Wireless Communications and Mobile Computing

Cloud Computing and IoT for Intelligent Applications

Lead Guest Editor: Maode Ma
Guest Editors: Ireneeusz Czarnowski and Shouyong Jiang

Cloud Computing and IoT for Intelligent Applications

Wireless Communications and Mobile Computing

Cloud Computing and IoT for Intelligent Applications

Lead Guest Editor: Maode Ma
Guest Editors: Ireneeusz Czarnowski and Shouyong Jiang

Chief Editor

Associate Editors

Academic Editors

Contents

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3833034, 6 pages
https://doi.org/10.1155/2022/3833034

Research Article

Evaluation of Perioperative Nursing for Patients with Hepatobiliary Pancreatic Diseases Combined with Diabetes Mellitus Based on Multilayer Perceptron Neural Network

Mengmeng Zhang [1] and Meng Yang [2]

[1] *Department of Hepatobiliary Surgery, The First People's Hospital of Lianyungang, Jiangsu 222000, China*
[2] *Department of Digestive Internal Medicine, The First People's Hospital of Lianyungang, Jiangsu 222000, China*

Correspondence should be addressed to Mengmeng Zhang; mengmengzhang1988@outlook.com

Received 29 April 2022; Revised 17 May 2022; Accepted 27 May 2022; Published 21 September 2022

Academic Editor: Maode Ma

Objective. To analyze the perioperative nursing methods for patients with hepatobiliary pancreatic diseases combined with diabetes mellitus and to evaluate the differences in nursing methods based on a multilayer perceptron (MLP) neural network. *Methods*. 80 patients with hepatobiliary and pancreatic diseases complicated with diabetes admitted to our hospital from January 2021 to January 2022 were selected as subjects. According to different nursing methods, two groups of hepatobiliary and pancreatic diseases with diabetes were randomly divided into the control group and the experimental group, 40 patients in each group, two groups were given routine nursing care, and the experimental group was given perioperative nursing intervention on the basis of routine nursing. The fasting blood glucose (FBG), postprandial two-hour blood glucose (2hPG), and glycosylated hemoglobin (HbA1c) index of two groups of patients with hepatobiliary pancreatic diseases and diabetes mellitus were observed and compared. The incidence of postoperative complications, the average hospitalization time of patients, and the clinical nursing satisfaction rate of patients were observed and compared. *Results*. The blood glucose level including FBG, 2hPG, and HbA1c in the experimental group were better than those in the control group, and the data comparison difference was statistically significant ($P < 0.01$). The incidence of postoperative complications in the experimental group was lower than that in the control group, and the data comparison difference was statistically significant ($P < 0.05$). The discharge time of patients in the experimental group was significantly shorter than that in the control group, and the satisfaction rate was significantly higher than that in the control group ($P < 0.05$). The comprehensive analysis based on MLP neural network model confirmed that perioperative surgical nursing can improve the treatment effect of patients. *Conclusion*. In the perioperative period of patients with hepatobiliary and pancreatic diseases combined with diabetes, the implementation of comprehensive and comprehensive nursing intervention will help to improve the effect of surgical treatment and postoperative recovery.

1. Introduction

Diabetics usually associate with other diseases (hepatobiliary stones or pancreatic tumors) [1]. In recent years, the incidence rate of diabetes and liver, biliary, and pancreatic diseases is higher. Hepatobiliary pancreatic diseases mainly refer to diseases related to the gallbladder, liver, and pancreas in clinic. They are more common in surgery, and there are many kinds, which makes nursing difficult. The most common hepatobiliary and pancreatic surgical diseases include liver cancer, liver cirrhosis, cholecystitis, gallstones, pancreatitis, and pancreatic tumors. In the clinical treatment of hepatobiliary and pancreatic diseases, drug treatment or surgical treatment can be adopted. Drug therapy is mainly used as a conservative treatment to improve the clinical symptoms of patients and control the progress of the disease. However, the effect of therapeutic drugs is relatively limited and it is difficult to cure the disease fundamentally. The application of surgical methods can directly remove the lesion tissue, thereby achieving the purpose of radical disease cure. However, the surgery is more complex, difficult, and risky, so higher requirements are put forward for the way of perioperative nursing intervention. Despite advances in patient selection and surgical techniques, the incidence of perioperative complications after pancreatectomy is still high, about 30%-40% [2, 3]. Studies have shown that targeted care during the perioperative period of diabetic patients can improve the patient's psychological state and reduce the occurrence of complications [4, 5]. In this study, 80 patients with hepatobiliary and pancreatic diseases combined with diabetes mellitus who will be admitted to the hospital in January 2021 and January 2022 were the research objects. Combined with the comprehensive evaluation function of the neural network model, the perioperative nursing effect of patients with hepatobiliary and pancreatic diseases combined with diabetes mellitus was analyzed.

The multilayer perception (MLP) model is an important nonlinear multifactor evaluation method and one of the most widely used artificial neural network models [6]. Compared with the single-layer perceptron, which only has the ability to classify linear tasks, the multilayer perceptron artificial neural network has a complex structure and is suitable for solving practical problems. The MLP neural network model is an objective and comprehensive method for evaluating and comparing drug efficacy. Compared with single-layer perceptual neural network, MLP is more suitable for medical data analysis to solve complex linear and indivisible multiclassification problems. For example, MLP combined with artificial neural network (ANN) model can be used to evaluate the efficacy of drugs [7]. The MLP model can also be used for the diagnosis of glomerular filtration rate in patients with polycystic kidney disease and to help judge the progression of the disease [8]. In recent years, MLP has been widely used to predict liver cirrhosis, hypertension, and other diseases [9, 10]. Therefore, we constructed the MLP–ANN model to evaluate the importance of relevant parameters and the effect of perioperative nursing.

2. Materials and Methods

2.1. Research Object and General Information. A total of 80 patients with hepatobiliary and pancreatic diseases combined with diabetes were selected as the research subjects, including 33 patients with acute severe pancreatitis, 21 patients with hepatic cyst, 15 patients with cholelithiasis, and 11 patients with cholecystitis. This experiment has been approved and agreed by the Ethics Committee of the First People's Hospital of Lianyungang City, Jiangsu Province, and all patients have obtained their informed consent. According to different nursing methods, the two groups of patients with hepatobiliary and pancreatic diseases combined with diabetes were randomly divided into a control group and an experimental group, with 40 patients in each group. Inclusion criteria: all patients were confirmed as hepatobiliary and pancreatic diseases complicated with diabetes by preoperative examination, without other serious heart, liver, or renal insufficiency, and without surgical contraindications. If the patient is delirious, or the treatment does not cooperate, it is excluded.

In the control group, there were 23 male patients and 17 female patients with hepatobiliary and pancreatic diseases combined with diabetes, aged between 51 and 72 years old, with an average age of 62.41 ± 4.63 years old, and the disease course of the patients was 1 to 8 years (6.90 ± 2.06 years). Among them, there were 17 patients with acute severe pancreatitis, 10 patients with liver cyst, 7 patients with cholelithiasis, and 6 patients with cholecystitis. In the experimental group, there were 21 male patients and 19 female patients with hepatobiliary and pancreatic diseases combined with diabetes, aged between 49 and 70 years old, with an average age of 60.50 ± 4.83 years old, and the disease course of the patients was 1 to 9 years (6.16 ± 2.25 years). Among them, there were 16 patients with acute severe pancreatitis, 11 patients with liver cyst, 8 patients with cholelithiasis, and 5 patients with cholecystitis. There was no significant difference in age, gender, course of disease, and other clinical data between the two groups of patients with hepatobiliary and pancreatic diseases combined with diabetes mellitus ($P > 0.05$), which were not comparable (Table 1, $P > 0.05$).

2.2. Nursing Methods. The control group received routine nursing, measured and recorded the patients' blood pressure, heart rate, and blood oxygen saturation, and guided the patients to participate in various preoperative examination activities. The comprehensive group implemented comprehensive nursing, and the specific methods were as follows.

Preoperative nursing: (1) psychological nursing: introduce the value of surgical treatment and successful treatment cases to patients before operation. Actively communicate with patients, explain the general process of operation and matters needing attention for patients, and alleviate the anxiety of patients. (2) Diet nursing: pay attention to diet matching before operation. You can eat some foods with high fiber, low fat, and high protein.

Intraoperative nursing: during the operation, carefully check the surgical instruments and cooperate with doctors to carry out surgical treatment activities. Accompany patients throughout the process, guide them, and encourage them. The blood glucose level should be recorded every 4 h during the operation to ensure the orderly operation.

Postoperative care: (1) complication care: anti-infective treatment and cleaning care should be done after surgery, regular monitoring of the patient's body temperature, heart rate, etc., and an ECG monitor should be set at the bedside. If the patient has any adverse reactions, the clinician needs to be notified immediately. Pay attention to the condition of the drainage tube to avoid problems such as bending and compression of the drainage tube. Regularly replace the drainage bag for the patient, and guide the patient to turn over once every 2 h to avoid the occurrence of pressure ulcer. (2) Diet nursing: take liquid food three days after operation. With the improvement of the patient's physical state, you can eat semiliquid food, gradually change the diet, and pay attention to the nutritional matching.

