or blocked by surrounding obstacles. The number of obstacles determines the fading and shadowing of the signal between receiver and transmitter [57]. To detect whether the communication of nodes is blocker by surrounding terrain, a line-of-sight (LOS) algorithm is introduced. In this article, we use the Bresenham LOS algorithm, because it has a faster calculation speed, requires fewer calculation points, and does not require interpolation calculations [58]. The sensor model in WSN is a mathematical formula that characterizes the connectivity of sensors as a function of distance and terrain obstacles. This paper adopts the straightforward binary sensing model to determine whether a node is connected to the WSN. If a sensor node is within the communication radius of another sensor node, no obstacles obstruct the communication signal between them. These two sensor nodes can communicate with each other and vice versa. In order to illustrate the signal propagation in 3D terrain more vividly, Figure 1 shows the details of the LOS algorithm.

In Figure 1, there are 7 nodes on uneven terrain, named $a, b, c, d, e, f,$ and g , which are represented by blue dots. As can be seen from the figure, nodes a and node b are connected by a green dashed line, indicating that they can communicate with each other, and this is because there is no obstacle between node a and node b that hinders signal propagation. Conversely, node b and node c are connected by a red dashed line, and they cannot communicate with each other. As there is a mountain peak between them, the signal propagation is hindered.

2.2. Black Hole Algorithm. The black hole (BH) algorithm is one of the population-based methods which is inspired by the black hole phenomenon. In this method, a population of candidate solutions to a given problem is generated and distributed randomly in the search space.

In the proposed BH algorithm, the evolution of the population is done by moving all the candidates towards the best candidate solution in each iteration as shown in Eq. (2), and the best candidate solution is named black hole. If the population finds a better candidate solution than black hole, this solution would replace the black hole.

$$X_i^{t+1} = X_i^t + \text{rand} \cdot (X_{BH}^t - X_i^t) \quad i = 1, 2, \dots, N, \quad (3)$$

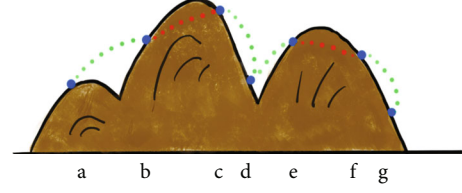


FIGURE 1: The signal propagation in 3D terrain.

where X_i^t and X_i^{t+1} are the locations of the i th star at iterations t and $t + 1$, respectively. X_{BH}^t is the location of the black hole in the search space at t th iteration. The rand is a random number between 0 and 1, and N is the number of stars (candidate solutions).

Otherwise, the black hole would engulf the candidate solution which within the radius of its event horizon. This mechanism guarantees that the population will not be trapped in local optimal value. In order to keep the population size, a new individual will be generated randomly in search space of a problem. It is worth noting that a small number of individuals will pass through the event horizon of the black hole, which can improve the exploitation performance of the algorithm. The radius of the event horizon in the BH algorithm is calculated by the following equation:

$$R = \frac{f_{BH}}{\sum_{i=1}^N f_i} \quad (4)$$

In the above formula, f_{BH} represents the fitness value calculated according to the position of the black hole. The fitness value calculated from the position of the i th individual is denoted by f_i . When the distance between an individual and the black hole is less than R , the individual will be swallowed, and the algorithm will randomly generate a new individual to maintain the population size.

3. Rotated Black Hole

In the original BH algorithm, the individual directly moves forward into the black hole. If the individual finds a better solution in the process, the black hole will be updated. However, in the multimodal problem, it is difficult for the intelligence computing algorithm to find a better solution on the straight path. [40] has introduced a formula which mimic the feeding behavior of whales and gotten great performance.

In order to improve the search ability of the BH algorithm in multimodal problems, the rotated model is introduced in this paper. The position update equation of population in the rotated model is presented in the following:

$$X_i^{t+1} = D \cdot e^{bl} \cdot \cos(2\pi l) + X_{BH}^t \quad (5)$$

X_i^{t+1} is the position of i th individual at $t + 1$ iteration, and the D indicates the distance between it and black hole. The position of the black hole at t iteration is represented by X_{BH}^t . The b is a constant, it is set to 1 in this article, and l is a random number between -1 and 1. In the novel algorithm, the original

```

Initialization:  $i = 1, t = 1, D = 50, n = 30, a = 0.7, T = 1000$ ;
randomly initial the position of population and calculate the function value:  $Pos, fun$ ;
 $fun\_BH = inf$ ;
while  $t \leq T$  do
     $Pos = (Pos_1, Pos_2, \dots, Pos_n), fun = (fun_1, fun_2, \dots, fun_n)$ ;
     $R\_BH = fun\_BH / sum(fun)$ ;
    while  $i \leq n$  do
        if  $rand < a$  then
            Update  $Pos_i$  according to Eq (3) and  $fun_i$ ;
        end
        else
            Update  $Pos_i$  according to Eq (5) and  $fun_i$ ;
        end
        if  $fun_i < fun\_BH$  then
             $Pos\_BH = Pos_i$ ;
             $fun\_BH = fun_i$ ;
        end
        if The distance between  $Pos_i$  and  $Pos\_BH$  smaller than  $R\_BH$  then
            Initial  $Pos_i$  and compute  $fun_i$ 
        end
         $i = i + 1$ ;
    end
     $t = t + 1$ ;
end

```

ALGORITHM 1: The rotated black hole algorithm.

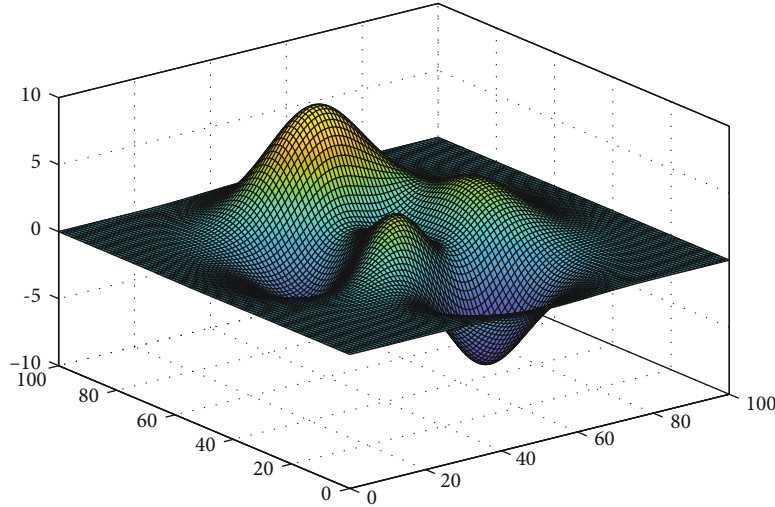


FIGURE 2: The flow chart of the novel algorithm.

position update way and the rotated model are combined to find the optimal solution. In each iteration, the algorithm generates a rand number, if the rand number is bigger than a special constant a , the algorithm utilizes a rotated model and vice versa. The detail of the novel algorithm is presented in Algorithm 1 and Figure 2.