2.3. Observation Indicators. The indexes of FBG, 2hPG, and HbA1c in the two groups, the incidence of postoperative

Table 1: Comparison of general data of patients.

	Experimental group ($n = 40$)	Control group ($n = 40$)	T/χ^2	P
Gender (n, %)			0.2020	0.6531
Man	23 (57.5%)	21 (52.5%)		
Female	17 (42.5%)	19 (47.5%)		
Age ($\bar{x} \pm s$)	62.41 ± 4.63	60.50 ± 4.83	3.2645	0.2829
Disease course ($\bar{x} \pm s$)	6.90 ± 2.06	6.16 ± 2.25	2.3781	0.3784
Complication (n, %)			0.2355	0.9717
Severe acute pancreatitis	17 (42.5%)	16 (40%)		
Liver cyst	10 (25%)	11 (27.5%)		
Cholelithiasis	7 (17.5%)	8 (20%)		
Cholecystitis	6 (15%)	5 (12.5%)		

Table 2: Comparison of blood glucose indexes in patients ($\bar{x} \pm s$).

	Case	FBG (mmol/L)	2hPG (mmol/L)	HbA1c (%)
Experimental group	40	6.96 ± 0.70	8.59 ± 0.87	6.80 ± 0.71
Control group	40	8.37 ± 0.75	11.63 ± 1.06	8.31 ± 0.79
t		0.2649	0.6348	0.2844
P		<0.01	<0.01	<0.01

Table 3: Comparison of average length of stay of patients ($\bar{x} \pm s$).

	Case	Average length of stay (days)
Experimental group	40	5.40 ± 0.89
Control group	40	8.63 ± 0.69
t		0.5724
P		<0.01

complications, the length of hospital stay, and the satisfaction rate of clinical nursing are the indicators.

2.4. Construction of MLP Neural Network. MLP is one of the most widely used artificial neural network models, which has good nonlinear system modeling ability [11]. The MLP neural network is composed of multilayer nonlinear neurons (also known as nodes), which is divided into three parts: the input layer composed of a group of source nodes, the implicit layer of one or more layers of computing nodes, and the output layer of one layer of computing nodes. The input vector enters each node of the first hidden layer, and then, the output of the first hidden layer node is sent to the neurons of the second hidden layer until the output. In order to establish an effective MLP model, 80% of the data was used as the training set to optimize the characteristics of the model, and then, use 20% of the data as the test set to evaluate the generalization ability. Finally, the MLP neural network model was established by blood glucose level, patient age and course of disease, satisfaction rate, complication rate, and other related indicators.

2.5. Statistical Analysis. SPSS 20.0 was used for analysis, $\bar{x} \pm s$ refers to the measurement data, and the comparison between groups is subject to t -test; the count data was represented by %, and the comparison between groups was performed by the χ^2 test. $P < 0.05$ was considered statistically significant. The comprehensive analysis adopts MLP neural network analysis.

3. Results

3.1. Comparison of Blood Glucose Indexes in Patients. All 80 patients underwent successful operations. The patients who received perioperative care (experimental group) had better blood glucose levels (FBG, 2hPG, and HbA1c) than those who received traditional care (control group), and the difference was statistically significant ($P < 0.01$); the results are shown in Table 2. It shows that giving perioperative care to patients with hepatobiliary and pancreatic diseases combined with diabetes can improve the therapeutic effect of patients.

3.2. Comparison of Average Length of Stay of Patients. After nursing, the average hospitalization time of hepatobiliary and pancreatic diseases with diabetes in the experimental group was 5.40 ± 0.89 d, and the average hospitalization time in the control group was 8.63 ± 0.69 d (Table 3, $P < 0.01$).

3.3. Comparison of Postoperative Complications. There were 2 cases of incision infection, 4 cases of pressure sore, and 3 cases of pulmonary infection in the control group, and the complication rate was 22.5%. There were 1 case of incision infection, 2 cases of pressure sore, and 0 cases of pulmonary infection in the experimental group, and the complication rate was 5%, which was significantly lower than that in the control group. The difference between the groups was statistically significant (Table 4, $P < 0.05$).

3.4. Comparison of Clinical Nursing Satisfaction Rate of Patients. The overall nursing satisfaction rate was 80% in the control group and 95% in the experimental group (Table 5, $P < 0.05$). It shows that perioperative surgical nursing can improve the clinical nursing satisfaction rate of patients.

Table 4: Comparison of postoperative complications (n, %).

	Incision infection	Pressure sore	Pulmonary infection	Complication rate
Experimental group ($n = 40$)	1 (2.5%)	1 (2.5%)	0 (0)	2 (5%)
Control group ($n = 40$)	2 (5%)	4 (10%)	3 (7.5%)	9 (22.5%)
χ^2				5.1647
P				0.0231

Table 5: Comparison of clinical nursing satisfaction rate of patients (n, %).

	Satisfied	Basically satisfied	Dissatisfied	Satisfaction rate
Experimental group ($n = 40$)	31 (77.5%)	7 (17.5%)	2 (5%)	38 (95%)
Control group ($n = 40$)	10 (25%)	22 (55%)	8 (20%)	32 (80%)
χ^2				4.1143
P				0.0425

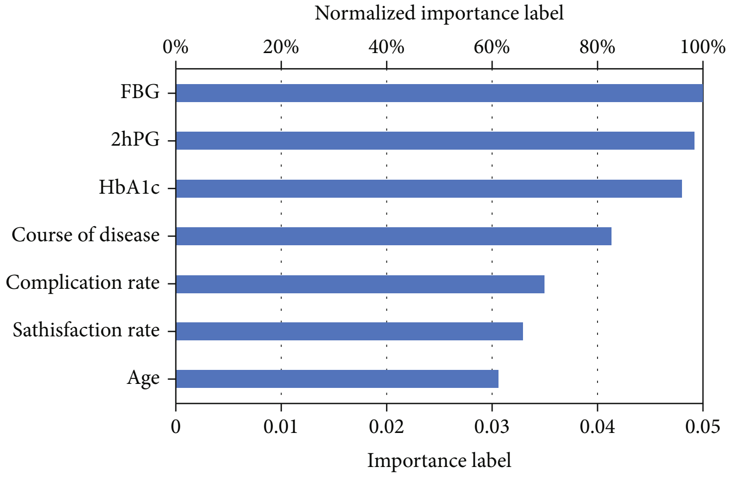


Figure 1: Importance analysis of MLP input variables.

3.5. Evaluation of Perioperative Nursing Effect Based on Neural Network Model. We use a three-layer neural MLP neural network model: taking the above factors as input layer, hidden layer, and output layer (average hospital stay). We evaluated and compared the comprehensive effects of perioperative nursing and traditional nursing according to MLP neural network model. The average hospital stay of perioperative nursing patients was 5.12 ± 0.68 days and that of traditional nursing patients was 9.24 ± 0.72 days, indicating that perioperative nursing can achieve better results ($P < 0.01$). According to the influence degree of input indexes on the network, FBG, 2hPG, and HbA1c are more important influencing factors in the neural network model (Figure 1).

4. Conclusion

Diabetes mellitus is a disease with a high clinical incidence, and hyperglycemia is its main clinical feature. In clinical treatment, the operation of hepatobiliary and pancreatic diseases is more complicated, and the symptoms of diabetic patients further increase the difficulty of surgical treatment. Therefore, it is particularly important to do a good job of preoperative, intraoperative, and postoperative nursing measures during the operation.

In this study, traditional nursing care was given to some patients with hepatobiliary and pancreatic diseases combined with diabetes, mainly including maintaining the balance of electrolytes and H_2O in the body, controlling the blood sugar level in the body, ensuring the life safety of the patients during the perioperative period, and avoiding the sudden occurrence of blood sugar in the patients. It is too high and has an impact on the operation, but it ignores the psychological impact of the disease on the patient. Due to the lack of awareness of the disease and the operation, the patient will be suspicious when receiving drug or surgical treatment, resulting in reduced compliance. Not only is it beneficial to surgical treatment, but it is also not conducive to subsequent nursing care. In addition, medical staff did not specifically control the blood glucose in the patient's body according to the patient's physical condition, resulting in poor nursing effect.

Perioperative nursing can control the blood glucose of patients, ensure the life safety of patients during perioperative period, and avoid the failure of operation caused by the sudden rise of blood glucose during operation. Psychological nursing for patients can make patients clearly know their condition and the importance of surgery, eliminate patients' unrealistic fantasy, improve patients' desire for survival, and make patients more cooperate with doctors and nurses in treatment and nursing. It can not only improve the psychological state of patients before operation but also improve the recovery of patients after operation [12]. However, the lack of knowledge about diabetes and the prevention and treatment of diabetes caused the poor compliance of diabetes treatment. The diabetes-related knowledge defect is relatively large in the elderly with special educational background and middle age without spouse and with a low education level and family history without diabetes [13]. This study shows that education through disease knowledge and medication knowledge can improve patients' understanding of diabetes and help patients adjust their lifestyle to a certain extent, thus improving blood sugar control level. Literature review found that most of the studies in the same period focused on the role of self-management education in patients' life and surgical prognosis [14, 15], but there were few studies on the impact of perioperative nursing on patients. In this study, the perioperative nursing of patients was studied, highlighting the importance of perioperative nursing for patients' intraoperative life safety and postoperative recovery.

Psychological problems related to long-term disease are another important factor affecting the treatment of diabetes. Investigations show that the incidence of psychological distress related to diabetes for 18 months is 38% to 48%, which affects the treatment compliance, self-management behavior, and blood sugar control of diabetic patients to varying degrees [16]. And studies on hospitalized diabetic patients show that different degrees of psychological distress are common in newly hospitalized diabetic patients. Therefore, it is possible to start with health education, improve patients' awareness of the disease, eliminate unnecessary fear and rejection, let patients treat diabetes with a positive attitude, reduce psychological pain, and naturally improve the treatment effect.

In our study, the accuracy between experimental data and MLP prediction is very high (100%). We found that FBG, 2hPG, and HbA1c were more important influencing factors in the neural network model, which were closely related to the length of hospital stay after nursing. In this study, only FBG, 2hPG, and HbA1c were counted, and the data were few. In the follow-up, further statistical analysis of inflammatory indicators and nutritional indicators after diabetes surgery will be conducted to further verify the importance of perioperative nursing. The MLP neural network has great application and development prospects in the future, so we can make more reasonable judgments and decisions on the diagnosis and treatment of diseases by predicting the results. The MLP neural network can also be used to predict the therapeutic effect of perioperative nursing on other diseases.

Data Availability

All data are publicly available and available from the corresponding author.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

References

[1] B. K. Bailes, "Diabetes mellitus and its chronic complications," *AORN Journal*, vol. 76, no. 2, pp. 265–282, 2002.

[2] V. F. Okunrintemi, F. Gani, and T. M. Pawlik, "National trends in postoperative outcomes and cost comparing minimally invasive versus open liver and pancreatic surgery," *Journal of Gastrointestinal Surgery*, vol. 20, no. 11, pp. 1836–1843, 2016.

[3] A. Ejaz, A. A. Gonzalez, F. Gani, and T. M. Pawlik, "Effect of index hospitalization costs on readmission among patients undergoing major abdominal surgery," *JAMA Surgery*, vol. 151, no. 8, pp. 718–724, 2016.

[4] T. Chen, S. Kumaran, G. Vigh et al., "Perioperative diabetes management of adult patients with diabetes: a best practice implementation project," *JBI Evidence Implementation*, vol. 20, no. 1, pp. 72–86, 2021.

[5] Q. Xiao, L. Lang, Z. Ma, Y. Zhang, and K. Xu, "Exploration of the curative effect of early enteral nutrition nursing on patients with severe acute pancreatitis and the improvement of patients' mental health and inflammation level," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8784905, 10 pages, 2021.

[6] M. Ahangarcani, M. Farnaghi, M. R. Shirzadi, P. Pilesjö, and A. Mansourian, "Predictive risk mapping of human leptospirosis using support vector machine classification and multilayer perceptron neural network," *Geospatial Health*, vol. 14, no. 1, 2019.

[7] Z. Ma, X. Li, Y. Chen et al., "Comprehensive evaluation of the combined extracts of Epimedii Folium and Ligustri Lucidi Fructus for PMOP in ovariectomized rats based on MLP-ANN methods," *Journal of Ethnopharmacology*, vol. 268, article 113563, 2021.

[8] L. Cong, Q. Q. Hua, Z. Q. Huang et al., "A radiomics method based on MR FS-T2WI sequence for diagnosing of autosomal dominant polycystic kidney disease progression," *European Review for Medical and Pharmacological Sciences*, vol. 25, no. 18, pp. 5769–5780, 2021.

[9] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, "Predicting hypertension without measurement: a non-invasive, questionnaire- based approach," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7601–7609, 2015.

[10] S. Er, S. Kara, and A. Güven, "Comparison of multilayer perceptron training algorithms for portal venous Doppler signals in the cirrhosis disease," *Expert Systems with Applications*, vol. 31, pp. 406–413, 2006.

[11] T. Fujita, A. Sato, A. Narita et al., "Use of a multilayer perceptron to create a prediction model for dressing independence in a small sample at a single facility," *Journal of Physical Therapy Science*, vol. 31, no. 1, pp. 69–74, 2019.

[12] N. Świątoniowska, K. Sarzyńska, A. Szymańska-Chabowska, and B. Jankowska-Polańska, "The role of education in type 2 diabetes treatment," *Diabetes Research and Clinical Practice*, vol. 151, pp. 237–246, 2019.

[13] A. Coppola, L. Sasso, A. Bagnasco, A. Giustina, and C. Gazzaruso, "The role of patient education in the prevention and management of type 2 diabetes: an overview," *Endocrine*, vol. 53, no. 1, pp. 18–27, 2016.

[14] A. Bukhsh, M. S. Nawaz, H. S. Ahmed, and T. M. Khan, "A randomized controlled study to evaluate the effect of pharmacist-led educational intervention on glycemic control, self-care activities and disease knowledge among type 2 diabetes patients: a consort compliant study protocol," *Medicine*, vol. 97, no. 12, article e9847, 2018.

[15] L. Marciano, A.-L. Camerini, and P. J. Schulz, "The role of health literacy in diabetes knowledge, self-care, and glycemic control: a meta-analysis," *Journal of General Internal Medicine*, vol. 34, no. 6, pp. 1007–1017, 2019.

[16] Association American Diabetes, "Updates to the standards of medical care in diabetes-2018," *Diabetes Care*, vol. 41, no. 9, pp. 2045–2047, 2018.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7827448, 8 pages
https://doi.org/10.1155/2022/7827448

Research Article

Elevator Leveling Failures Monitoring Device and Method

R. Z. Sun,[1,2] X. A. Wang,[1] Y. Z. Cai,[2] and J. M. Cao[2]

[1] *School of Software & Microelectronics, Peking University, Beijing, China*
[2] *College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China*

Correspondence should be addressed to J. M. Cao; caojianmin@sztu.edu.cn

Received 7 April 2022; Revised 11 August 2022; Accepted 24 August 2022; Published 17 September 2022

Academic Editor: Amrit Mukherjee

Elevators are highly susceptible to safety incidents in the event of leveling failures, so the ability to monitor related failures must be strengthened. This paper proposed a new elevator leveling failures monitoring device and method in which elevator signals are obtained from the elevator CAN bus interface, transmitted to a remote monitoring platform via NB-IoT, and stored in our private data center. The leveling sensor sensing signal, the door signal, the car call signal, the target floor signal, and the running signal are obtained by analyzing the data extracted from the elevator. Logical analysis could be used to determine the elevator's running status and leveling-related failures. The device and method could identify and also predict leveling-related failures and have advantages in terms of universality, accuracy, and economy.

1. Introduction

In recent years, the number and service time of elevators have increased significantly; meanwhile, elevator failures have inevitably increased [1, 2]. As a result of this situation, elevator monitoring technologies and systems have been designed and implemented in the industry. Remote monitoring systems developed by elevator giants, represented by Otis's ONE™ system [3] in America, Mitsubishi's MelEye system [4] in Japan, and KONE's E-Link™ system [5] in the United Kingdom. These systems continuously monitor the elevator's running status in order to detect or even predict elevator failures in real time. However, these systems are expensive, and because the manufacturers do not publish the protocols [6–8], we do not know how the system monitors, what data is used for monitoring, or what the data structure is, implying that these systems are limited in universality as they could only be applied to their own brand of products. The promotion of monitoring systems has been stymied by these issues.

When the elevator is leveled normally, the car pedal and the external hall door pedal are in the same plane, and elevator leveling helps to facilitate peoples' coming in and out while reducing unnecessary damage to the elevator [9]. The incorrect judgment of elevator car position not only affects elevator efficiency, but it may also cause a series of leveling-related failures. When an elevator car position is incorrectly obtained, the consequences are severe if an accident occurs. Skog et al. [10] used signal processing to achieve elevator safety warning and monitoring. Abnormal stops were identified by monitoring the deceleration of the elevator. Luo and Feng [11] established a failure-tolerant control strategy based on neural networks and used a photoelectric encoder to realize elevator leveling. Lai and Liu [12] could analyze and calculate the target images of the elevator car floor and the elevator floor, and obtain the difference in viewpoint between the two to determine whether the elevator leveling failure occurs and alarms in time. However, these studies are relatively isolated on the leveling failure and do not investigate the internal causes or combine with the elevator's running status. In addition, some scholars have also proposed their own approaches, such as an additional leveling gauge [13] or altimeter [14] being available to monitor leveling failures. In essence, they all install sensors in the elevator shaft, which has the disadvantages of complicated installation, high cost, and the possibility of sensor false alarms, making it difficult to promote the application.

In this paper, we obtained elevator data through the CAN bus interface between the elevator controller and the control box inside the elevator, obtained the leveling sensor sensing signal, target floor signal, and operation signal of the elevator through data analysis, and used logic analysis to determine the elevator's running status and leveling-related failures. Since elevator data is transmitted in real time to the remote monitoring platform, prompt intervention could be requested in the event of leveling failures.