4. Rotated Black Hole Algorithm Applied on DV-Hop in 3D Terrain

In this section, the novel algorithm is applied to reduce locate error of unknown nodes in 3D terrain. [59] has

proved that the intelligence computing algorithm has a distinct effect on WSN in 2D plane. In this paper, the 2D plane is extended to a 3D terrain as shown in Figure 3. This terrain is generated by the “peak” function of Matlab 2018a, and the connective state between any two nodes can be estimated by the LOS algorithm. The locate error of DV-Hop can be calculated by Eq. (6).

$$e = \left(\sum_{i=1}^m \left(\sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} - \text{dist}_i \right) \right)^2. \quad (6)$$

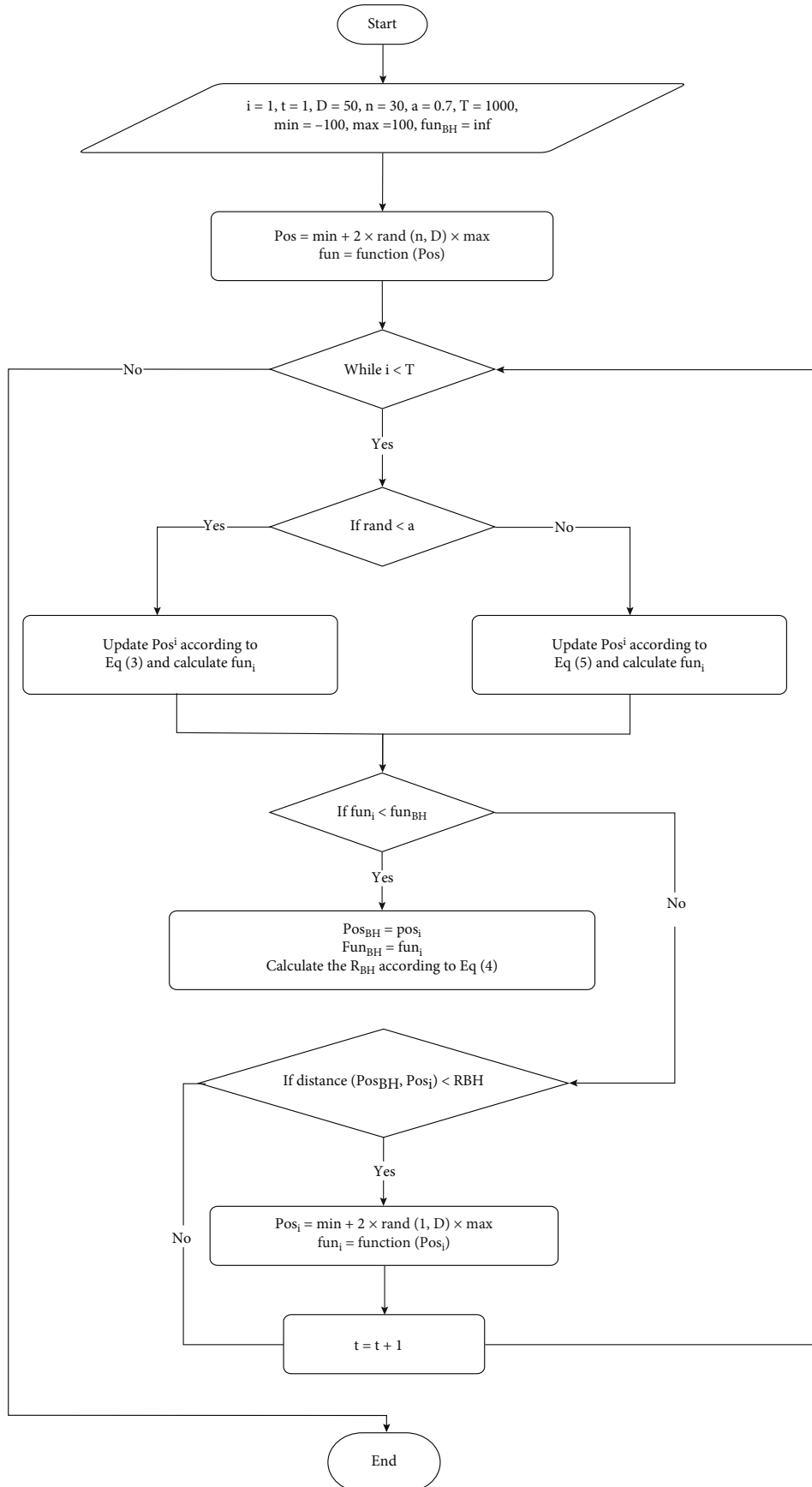


FIGURE 3: Terrain for deploying sensor nodes.

TABLE 1: Simulation results of CEC 2013 benchmark functions.

Algorithms	GA	WOA	PSO	ABC	MSA	BH	RBH
f_1	1.70E+05	3.25E+03	-1.02E+03	2.76E+04	1.81E+05	2.97E+04	6.74E+03
f_2	6.52E+09	2.04E+08	4.73E+07	8.19E+08	6.93E+09	3.58E+08	3.06E+08
f_3	2.73E+20	1.75E+11	5.80E+10	2.19E+15	4.10E+21	3.20E+12	1.15E+11
f_4	1.11E+06	1.01E+05	7.12E+04	1.01E+05	1.48E+07	9.48E+04	1.48E+05
f_5	1.04E+05	6.15E+02	-8.03E+02	1.05E+04	1.16E+05	2.09E+03	1.24E+03
f_6	3.56E+04	-2.14E+02	-7.04E+02	3.22E+03	3.59E+04	1.25E+03	-1.26E+02
f_7	7.89E+06	1.84E+03	-4.00E+02	1.94E+04	2.35E+07	1.70E+02	-5.60E+02
f_8	-6.79E+02	-6.79E+02	-6.79E+02	-6.79E+02	-6.78E+02	-6.79E+02	-6.79E+02
f_9	-5.18E+02	-5.27E+02	-5.27E+02	-5.28E+02	-5.12E+02	-5.28E+02	-5.33E+02
f_{10}	2.63E+04	8.64E+02	-2.58E+02	4.39E+03	2.98E+04	3.48E+03	1.46E+03
f_{11}	2.35E+03	4.65E+02	3.35E+02	3.76E+02	2.58E+03	4.41E+02	2.72E+02
f_{12}	2.26E+03	7.72E+02	4.96E+02	5.31E+02	2.35E+03	6.05E+02	4.77E+02
f_{13}	2.31E+03	9.01E+02	8.34E+02	6.37E+02	2.50E+03	7.35E+02	6.07E+02
f_{14}	1.73E+04	1.14E+04	1.08E+04	1.01E+04	1.77E+04	1.36E+04	9.91E+03
f_{15}	1.65E+04	1.30E+04	1.28E+04	1.48E+04	1.83E+04	1.39E+04	1.45E+04
f_{16}	2.05E+02	2.03E+02	2.03E+02	2.04E+02	2.09E+02	2.03E+02	2.04E+02
f_{17}	5.72E+03	1.51E+03	1.48E+03	1.31E+03	6.07E+03	1.36E+03	1.31E+03
f_{18}	5.82E+03	1.62E+03	1.55E+03	1.51E+03	6.03E+03	1.46E+03	1.44E+03
f_{19}	2.83E+07	2.15E+03	1.10E+03	5.49E+05	3.44E+07	4.74E+04	3.44E+03
f_{20}	6.25E+02	6.25E+02	6.25E+02	6.25E+02	6.25E+02	6.25E+02	6.24E+02
f_{21}	1.34E+04	3.84E+03	1.73E+03	5.26E+03	1.37E+04	4.53E+03	3.72E+03
f_{22}	1.93E+04	1.43E+04	1.43E+04	1.20E+04	1.94E+04	1.61E+04	1.30E+04
f_{23}	1.84E+04	1.52E+04	1.52E+04	1.67E+04	1.98E+04	1.65E+04	1.56E+04
f_{24}	2.15E+03	1.42E+03	1.53E+03	1.40E+03	2.36E+03	1.58E+03	1.40E+03
f_{25}	1.80E+03	1.55E+03	1.67E+03	1.52E+03	1.82E+03	1.68E+03	1.52E+03
f_{26}	1.84E+03	1.67E+03	1.67E+03	1.67E+03	1.90E+03	1.67E+03	1.67E+03
f_{27}	5.16E+03	3.67E+03	4.03E+03	3.55E+03	5.30E+03	3.93E+03	3.46E+03
f_{28}	1.73E+04	1.04E+04	1.01E+04	9.09E+03	1.95E+04	9.85E+03	4.13E+03

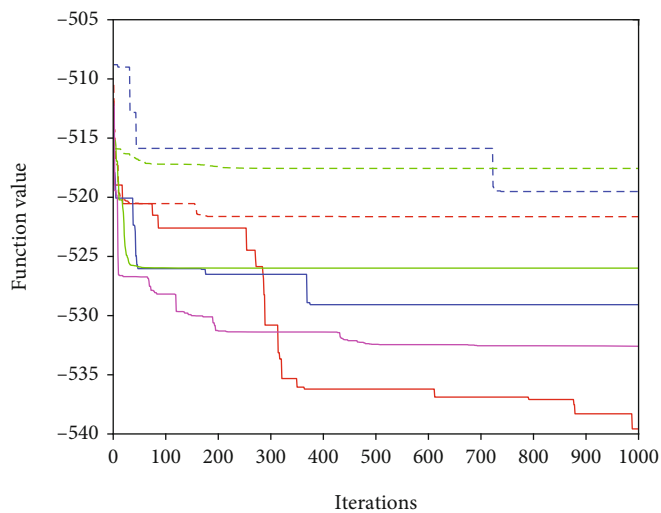
The position of the i th anchor node is represented by (x_i, y_i, z_i) , and (x, y, z) is the estimation position of unknown node. The $dist_i$ means the real distance between i th anchor node and unknown node; so, e is the error between the real position and estimation position of unknown node in WSN. This section applies the novel algorithm to reduce the error, and the fitness function is shown as follows:

$$f(x, y) = \min \left(\sum_{i=1}^m \left(\frac{1}{\text{hop}_{ii}} \right)^2 \left(\sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} - \text{dist}_i \right)^2 \right), \quad (7)$$

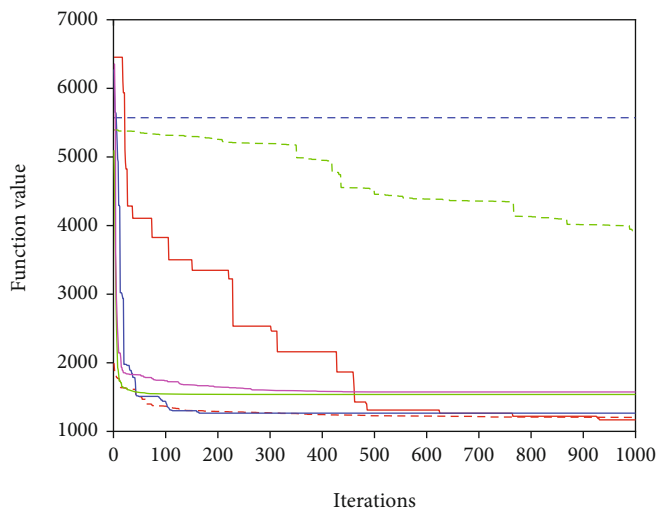
where the hop count between unknown node and i th anchor node is represented by hop_{ii} . The purpose of this article is to use the new algorithm to weigh trade-off all anchor nodes and find a suitable unknown node location.

5. Results and Discussion

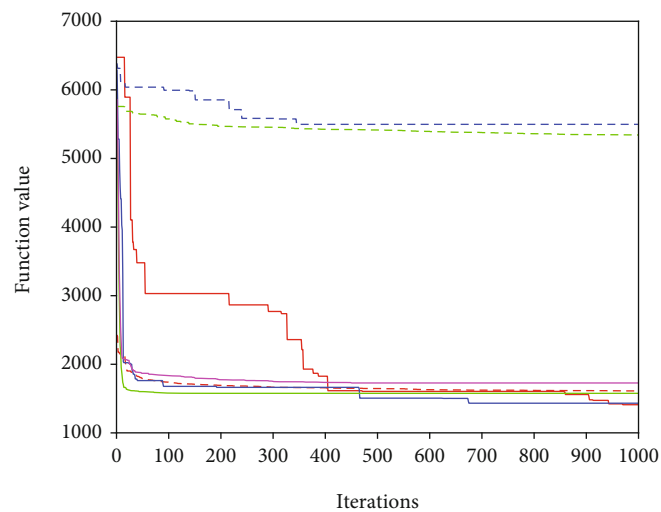
5.1. CEC 2013 Simulation Results. To test the performance of the RBH algorithm, CREST was carried out using the CEC 2013 benchmark function, which is a convincing function for the testing optimization algorithm. There are 28 test functions in the CEC 2013 benchmark function. The f_1 to f_5 are unimodal functions, mainly check out the convergence rate of the optimization algorithm. The f_6 to f_{20} are multimodal functions, which are used to verify the performance of avoiding local optimal values of algorithms. The f_{21} to f_{28} are composition functions, and their simulation results reveal the comprehensive performance of optimization algorithms. CREST was compared with some frequently used optimization algorithms, such as WOA, GA, BH, PSO, ABC, and MSA algorithms. CREST in this study is implemented on the same notebook computer which equips with an i5-7300HQ CPU @2.5GHz. CREST results were