The failure monitoring device and method have the following advantages:

(i) universality. Modern elevators generally use the CAN bus interface, and the device and method described in this paper are not limited by the brand and model of the elevator

(ii) accuracy. Since the running data of the elevator is truly collected through the elevator serial port, and the system logically combines the leveling related failures of the elevator by itself; the failure judgment has a high success rate and is characterized by accuracy

(iii) economical. Since a large number of sensors are not used to collect elevator running data, but the elevator data is collected from the CAN bus interface, the construction cost is low and the installation is easy, which is conducive to the promotion of the system

(iv) it is possible to identify and also predict failures. When the faulted floor is the target floor or passing floor, its leveling sensor signal will show corresponding abnormal variation, thus enabling identifies and prediction of different types of leveling failures

2. Architecture and Methods

2.1. Elevator Data Collection Scheme and Transmission Method. Elevator floors are getting higher and higher nowadays, such as in residential community elevators, which are often above 30 floors. Regardless of the brand or model of elevator, data is usually transferred between the controller and the car communication board using a serial method, such as the CAN bus interface [15]. In this paper, we obtained the running data from the CAN bus interface. The specific connection method of the data collector designed by our research group [8] is shown in Figure 1.

In addition to power and ground (Vcc and Gnd), the elevator CAN bus has two data lines, Can+ and Can-. The data collector gets the elevator signal from the elevator CAN bus interface, and then transmits it to the remote monitoring platform via the NB-IoT module after processing by the microcontroller. Compared with commonly used wireless communication methods, NB-IoT has lower device costs, longer battery life, and expanded coverage [16]. It is especially advantageous for meeting wireless communication requirements for long time, light weight, high stability, and wide coverage, which is consistent with our needs. Once a failure is detected, maintenance personnel can be notified in time to intervene. The elevator monitoring system is shown in Figure 2.

2.2. Elevator Signals Could Be Obtained. Typical elevator CAN bus data includes the STEP elevator used in the experiment, including index, time, name, ID, type, format, Len and data. Although the exact form of the raw CAN bus data varies between elevators, our approach to extraction and analysis remains consistent. In general, the analysis approach is based on the principle of control variables, for example, comparing the data of an elevator at rest on the 1st floor with that of an elevator at rest on the 2nd floor, without interference from other states such as doors and running, to find the data representing the floor. Based on the above principle and methods, the following elevator signals could be obtained by analyzing the elevator CAN bus data.

Door signal is sent from the car to the elevator controller, includes the door closed and door opened signals, which represent the door closed or opened in place, respectively. And door movement signals, which are sometimes refined to opening and closing signals to indicate that the door is in movement.

Car call signal is sent from the car to the elevator controller. Under normal circumstances, whenever a car call signal is generated, it means that someone is summoning the floor inside the elevator.

Target floor signal, generally sent from the car to the elevator controller, is usually in the form of a bit to indicate the target floor. For example, 01 means the first floor is calling, 10 means the second floor is calling, and 11 means the first and second floors are calling at the same time.

Running signal is sent from the elevator controller to the car to indicates that the elevator is running up or down. The elevator's running signal is reset (low level), indicating that the elevator has stopped in place.

Leveling sensor signal is sent from the car to the elevator controller, indicating the sensing relationship of the leveling sensor to the baffle. This includes the upper sensing signal, when the leveling sensor's upper sensing node detects the leveling baffle, the upper sensing signal is activated. Similarly, there are signals with lower sensing, full sensing, and no sensing. These signals are contained in the leveling data frame's 7th byte (from high to low), beginning with 00 01. The format of these signals varies depending on the elevator's up/down/static state. With the work of data analysis, the leveling sensor signals for the STEP elevator used in the experiment are shown in Table 1.

3. Results and Discussion

We use leveling sensor signal to determine the elevator failure method as follows. Following normal elevator leveling, the car pedal and the external hall door pedal are in the same plane, and the acquisition of the car position is critical to determining the elevator leveling status. At present, elevators generally rely on the floor encoder and leveling device to determine the car's position. The leveling device generally includes leveling sensor and leveling baffle; the leveling sensor is installed on the car and runs up and down with it, while the leveling baffle is installed at a fixed position in the elevator shaft. The specific installation is shown in Figure 3.

Elevator leveling sensors generally have photoelectric sensing type and magnetic sensing type, both of which are used to determine the elevator position through the sensing

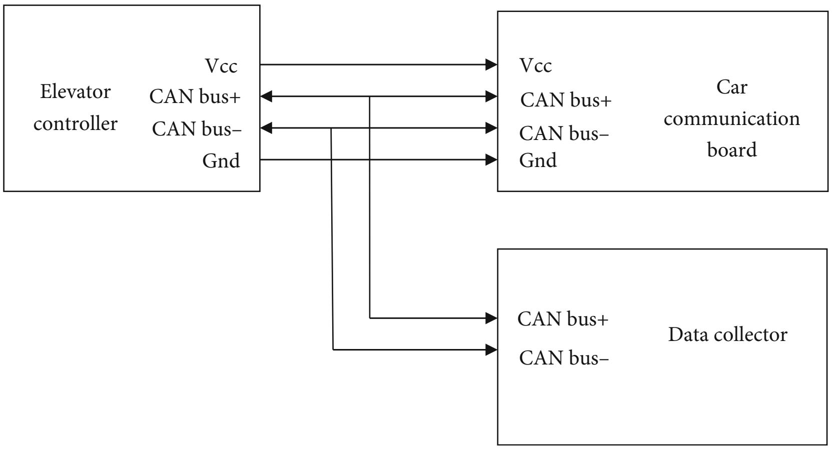


Figure 1: Connection diagram of the data collector and elevator serial interface [8].

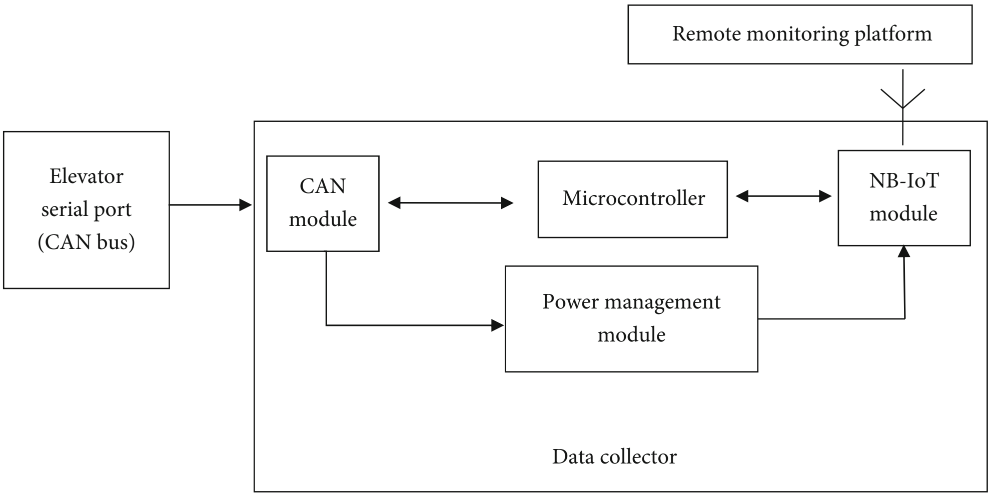


Figure 2: The structure diagram of the monitoring system.

Table 1: Correspondence between leveling sensor sensing signal and elevator running status.

Elevator/leveling sensors	No sensing	Lower sensing	Upper sensing	Full sensing
Static	00	04	08	0C
Upward	01	05	09	0D
Downward	02	06	0A	0E

(Note: The high level of the leveling sensor signal may be 0 or 1, 2 ...... its specific number has no effect on the leveling signal judgment, this paper is unified to 0.).

signal between the leveling sensor and the leveling baffle. The photoelectric sensing type sensor, which is widely used today, has two sensing nodes, upper and lower. When both sensing nodes detect the baffle, the elevator door could be accurately aligned with the elevator exit to allow personnel safe access.

Take the 2nd floor leveling sensor signal as an example in the upward movement (the downward situation is similar to the upward direction). When the elevator is inside summoned from the 1st floor to the 2nd floor, the variation of the leveling sensor signal of the 2nd floor (as the target floor) is shown in Table 2. When the elevator is inside recruited from the 1st floor to the 3rd floor (or higher floors), i.e., the 2nd floor as the passing floor; the variation of the leveling sensor signal is shown in Table 3.

However, leveling failures still occur from time to time due to signal loss, leveling baffle displacement, etc. At this time, the variation of the leveling sensor signal will differ from that shown in Tables 2 and 3.

3.1. Leveling Stopping Failure. If the target floor leveling baffle falls off or cannot be sensed, the elevator will not stop and open the door normally, but will instead continue to run up/down to find the leveling state, which is referred to as "leveling stopping failure" below.