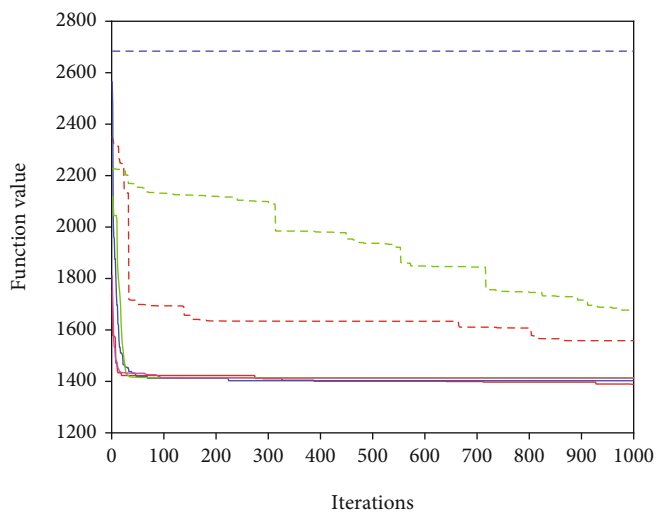
(a) f_9



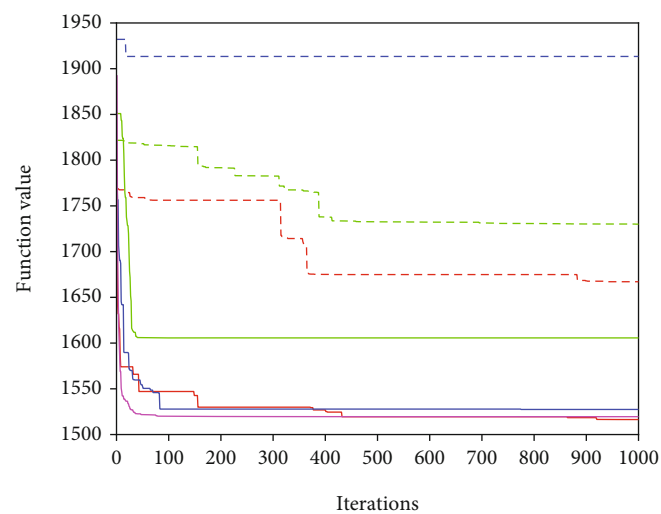
(b) f_{17}



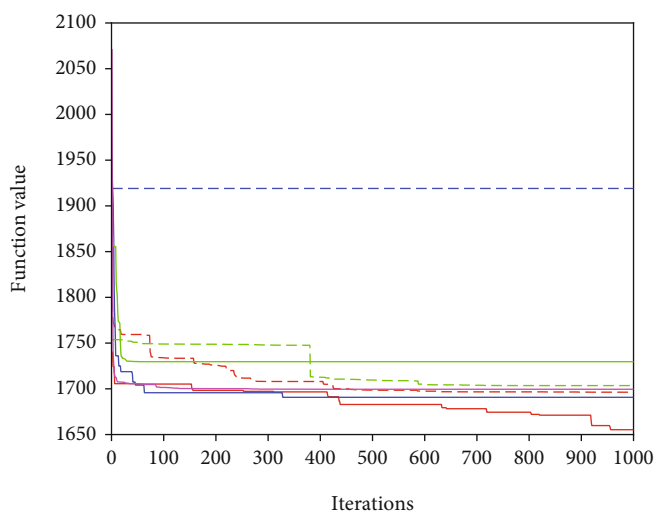
(c) f_{18}



(d) f_{24}



(e) f_{25}



(f) f_{26}

FIGURE 4: Continued.

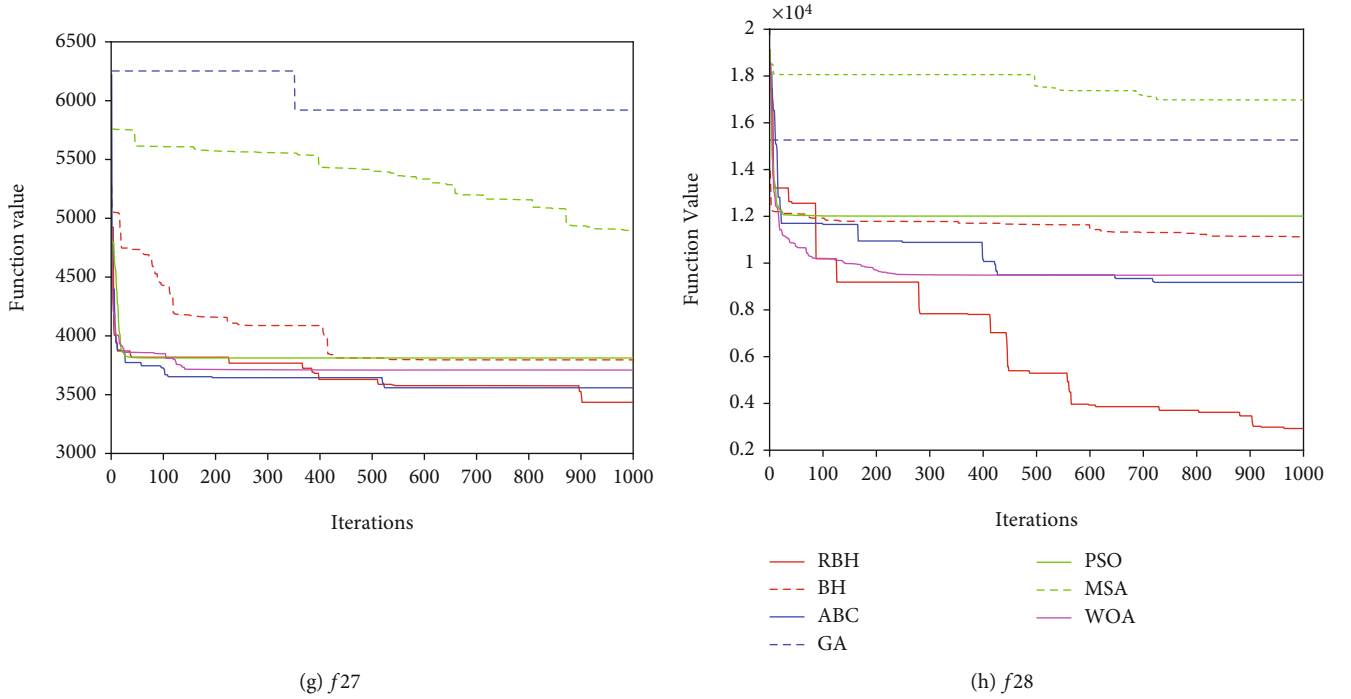


FIGURE 4: Simulation results of CEC 2013.

TABLE 2: Localization error of WSN with different numbers of nodes in 3D terrain.

Algorithm	Node number						
	200	250	300	350	400	450	500
DV-Hop	27.12 m	53.94 m	44.20 m	27.12 m	29.22 m	24.63 m	28.53 m
GA	7.29 m	6.94 m	8.00 m	7.27 m	7.01 m	7.33 m	7.99 m
WOA	6.14 m	5.35 m	6.33 m	5.19 m	5.43 m	5.68 m	6.15 m
BH	6.17 m	5.45 m	6.26 m	5.01 m	5.45 m	5.85 m	5.90 m
RBH	5.94 m	5.23 m	6.22 m	5.00 m	5.30 m	5.59 m	5.90 m
PSO	7.88 m	8.34 m	7.74 m	7.01 m	6.99 m	7.01 m	7.62 m
ABC	16.47 m	17.86 m	16.15 m	23.14 m	26.67 m	24.73 m	24.83 m
MSA	10.88 m	10.63 m	10.15 m	9.69 m	10.02 m	10.23 m	9.88 m

presented in Table 1, every data is the average of 48 test results, and the best results are marked by underline and bold font for each function.

The optimal results of intelligence computing algorithms under the CEC 2013 were showed in Table 1, and the best result of each test problem is marked in bold. The search ability of RBH algorithm on 14 benchmark functions is better or equal than other algorithms.

In the unimodal problems, as the excellent local search ability of PSO, other algorithms are far from the PSO algorithm. In the multimodal problems, the new algorithm proposed in this paper has obtained 8 best results. The results show that the new algorithm has a strong global search ability than other comparison algorithms. Out of a total of eight composition problems, the new algorithm performed best in five, and this further verifies the novel algorithm with great global search ability. Through CREST, one conclusion was

achieved that the RBH algorithm has good global search capabilities and is good at solving complex optimization problems.

In Figure 4, the details of intelligence computing algorithms to solve some optimal problems are given. Although the new algorithm cannot find a better solution than other algorithms in the early stage, it has a strong global search ability and is easy to avoid the local optimal value and finally find the optimal solution. Contrast other algorithms, the PSO has the fastest convergence rate but it has difficult jumping local optimal value; so, it is surpassed by other algorithms in the last stage. It can be seen that for complex problems, the new algorithm has an advantage due to its excellent global search ability.

5.2. Localization of WSN in 3D Terrain. In this section, the sensor nodes are randomly arranged in a $100\text{ m} \times 100\text{ m}$

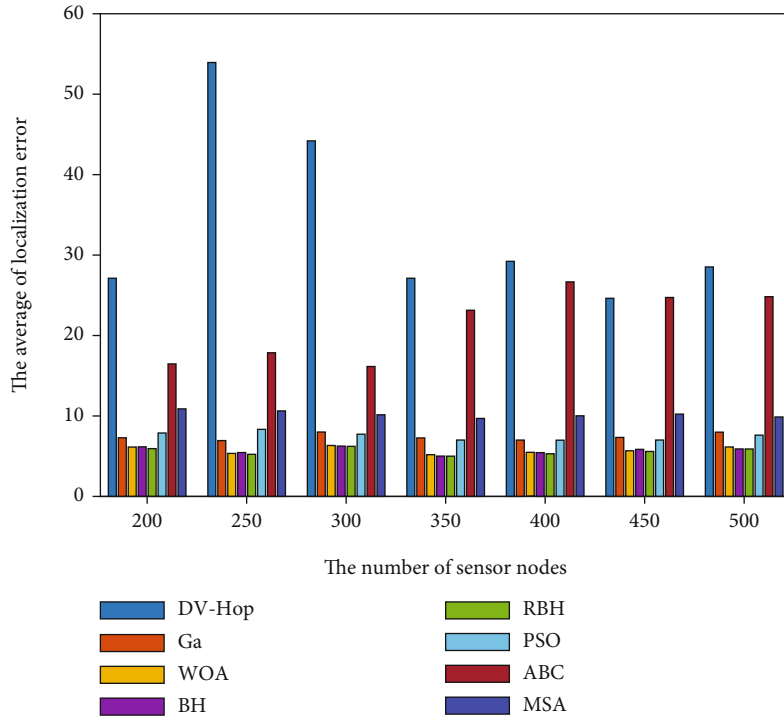


FIGURE 5: Localization error of WSN with different numbers of nodes in 3D terrain.

TABLE 3: Localization error of WSN with different numbers of anchor nodes in 3D terrain.

Algorithm	Anchor node number						
	15	20	25	30	35	40	45
DV-Hop	59.32 m	49.05 m	50.04 m	67.02 m	74.02 m	83.01 m	83.06 m
GA	7.10 m	8.05 m	7.72 m	8.88 m	10.31 m	11.40 m	13.14 m
WOA	6.15 m	6.47 m	6.64 m	7.77 m	9.16 m	10.86 m	12.25 m
BH	6.05 m	6.40 m	6.64 m	7.74 m	9.31 m	10.72 m	12.21 m
RBH	6.04 m	6.24 m	6.42 m	7.68 m	9.20 m	10.65 m	12.04 m
PSO	7.85 m	8.48 m	8.05 m	8.68 m	10.44 m	11.48 m	12.97 m
ABC	16.23 m	16.55 m	19.91 m	23.07 m	23.80 m	25.78 m	27.99 m
MSA	10.40 m	10.88 m	10.58 m	11.50 m	13.29 m	13.13 m	15.13 m

3D terrain as shown in Figure 3. In order to comprehensively check the performance of all positioning algorithms, we design three simulation experiments with different number of nodes, different number of anchor nodes, and communication radius of sensor nodes. The best experimental results for each experiment are underlined.