When the leveling stopping failure occurs on the 2nd floor as the target floor, after the elevator enters the leveling range of the 2nd floor, it will not stop and open the door normally due to the lack of leveling full sensing signal, but it will be static for a short time and leveling at the floor

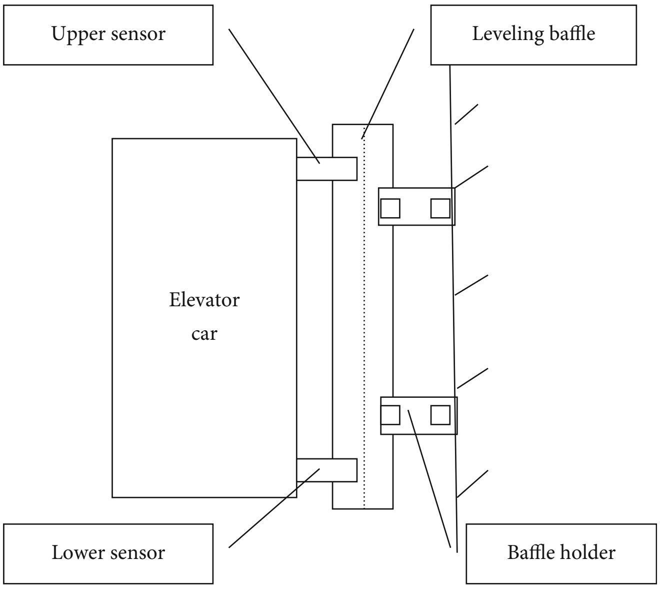


Figure 3: Elevator leveling sensor installation schematic.

Table 2: Variation of leveling sensor signal of the target floor (2nd floor) under normal condition.

Content of the 7th byte	The 7th byte means	Corresponding elevator running status
01	Upward and no sensing	The elevator enters the 2nd floor range, but has not yet sensed the 2nd floor leveling baffle
09	Upward and upper sensing	The upper sensor starts to sense the 2nd floor leveling baffle
0D	Upward and full sensing	Full sensor sensing, still running upward
0C	Static and full sensing	Elevator leveling normaly

Table 3: Variation of leveling sensor signal of the passing floor (2nd floor) under normal condition.

Content of the 7th byte	The 7th byte means	Corresponding elevator running status
01	Upward and no sensing	The elevator enters the 2nd floor range but has not yet sensed the 2nd floor leveling baffle
09	Upward and upper sensing	The upper sensor starts to sense the 2nd floor leveling baffle
0D	Upward and full sensing	Full sensor sensing, still running upward
05	Upward and lower sensing	The elevator continues to go upward; the upper leveling sensor leaves the leveling baffle, and the lower leveling sensor is still in sensing
01	Upward and no sensing	The leveling sensor completely leaves the 2nd floor leveling baffle

nearby. Correspondingly, the leveling sensor signal variation will change from “01-09-0D-0C” to “01–00”. That is, more “00” and missing “09”, “0D” and “0C”. The specific meaning could be correlated to Tables 1 and 2. Similarly, it is possible to predict the leveling failure. When the leveling stopping failure occurs on the 2nd floor as a passing floor, the 7th byte of the leveling data will change to: “01-01”, continuously going up and no sensing, missing “09” “0D “05”, but since the 2nd floor is not the target floor, the elevator will still running upward normally. Elevator data is transmitted to the data center of the monitoring system in real time through the data acquisition board, so that in the event of leveling without stopping failure, failure judgment and prediction could be made based on the sequence variation of signals.

Traditionally, leveling stopping failure could only be identified when the elevator reaches the floor with failure, which means the passengers have to experience the leveling failure. With the elevator monitoring system, however, once

Table 4: Variation of leveling sensor signal at the target floor (2nd floor) in case of leveling alignment failure.

Content of the 7th byte	The 7th byte means	Corresponding elevator running status
09	Upward and upper sensing	The upper sensor starts to sense the 2nd floor leveling baffle
0D	Upward and full sensing	Full sensor sensing, still running upward
05	Upward and lower sensing	The elevator continues to go upward; the upper leveling sensor leaves the baffle, and the lower leveling sensor still senses
04	Static and lower sensing	Elevator static for a short time
06	Downward and lower sensing	The elevator reruns and goes down to find the full sensing state
0E	Downward and full sensing	Full sensor sensing, still running downward
0C	Static and full sensing	Elevator ends running downward, and the inner door opens

Table 5: Variation of leveling sensor signal at the target floor (2nd floor) in case of leveling cycling failure.

Content of the 7th byte	The 7th byte means	Corresponding elevator running status
01	Upward and no sensing	The elevator enters the 2nd floor range but has not yet sensed the 2nd floor leveling baffle
09	Upward and upper sensing	The upper sensor starts to sense the 2nd floor leveling baffle
05	Upward and lower sensing	The elevator continues to go upward; the upper leveling sensor leaves the baffle, and the lower leveling sensor senses
04	Static and lower sensing	Elevator static for a short time, lower leveling sensor senses
06	Downward and lower sensing	The elevator reruns and goes down to find the full sensing state
0A	Downward and upper sensing	Elevator running downward; lower leveling sensor leaves the baffle, and upper leveling sensor senses
......	......	......
……	……	……

Table 6: Variation of leveling sensor signal at the passing floor (2nd floor) in case of leveling cycling failure.

Content of the 7th byte	The 7th byte means	Corresponding elevator running status
01	Upward and no sensing	The elevator enters the 2nd floor range but has not yet sensed the 2nd floor leveling baffle
09	Upward and upper sensing	The upper sensor starts to sense the 2nd floor leveling baffle
05	Upward and lower sensing	The elevator continues to go upward; the upper leveling sensor leaves the baffle, and the lower leveling sensor still senses
01	Upward and no sensing	The leveling sensor completely leaves the 2nd floor leveling baffle

the data is abnormal as described above, it will be detected in real time by the system, and it only requires the elevator to pass through the floor to determine the failure and provide an early warning. Taking a 30-floor community building as an example, the probability of which the failure floor happens to be the target floor is only about 3%, while 97% of the failures could be detected in advance, thus passengers experiencing the relevant failure could be greatly reduced.

3.2. Leveling Alignment Failure. If the leveling baffle is misaligned due to loose screws or other factors, it will usually move down relative to the normal position due to gravity. When the

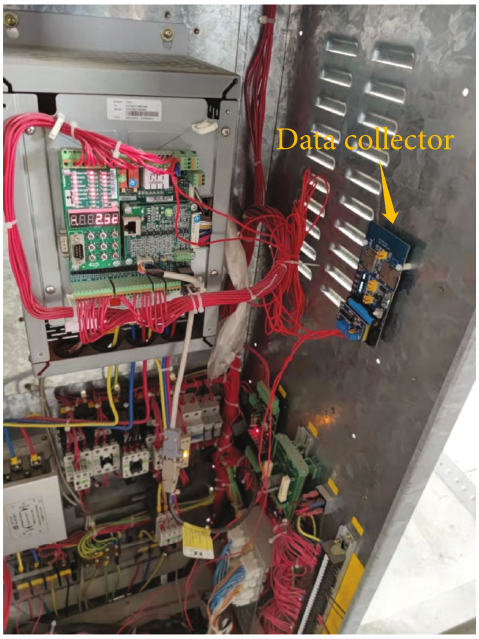
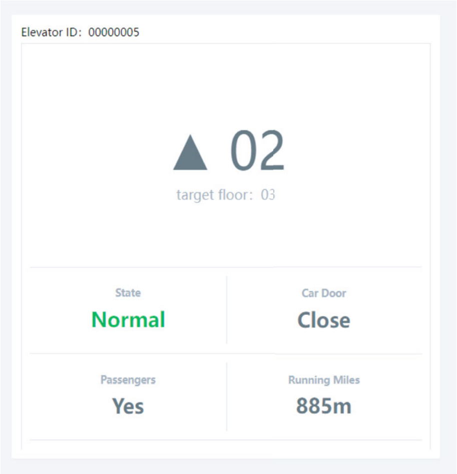


Figure 4: The installation of the data collector in the actual elevator.



Elevator ID: 00000005

▲ 02

target floor: 03

State
Normal

Car Door
Close

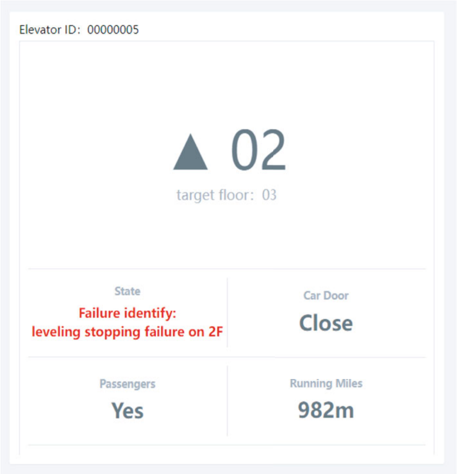
Passengers
Yes

Running Miles
885m

Figure 5: Elevator in normal condition.

elevator reaches the normal levelling range, due to the downward shift of the leveling baffle, the leveling sensor is lower sensing instead of full sensing, and the elevator will not open the door but continue to go down to find the full sensing state of the leveling sensor. Because of the deviation from the leveling range, the floor encoder will judge that the elevator is not in a leveling state, and the outer door will remain closed and the power off. In this case, "leveling alignment failure" described below occurs. Take the 2nd floor as the target floor as an example, when the 2nd floor leveling baffle moves down and leveling alignment failure occurs, the variation of the leveling sensor signal is shown in Table 4.