5.2.1. Localization Error of WSN with Different Numbers of Nodes in 3D Terrain. In this simulation experiment, we have configured 200, 250, 300, 350, 400, 450, and 500 nodes in the 3D terrain, respectively, the communication radius of all nodes is set to 20 m, and there are 30 anchor nodes. The simulation results are shown in Table 2 and Figure 5.

The localization error is the average value of localization error of all unknown nodes which is calculated by Eq. (6). From this table, we can see that the intelligence computing algorithm effectively improves the accuracy of the DV-Hop

localization algorithm, and the novel algorithm has the strongest performance in all participated compared algorithms. The biggest localization error of the original DV-Hop algorithm appears 250 nodes. Generally speaking, the positioning error is positively related to the number of nodes. The reason for this phenomenon is that each node deployment is independent and random. In the experiment of deploying 250 nodes, the signals of many nodes are blocked by the terrain; so, the experimental results of the original DV-Hop are not ideal. However, in this experimental environment, intelligent computing shows strong optimization capabilities, which greatly reduces the positioning error.

5.2.2. Localization Error of WSN with Different Numbers of Anchor Nodes in 3D Terrain. In this simulation experiment, there are 15, 20, 25, 30, 35, 40, and 45 anchor nodes,

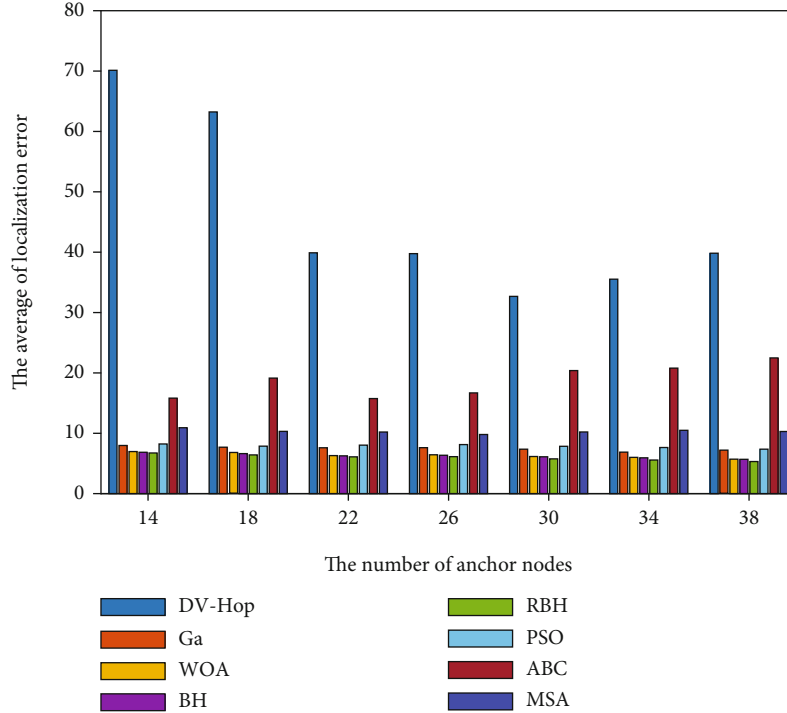


FIGURE 6: Localization error of WSN with different numbers of anchor nodes in 3D terrain.

TABLE 4: Localization error of WSN with different communication radii in 3D terrain.

Algorithm	Communication radius						
	14 m	18 m	22 m	26 m	30 m	34 m	38 m
DV-hop	59.32 m	49.05 m	50.04 m	67.02 m	74.02 m	83.01 m	83.06 m
GA	7.10 m	8.05 m	7.72 m	8.88 m	10.31 m	11.40 m	13.14 m
WOA	6.15 m	6.47 m	6.64 m	7.77 m	9.16 m	10.86 m	12.25 m
BH	6.05 m	6.40 m	6.64 m	7.74 m	9.31 m	10.72 m	12.21 m
RBH	6.04 m	6.24 m	6.42 m	7.68 m	9.20 m	10.65 m	12.04 m
PSO	7.85 m	8.48 m	8.05 m	8.68 m	10.44 m	11.48 m	12.97 m
ABC	16.23 m	16.55 m	19.91 m	23.07 m	23.80 m	25.78 m	27.99 m
MSA	10.40 m	10.88 m	10.58 m	11.50 m	13.29 m	13.13 m	15.13 m

respectively, the number of nodes is 300, and the communication radius of sensor node is set to 20 m. Table 3 and Figure 6 show the simulation results.

Generally, the positioning error decreases as the number of anchor nodes increases, but when the number of anchor nodes is greater than 35, a special situation occurs in the traditional DV-Hop. Excessive location information causes trouble to the traditional DV-Hop and reduces the accuracy of positioning. DV-Hop combined with intelligent computing algorithm, especially combined with RBH algorithm, can adapt to any conditions, and no matter the number of anchor nodes, the simulation results are equally excellent.

5.2.3. *Localization Error of WSN with Different Communication Radii in 3D Terrain.* The communication

radius of the sensor node can determine the anchor node that provides location information to how many sensor nodes. So, the value of communication radius of sensor nodes is an important factor for simulation results. We install the communication radius as 14, 18, 22, 26, 30, 34, and 38 to test the performance of different localization algorithms. There are 300 sensor nodes and 30 anchor nodes in this experiment, and the simulation results are presented in Table 4 and Figure 7.

From this table, we can see the influence of the communication radius on the experimental results. Different from what we originally envisioned, the increase of communication radius would increase the localization error of WSN. And without exception, all localization algorithms show this feature. The novel algorithm gets the best result at each experiment except communication radius is 30.