It can be seen that when the elevator has a leveling alignment failure due to the leveling baffle moving down, there is a variation of "09—0D—05—04—06—0E—0C" (Table 4) compared to the normal leveling sensor signal variation of "09—0D—0C" (Table 2). When the failure occurs, the inner door opens, while the outer door remains closed, which could result in trapping if there are passengers inside.



Elevator ID: 00000005

▲ 02

target floor: 03

State
Failure identify:
leveling stopping failure on 2F

Car Door
Close

Passengers
Yes

Running Miles
982m

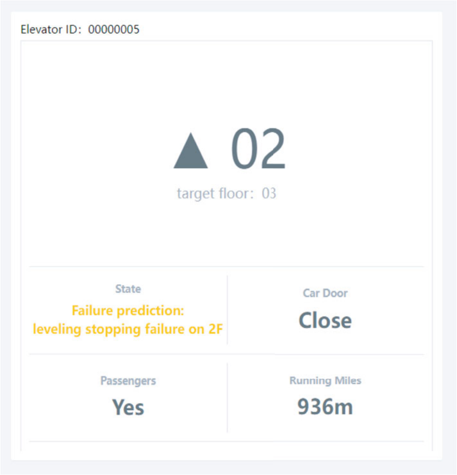
Figure 6: Failure identify.

The leveling sensor signal is transmitted in real time to the monitoring system's data center, and the elevator leveling alignment failure could be identified based on its variation sequence and timely maintenance could be carried out.

3.3. Leveling Cycling Failure. The elevator opens only when the leveling sensor finds the full sensing state. However, if the leveling baffle becomes shorter due to corrosion, fracture, or other reasons, and the length is less than the distance between the upper and lower sensing nodes of the leveling sensor; the sensor would be unable to reach the full sensing state, and the elevator would move up and down in the normal leveling range cycle, unable to leveling the floor. It is the "leveling cycle failure," which is described as follows:

As an example, consider the 2nd floor. If the leveling baffle on the 2nd floor becomes too short to allow the leveling sensor to reach its full sensing state, the procedure of the elevator inside summoned from the 1st floor to the 2nd floor is as follows: After the elevator senses the upper leveling sensor, it continues to go up normally to the lower leveling sensor, but the upper leveling sensor no longer senses the baffle at this time. After a brief pause, the elevator moves downward to find the full sensing state. When the lower leveling sensor is sensing, the upper level sensor is no longer sensing the baffle. After a brief pause, the elevator moves up. So on and so forth, cycling upward and downward until the leveling full sensing state is reached. The variation of the 2nd floor leveling sensor signal under this failure is shown in Table 5.

This is a rare occurrence, but when it occurs, the elevator will cycle up and down in the vicinity of the failure floor leveling range without opening the door, causing serious physical and mental harm to passengers. Furthermore, since it is the normal operation logic of the elevator to look for the full sensing state, the elevator will not actively determine the failure.



Elevator ID: 00000005

▲ 02

target floor: 03

State: Failure prediction: leveling stopping failure on 2F

Car Door: Close

Passengers: Yes

Running Miles: 936m

Figure 7: Failure prediction.

The leveling sensor is missing full sensing “0D” between upper sensing “09” and lower sensing “05.” Once the monitoring system detects such an anomaly in the data, it could identify the leveling cycling failure and issue an alarm.

Similarly, the leveling sensor signal could be used to predict the leveling cycling failure. When the 2nd floor is used as a passing floor, that is, when a passenger runs from the 1st floor to the 3rd floor (or higher) and passes through the 2nd floor, the leveling sensor signal variation shown in Table 6.

Compared to the normal state of Table 3, it can be found that the leveling sensor is missing the full sensing “0D” between the upper sensing “09” and the lower sensing “05” in Table 6. Once the monitoring system detects such anomalies in the data, it could make failure predictions and notify maintenance personnel to intervene in advance.

3.4. The Leveling-Related Failures Monitoring Method. To summarize, the judgment method of leveling-related failures is as follows. When the elevator arrives at a particular floor, it first determines whether it is the target floor or the passing floor based on the target floor signal, and then obtains the sequence of leveling signal variations for the current floor and compares it to the variations under normal condition in Tables 2 and 3. If the variation differs from the normal state, it is possible to conclude that the elevator has a leveling failure.

Furthermore, the specific failure type is determined, and the corresponding identification or prediction is generated in conjunction with Tables 4–6. What is more, if a new type of unknown failure emerges, we could add its sequence variations to the monitoring system and constantly update and optimize the failure judgment strategy. The above description is for the monitoring method when the elevator is in the upward movement; the specific signal when the elevator is in the downward movement is different (as shown in Table 1), but the method is the same.

The installation of the data collector in the actual elevator is shown in Figure 4, which can be corresponded to Figure 1, where the data collector connects the CAN+ and CAN- data lines of the elevator and sends them to the remote monitoring platform via the NB-IoT module. In the experiment, we used the leveling stopping failure of the 2nd floor as an example. The elevator’s running status and level-related faults can be reflected visually. The page of the monitoring platform when the elevator is in the normal state is shown in Figure 5. The page of the monitoring platform for identifying the leveling stopping failure is shown in Figure 6. The page of the monitoring platform for predicting the leveling stopping failure is shown in Figure 7. The experimental results support the efficacy of our device and method.

4. Conclusions

This paper extracts the elevator’s door signal, car call signal, target floor signal, running signal, and leveling sensor signal, and then monitors leveling failures by analyzing the real-time variation sequence of the elevator leveling sensor signal. Whether the elevator is leveling normally is determined. If a leveling failure occurs, the type of failure is identified based on the leveling variation. Based on the failure judgment, timely maintenance is possible.

The device and method are not limited by elevator brand and signal, so they have advantages in universality. Furthermore, raw CAN bus data is collected by data collectors and stored in our private data center, and it could accurately identify failures by logical analysis and is inexpensive because no additional sensors are required. Since the monitoring of the elevator’s CAN bus interface, leveling failures can be identified and predicted once the leveling sensor signal show the corresponding abnormal variations.

Data Availability

Raw CAN bus data is collected by the data collector and stored in our private data center. If you would like more detailed information about CAN bus data, please contact our corresponding mail and state your intention and purpose. We will sincerely consider your request at our discretion and try to accommodate you.

Conflicts of Interest

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

Acknowledgments

First and foremost, we want to express our gratitude to Engineer Zhiqun Luo and his colleagues at the Guangdong Institute of Special Equipment Inspection and Research for providing the actual elevator for our experiments. We would like to thank Dr. Lei Ning for his help in building up the NB-IoT environment as well as Engineer Jun Yu for his assistance in writing this paper. The research was supported by the Science and Technology Plan Project of Guangdong Administration for Market Regulation [2020ZT02], Scientific Research Capacity Improvement Project of Key

Construction Disciplines in Guangdong Province of China [2021ZDJS109] and Industry University Cooperative Education Program of the Chinese Ministry of Education [202002321012].

References

[1] S. L. Tu, Z. Y. Wu, and B. Qian, “Research of the elevator monitoring system based on the internet of things,” *Applied Mechanics and Materials*, vol. 423-426, pp. 2380–2385, 2013.

[2] K. P. Subbu and A. V. Vasilakos, “Big data for context aware computing – perspectives and challenges,” *Big Data Research*, vol. 10, pp. 33–43, 2017.

[3] “Otis ONE™ [EB/OL],” https://www.otis.com/zh/cn/products-services/otis-signature-service/otis-one.

[4] “MelEye Monitoring and Control System[EB/OL],” http://www.mitsubishielectric.com/elevator/products/basic/elevators/control_system/index.html.

[5] “Kone Monitoring Solutions[EB/OL],” https://www.kone.cn/zh/new-buildings/advanced-people-flowsolutions/monitoringsolutions.aspx.

[6] W. G. Bao and Q. Zong, *CAN bus-based Remote Monitoring System for Elevators*, Modern Machinery, China, 2005.

[7] J. Chen, X. Li, S. Zhang, L. Li, H. Yang, and X. Wang, “Remote monitoring system of elevator energy consumption in green building based on ARM,” *Modern Manufacturing Engineering*, 2018.

[8] Q. Huang, J. M. Cao, and R. Z. Sun, “Design and implementation of an elevator power failure warning system,” *Journal of Physics: Conference Series*, vol. 1621, article 012050, 2020.

[9] X. H. Pan, “Research on elevator safety monitoring system based on internet of things technology,” *Wireless Connected Technology*, vol. 20, pp. 62-63, 2016.

[10] I. Skog, I. Karagiannis, A. B. Bergsten, J. Harden, L. Gustafsson, and P. Handel, “A smart sensor node for the internet-of-elevators—non-invasive condition and fault monitoring,” *IEEE Sensors Journal*, vol. 17, no. 16, pp. 5198–5208, 2017.

[11] J. W. Luo and S. C. Feng, “Fault diagnosis and fault-tolerant control of high-speed elevator leveling sensor,” *Mechanical Engineering and Automation*, vol. 1, pp. 158–160, 2016.

[12] M. G. Lai, X. Z. Liu, L. M. Huang, Z. F. Zhong et al., “Method and system for detecting elevator leveling faults, China,” 2018, CN107539855A.

[13] S. Jin, “A kind of elevator preventive detection and maintenance method, China,” 2020, CN112061926A.

[14] Z. W. Hua, M. Z. Ling, and C. Chen, “A kind of elevator car position monitoring system, China,” 2020, 202021991426.

[15] Z. Yang and Q. F. Li, “Serial communication of elevator based on CAN bus,” *Microcomputer Information*, vol. 116, no. 20, pp. 56-57, 2005.

[16] J. Liu, R. L. Du, L. D. Wu, Y. F. Yang et al., *Research on NB-IoT Network Structure Optimization Method*, Telecom Engineering Technics and Standardization, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 6604286, 14 pages
https://doi.org/10.1155/2022/6604286

Research Article

Hybrid Genetic Algorithm for IOMT-Cloud Task Scheduling

Adedoyin A. Hussain [1] **and Fadi Al-Turjman** [2,3]

[1] *Computer Engineering Dept and Research Centre for AI and IoT, Near East University, Nicosia, Mersin 10, Turkey*
[2] *Artificial Intelligence Engineering Dept and AI and Robotics Institutes, Near East University, Nicosia, Mersin 10, Turkey*
[3] *Research Centre for AI and IoT, Faculty of Engineering, University of Kyrenia, Kyrenia, Mersin 10, Turkey*

Correspondence should be addressed to Adedoyin A. Hussain; hussaindoyin@gmail.com

Received 21 March 2022; Revised 6 June 2022; Accepted 16 August 2022; Published 5 September 2022

Academic Editor: Ashish Bagwari

Task scheduling for the cloud is one of the main advances in IoMT stage, which impacts the whole execution of the cloud resource. Cloud is a proficient headway for computation, and it encompasses data storage, management, and manipulation in large volumes. Thus, a proposition is being made a better approach to proffer task scheduling in the cloud. In this case, a new hybrid genetic algorithm (HGA) is proposed. The proposed HGA method will be justified by contrasting it with the previous researches and approaches. The CloudSim is utilized to quantify their effect on various metrics like timing factors and resource utilization. The proposed HGA technique enhanced the viability of task scheduling with a better execution rate of 32.57 ms. Thus, the experimented outcomes show that the HGA also reduces cost profoundly.

1. Introduction

A reasonable scheduling technique is required to schedule these IoMT requests to cloud resources. Scheduling task is classed as one of the focal issues for computing in IoMT-cloud. The IoMT-cloud is progressed with the improvement of PC and association advancement. This prompts the execution of all tasks efficiently and also provides patients with formidable QoS [1, 2]. Task scheduling for the cloud is one of the main advances in IoMT stage, which impacts the whole execution of the cloud resource. Authors in [3] proposed a scheduling computation for tending to the cloud task to further develop scheduling estimation, which can get the more unobtrusive time and lower cost for doing for each process. Numerous investigations show that IoMT-cloud task scheduling problem is termed as a NP-hard problem, which has been concentrated by various analysts. The work in [4] proposed particle swarm optimization (PSO) to deal with the idea of the organization of users. It has achieved extraordinary results in the field of arranging resources to cloud tasks ensuing in finishing a huge number of coherent tasks. It incorporates a moderate speed of processing and is essentially caught in more waiting time. Authors in [5] prepared a hereditary reenacted tempering estimation for task arrangement with twofold fitness, and this can effectively change the solicitations of the clients for the properties of tasks and work on the clients' satisfaction appropriately. Authors in [6] use the procedure for tending to the cloud task scheduling by exceptional self-changing underground ant colony optimization (ACO) in handling the scheduling of tasks.

To work on scheduling issues adequately in the IoMT-cloud stage, the environment has to be viewed and studied. Figure 1 provides a proper view of the IoMT-cloud. Cloud is a proficient headway for computation, and it encompasses data storage, management, and manipulation in large volumes and uses that data to understand a given outcome [7–11]. This reduces the outright period of manpower and lessens the cost, in the health system. This is a foremost advancement that makes use of the probability of business execution of computer programming with patients publicly [12]. IoMT-cloud is another progression gotten from grid computing, and it suggests involving enrolling assets in an association and accommodating recipients on demand through the Internet [13]. Scheduling in the cloud is one of the major factors in IoMT that it is considered to be the

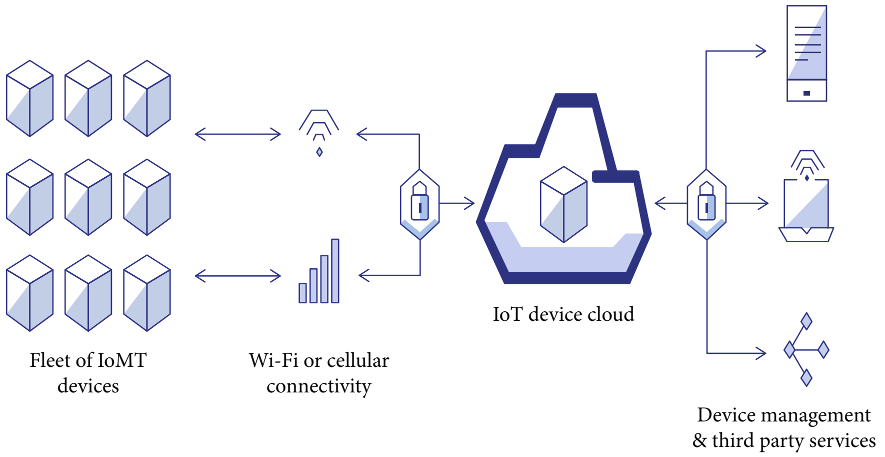

Fleet of IoMT devices
Wi-Fi or cellular connectivity
IoT device cloud
Device management & third party services

Figure 1: IoMT-cloud platform illustration.

essential factor that controls several operations like flexibility, patient resource sharing, and power use. Regardless, there are various troubles normal in IoMT scheduling. High execution rate can be caused by the scheduling technique, and task weights for each process will be scattered across all resources adequately and effectively to get less hold-up time, execution time, and most outrageous throughput. This process can solve a segment of the troubles faced in IoMT computing.

The critical ideal process of IoMT experimentation is that it propels authentic use of resources [14]. Each impacts the other. Fitting these IoMT tasks adequately might achieve efficient utilization of resources. With this, patients can get content wherever and without hoping to contemplate the working of the establishment. IoMT works on no limit provided that there is an internet connection. Therefore, task sharing and resource utilization in the IoMT stage are two sides of a lone coin. The cloud propels organizations to breach the gap between users or patients [15]. Cloud organization can scale up or down resources in the IoMT stage, per the solicitations of the applications. The cloud organization client can rent the resources at whatever point and release them with no difficulty. The cloud organization provides remote assistance regarding any application or resource to the users. This is a central purpose of the IoMT-cloud computation; nonetheless, the organization may be responsible for paying additional costs for this proposition. The example of the IoMT-cloud trends is depicted in Figure 2. Consequently, resource management and task scheduling are required bits of IoMT-cloud research [16]. In handling complex task scheduling-related issues, the usage of scheduling computation is recommended. The adequacy of resource uses depends upon the scheduling and resource weight, rather than the unpredictable designation of resources. Scheduling in IoMT-cloud is for the most part used for handling complex endeavors (client requests). Such arranging computations impact the resources.

In this work, the commitment provided has described major factors that are required for task scheduling, which are as follows: a survey on task scheduling optimization; a portrayal and analysis of the result gotten from the examination; proposing an HGA for IoMT-cloud compared to previous studies; and summing major points and issues in this paper.

This work is segmented as follows. Section 2 gives the background about scheduling procedures and techniques in the IoMT-cloud stage. An introduction to various literature reviews which add to the idea of the method and experimentation utilized is introduced in Section 3. Section 4 provides the problem statement proposed in the research, while in Section 5, the used technique, materials, and the proposed method utilized in experimenting are discussed. Section 6 discusses the outcomes of the experiment. Section 7 presents the conclusion and the closing remarks. Table 1 shows the list of abbreviations used.

2. Background

Here, it discusses the notable optimization scheduling techniques. More light will be given in the literature concerning related works in the cloud environment.