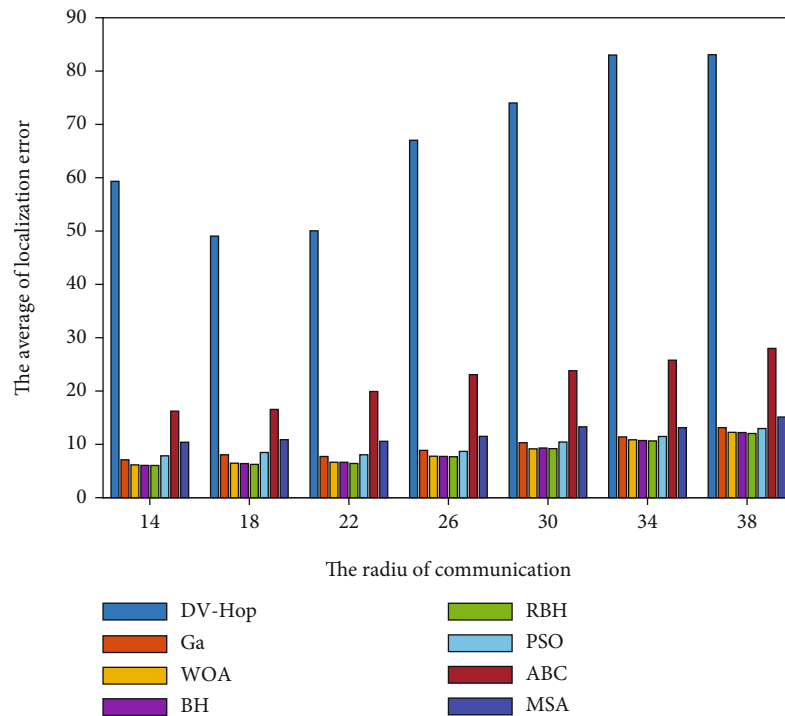


FIGURE 7: Localization error of WSN with different communication radii in 3D terrain.

6. Conclusion

In this paper, a novel algorithm is proposed, which utilizes a rotated optimal path to greatly improve the optimal ability of the original BH algorithm, especially in multimodal problems and composition problems. The CEC 2013 test suits show that the novel algorithm has outstanding global search ability; so, the novel algorithm can easily avoid local optimal value. In addition, this paper uses a new algorithm to optimize the position of the unknown node based on the DV-Hop positioning method, thereby reducing the error of the position estimation of the unknown node. The experiments in this paper have performed simulation experiments on the number of nodes, the number of anchor nodes, and the communication radius, respectively, and all prove that the new algorithm has a good effect on the localization problem of WSN in 3D terrain.

Data Availability

The related source code is uploaded in the following URL: https://mp.csdn.net/mp_download/manage/download/UpDetailed, which is available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] I. F. Akyildiz, Weilian Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.
- [2] S.-M. Lee, H. Cha, and R. Ha, "Energy-aware location error handling for object tracking applications in wireless sensor networks," *Computer Communications*, vol. 30, no. 7, pp. 1443–1450, 2007.
- [3] F.-C. Chang and H.-C. Huang, "A survey on intelligent sensor network and its applications," *Journal of Network Intelligence*, vol. 1, no. 1, pp. 1–15, 2016.
- [4] A. Awad, T. Frunzke, and F. Dressler, "Adaptive distance estimation and localization in wsn using rssi measures," in *10th Euromicro Conference on Digital System Design Architectures, Methods and Tools (DSD 2007)*, pp. 471–478, Lubeck, Germany, 2007.
- [5] I. Guvenc and C.-C. Chong, "A survey on toa based wireless localization and nlos mitigation techniques," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 3, pp. 107–124, 2009.
- [6] E. Xu, Z. Ding, and S. Dasgupta, "Source localization in wireless sensor networks from signal time-of-arrival measurements," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2887–2897, 2011.
- [7] R. Peng and M. L. Sichertiu, "Angle of arrival localization for wireless sensor networks," in *2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, pp. 374–382, Reston, VA, USA, 2006.
- [8] Z. Tian, W. Yang, Y. Jin, L. Xie, and Z. Huang, "MFPL: multi-frequency phase difference combination based device-free localization," *Computers, Materials & Continua*, vol. 62, no. 2, pp. 861–876, 2020.

- [9] J. Wang, P. Urriza, Y. Han, and D. Cabric, "Weighted centroid localization algorithm: theoretical analysis and distributed implementation," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3403–3413, 2011.
- [10] D. Niculescu and B. Nath, "Dv based positioning in ad hoc networks," *Telecommunication Systems*, vol. 22, no. 1/4, pp. 267–280, 2003.
- [11] W.-W. Ji and Z. Liu, "An improvement of dv-hop algorithm in wireless sensor networks," in *2006 International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4, Wuhan, China, 2006.
- [12] J. Z. Wang and H. Jin, "Improvement on apit localization algorithms for wireless sensor networks," in *2009 International Conference on Networks Security, Wireless Communications and Trusted Computing*, pp. 719–723, Wuhan, China, 2009.
- [13] X. Chen and B. Zhang, "Improved dv-hop node localization algorithm in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 8, no. 8, Article ID 213980, 2012.
- [14] B. F. Gumaida and J. Luo, "A hybrid particle swarm optimization with a variable neighborhood search for the localization enhancement in wireless sensor networks," *Applied Intelligence*, vol. 49, no. 10, pp. 3539–3557, 2019.
- [15] Y. Chen, X. Li, Y. Ding, J. Xu, and Z. Liu, "An improved dvhop localization algorithm for wireless sensor networks," in *2018 13th IEEE conference on industrial electronics and applications (ICIEA)*, pp. 1831–1836, Wuhan, China, 2018.
- [16] H. Chen, K. Sezaki, P. Deng, and H. C. So, "An improved dv-hop localization algorithm with reduced node location error for wireless sensor networks," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E91-A, no. 8, pp. 2232–2236, 2008.
- [17] S. Kumar and D. K. Lobiyal, "An advanced dv-hop localization algorithm for wireless sensor networks," *Wireless Personal Communications*, vol. 71, no. 2, pp. 1365–1385, 2013.
- [18] C. Peng and Y. Z. Zhenglin, "Modeling analysis for positioning error of mobile lidar based on multiu2010 body system kinematics," *Intelligent Automation and Soft Computing*, vol. 25, no. 4, 2019.
- [19] B. Tang, Z. Chen, G. Hefferman et al., "Incorporating intelligence in fog computing for big data analysis in smart cities," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2140–2150, 2017.
- [20] C.-W. Tsai, P.-W. Tsai, J.-S. Pan, and H.-C. Chao, "Metaheuristics for the deployment problem of wsn: a review," *Microprocessors and Microsystems*, vol. 39, no. 8, pp. 1305–1317, 2015.
- [21] Z. Lin, X. Chen, Y. Hao, C. Lv, and L. Yu, *A new three dimensional assessment model and optimization for acoustic positioning system*, 2018.
- [22] H. Xing, Y. Zhao, Y. Zhang, and Y. Chen, "3d trajectory planning of positioning error correction based on PSO-A* algorithm," *Computers, Materials & Continua*, vol. 65, no. 3, pp. 2295–2308, 2020.
- [23] B. Cao, J. Zhao, P. Yang, P. Yang, X. Liu, and Y. Zhang, "3-D deployment optimization for heterogeneous wireless directional sensor networks on smart city," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1798–1808, 2019.
- [24] J.-S. Pan, Q.-W. Chai, S.-C. Chu, and N. Wu, "3-D terrain node coverage of wireless sensor network using enhanced black hole algorithm," *Sensors*, vol. 20, no. 8, 2020.
- [25] Q.-W. Chai, S.-C. Chu, J.-S. Pan, and W.-M. Zheng, "Applying adaptive and self assessment fish migration optimization on localization of wireless sensor network on 3-D terrain," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 11, no. 2, pp. 90–102, 2020.
- [26] G. Sharma and A. Kumar, "Improved range-free localization for three-dimensional wireless sensor networks using genetic algorithm," *Computers & Electrical Engineering*, vol. 72, pp. 808–827, 2018.
- [27] S.-C. Chu, H.-C. Huang, J. F. Roddick, and J.-S. Pan, "Overview of algorithms for swarm intelligence," in *International Conference on Computational Collective Intelligence*, pp. 28–41, Gdynia, Poland, 2011.
- [28] D. Michael, Vose. *The simple genetic algorithm: foundations and theory*, MIT press, 1999.
- [29] H. John, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–72, 1992.
- [30] S. Das and P. N. Suganthan, "Differential evolution: A survey of the state-of-the-art," *IEEE Transactions On Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.
- [31] R. Gämperle, S. D. Müller, and P. Koumoutsakos, "A parameter study for differential evolution," *Advances in Intelligent Systems, Fuzzy Systems, Evolutionary Computation*, vol. 10, no. 10, pp. 293–298, 2002.
- [32] Z. Meng, J.-S. Pan, and L. Kong, "Parameters with adaptive learning mechanism (palm) for the enhancement of differential evolution," *Knowledge-Based Systems*, vol. 141, pp. 92–112, 2018.
- [33] I. Rechenberg, "Evolutionstrategien," in *Simulationsmethoden in der Medizin und Biologie*, Springer, 1978.
- [34] D. Dasgupta and Z. Michalewicz, *Evolutionary Algorithms in Engineering Applications*, Springer Science & Business Media, 2013.
- [35] J. R. Koza and R. Poli, "Genetic programming," in *Search Methodologies*, Springer, 2005.
- [36] P.-C. Song, S.-C. Chu, J.-S. Pan, and H. Yang, "Phasmatodea population evolution algorithm and its application in length-changeable incremental extreme learning machine," in *2020 2nd International Conference on Industrial Artificial Intelligence (IAI)*, pp. 1–5, Shenyang, China, 2020.
- [37] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.
- [38] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, pp. 1942–1948, Perth, WA, Australia, 1995.
- [39] C. Yang, T. Xuyan, and J. Chen, "Algorithm of marriage in honey bees optimization based on the wolf pack search," in *The 2007 International Conference on Intelligent Pervasive Computing (IPC 2007)*, Jeju, Korea (South), 2007.
- [40] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
- [41] T.-K. Dao, T.-S. Pan, and J.-S. Pan, "A multi-objective optimal mobile robot path planning based on whale optimization algorithm," in *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, pp. 337–342, Chengdu, China, 2016.
- [42] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

- [43] P. Hu, J.-S. Pan, S.-C. Chu, Q.-W. Chai, T. Liu, and Z.-C. Li, "New hybrid algorithms for prediction of daily load of power network," *Applied Sciences*, vol. 9, no. 21, 2019.
- [44] M. Dorigo and C. Blum, "Ant colony optimization theory: a survey," *Theoretical Computer Science*, vol. 344, no. 2-3, pp. 243–278, 2005.
- [45] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature inspired cooperative strategies for optimization (NICSO 2010)*, Springer, 2010.
- [46] A. Hatamlou, "Black hole: a new heuristic optimization approach for data clustering," *Information Sciences*, vol. 222, pp. 175–184, 2013.
- [47] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," *Information Sciences*, vol. 179, no. 13, pp. 2232–2248, 2009.
- [48] G.-G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memeetic Computing*, vol. 10, no. 2, pp. 151–164, 2018.
- [49] A.-Q. Tian, S.-C. Chu, J.-S. Pan, H. Cui, and W.-M. Zheng, "A compact pigeon-inspired optimization for maximum short-term generation mode in cascade hydroelectric power station," *Sustainability*, vol. 12, no. 3, 2020.
- [50] X. Xue and J.-S. Pan, "A compact co-evolutionary algorithm for sensor ontology meta-matching," *Knowledge and Information Systems*, vol. 56, no. 2, pp. 335–353, 2018.
- [51] G. R. Harik, F. G. Lobo, and D. E. Goldberg, "The compact genetic algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 287–297, 1999.
- [52] C. Sun, Y. Jin, R. Cheng, J. Ding, and J. Zeng, "Surrogate-assisted cooperative swarm optimization of high-dimensional expensive problems," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 644–660, 2017.
- [53] H. Wang, L. Feng, Y. Jin, and J. Doherty, "Surrogate-assisted evolutionary multitasking for expensive minimax optimization in multiple scenarios," *IEEE Computational Intelligence Magazine*, vol. 16, no. 1, pp. 34–48, 2021.
- [54] P. Liao, C. Sun, G. Zhang, and Y. Jin, "Multi-surrogate multitasking optimization of expensive problems," *Knowledge-Based Systems*, vol. 205, article 106262, 2020.
- [55] F. Guoxia, C. Sun, Y. Tan, G. Zhang, and Y. Jin, "A surrogate-assisted evolutionary algorithm with random feature selection for large-scale expensive problems," in *International Conference on Parallel Problem Solving from Nature*, pp. 125–139, Leiden, The Netherlands, 2020.
- [56] B. Van Stein, H. Wang, W. Kowalczyk, M. Emmerich, and T. Bäck, "Cluster-based kriging approximation algorithms for complexity reduction," *Applied Intelligence*, vol. 50, no. 3, pp. 778–791, 2020.
- [57] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge university press, 2005.
- [58] H. R. Topcuoglu, M. Ermis, and M. Sifyan, "Positioning and utilizing sensors on a 3-D terrain part I—theory and modeling," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 3, pp. 376–382, 2011.
- [59] Q.-w. Chai, S.-C. Chu, J.-S. Pan, P. Hu, and W.-m. Zheng, "A parallel WOA with two communication strategies applied in dv-hop localization method," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, 10 pages, 2020.