2.1. Shortest Job First (SJF) in IoMT-Cloud. In this conventional methodology, a need is given the length of the task process. It begins from the least to the task with the highest process. In this model, the task is organized on their necessities. The mentioned resource is then allocated to the task process that has the littlest time [18]. It is a rule of a medium waiting time among all other computations. The model is known as a precautionary methodology that picks on cycles that have the least execution time. It does not guarantee task fairness when tasks are distributed to VM [19]. Be that as it may, it has a more drawn out finish time. With this, this procedure is said to be a static scheduling procedure. This is a direct result of tasks with high processes being left unattended to, while little processes are taken care of. It has these processive traits:

(i) It will always be aware of the next task process

(ii) It lessens the waiting time for the task process as it processes little task before huge ones

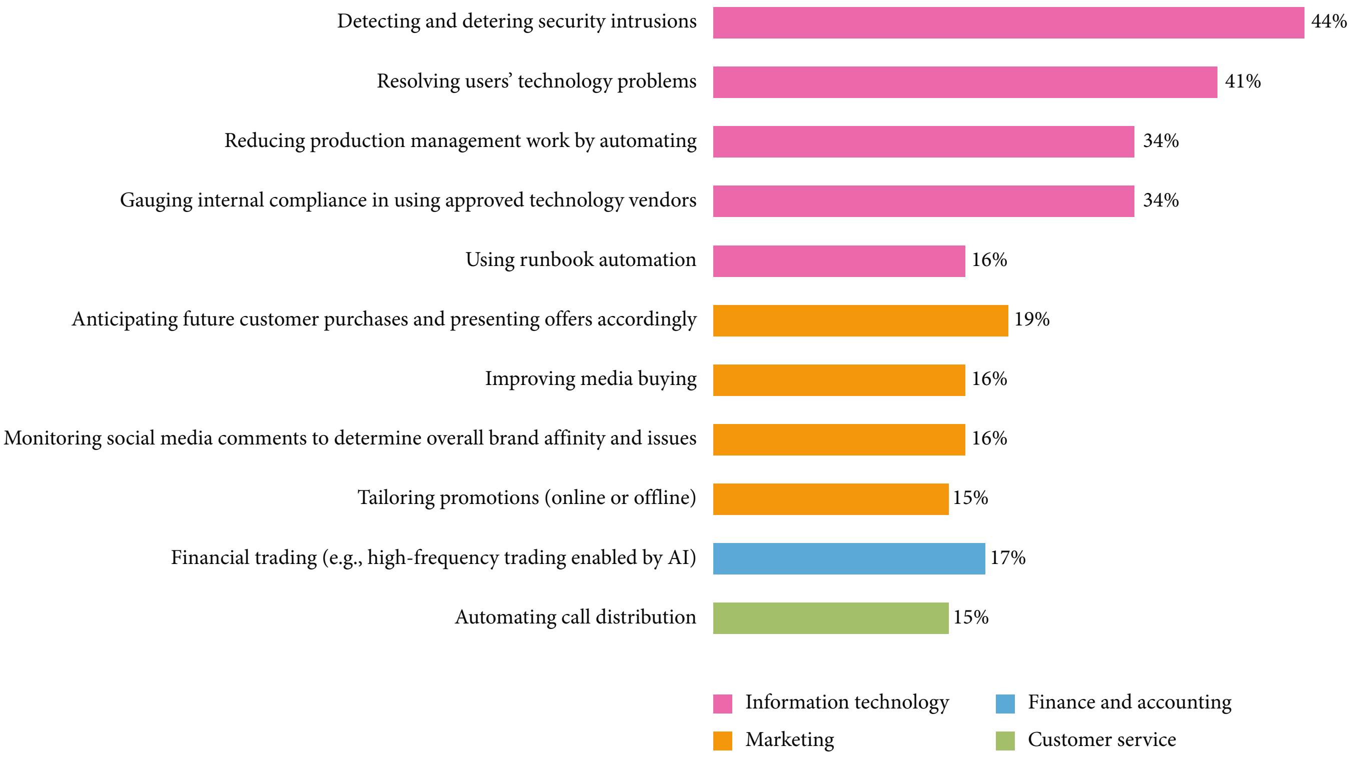


Figure 2: Trends of IoMT-cloud [17].

Table 1: Abbreviations used in the work.

Terms	Meaning
SJF	Shortest job first
PSO	Particle swarm optimization
ICT	Information and communication technologies
QoS	Quality of service
RR	Round robin
IaaS	Infrastructure as a service
TET	Total execution time
AI	Artificial intelligence
ML	Machine learning
NP	Nondeterministic polynomial
GA	Genetic algorithm
ACO	Ant colony optimization
TFT	Total finish time
TWT	Total waiting time
PaaS	Platform as a service
FCFS	First come first serve
PC	Personal computer

2.2. First Come First Serve (FCFS) in IoMT-Cloud. FCFS is a customary methodology, a task that shows up first is served first. The latest request from the patients is installed into the tail of the line. The solicitation of assets relies upon the time of task arrival. This is one of the standard methodologies, and it is more alluring than different methodologies [20]. It depends upon the standard of FIFO with lesser complexity than other computations techniques [21]. This process is immediate and expedient. Whenever we have immense requests, all requests delay until the primary occupation is done. To evaluate the achievement by this technique, we will test them and subsequently gauge their impact on a few legitimate rules in the methodology. With this, this booking strategy is the static methodology. The FCFS has these qualities:

(i) Prioritization depends on the main request and each cycle towards the end finish before new cycle

(ii) This kind of computation does not work honorably with postponing traffic as holding time and mapping are for the most part on the higher side

2.3. Round Robin (RR) in IoMT-Cloud. In this conventional methodology, all task process is executed with fairness. In a general sense, this approach is differentiated to the static type as a result of its dynamicity [22]. Right away, all task processes are given equivalent time for execution which is once in a while called the quantum time. All processes are kept in the solicitation as they show up [23]. In light of the model utilized in this work, the quantum is picked given the mean of the cumulative process time. Right after deciding the mean, it will portray the finish time at the same time. It usually has these properties:

(i) Assuming we apply a more restricted quantum, by then, productivity might become low

(ii) Juggling the quantum to get a decent time will increase time process efficiency

2.4. Genetic Algorithm (GA) in IoMT-Cloud. The GA is an AI strategy that has gained ascend in execution lately. The GA is a metaheuristic approach that deals with the foundations of hereditary qualities and regular determination. The GA approach begins in light of its underlying population [24]. The general population is taken self-assertively to fill in as the early phase for this procedure. A fitness calculation is always used to get the fittest of the chromosome for a general population. Given these factors, chromosomes are picked, and mating operations are carried out on them for the new generational population. The fitness variable surveys the idea of each successor [25]. This paper will utilize an HGA approach which is an adjusted GA approach for greater legitimacy. It will be examined further in the next segment. The fundamental GA approach is exhibited below:

(i) Initialization: Generates an initial populace

(ii) Fitness: Based on the fitness value, calculate for each chromosome

(iii) Mating pool: Select the 2 best chromosomes after wellness handling, and this is otherwise called the guardians. Hybrid produces results by choosing chromosomes to play out this activity to deliver new chromosomes known as posterity. At long last, mutation happens by playing out the change strategy on the chromosomes for a superior chromosome

(iv) Fitness: Based on the fitness value, calculate for each chromosome

(v) Repeat 2 to 5 until meeting the end condition. A stopping condition may be the number of cycles

(vi) End procedure by giving the result of the best chromosome as the last outcome

3. Literature Review

This section gives an overview of several studies on scheduling arrangement and resource distribution. Various experts put forth replies to solve the problem of scheduling. Authors in [26] put forth a multiobject technique that applies better differential progression computation. However, task types are not emphasized in this philosophy. Thus, further improvement can even presently be made. However, this current technique provides a cost and time model for conveyed scheduling. This process does not depict the genuine utilization of resources. Authors in [27] put forth a queue arranging and changing estimation that does not emphasis on work sizes. A programming nonlinear model was used to disperse assets for tasks. Likewise, in [28], they introduced the preparation of tasks reliant upon an excursion lining model. Nonetheless, writers in [29] proposed the scheduling of users request while pondering transmission of information as a resource. In [30], they proposed analytic hierarchy process (AHP) situating based endeavor arrangement. The proposed system does not focus on rashly finishing the processes and starvation. Authors in [31] proposed an acquainted moving skyline approach with planned tasks. They considered the FCFS process for managing demands when assets are free. Subsequently, in [32], they anticipated equivalent extraordinary weights based on incoming demands. Writers in [33] put forth a need-based business scheduling estimation for use in disseminated registering. Authors in [34] put forth the use of a metaheuristic upgrade to diminish costs through task arranging. In [35], they introduced the high-level cost of energy and coating delay goals. This system does not ponder the availability of resources or the weight of tasks. Be that as it may, in [36], they proposed the usage of modified bug area upgrade in load changing. Authors in [37] proposed a system subject to a multiguidelines computation for arranging specialist load. This strategy works on the makespan of a work. Thus, in [38], they put forth a resource assignment problem that means to restrict the full-scale energy cost of appropriated scheduling structures while meeting the foreordained client level SLAs according to a probabilistic point of view [39–41]. Here, they have applied a contrary philosophy that applies a disciplined approach on the off chance that the client does not meet the SLA plans. Consequently, in [42], they proposed a structure subject to the requirement for performing a distinguishable weight schedule that uses coherent movement measures. The technique robotizes the cycle and diminishes the piece of human management, while in [43], they introduced a central weight changing the decision model in the cloud. A couple of makers have proposed a heuristic estimation to handle task arranging and resource task issues portrayed already. Regardless, the technique is lacking in choosing the weight of center points and, arrangement nuances, and the complete phase has no support, as needs are achieving a lone reason for dissatisfaction. Moreover, in [44, 45], they focused on arranging endeavors while contemplating various goals. This technique coordinates additional examination on task planning and resource distribution.

Another approach scheduling approach is using the ant colony optimization (ACO) planning computation. Authors in [46] proposed using hybridization of bug region movement strategy for reasonable weight-changing process, using bug settlement min-max methodology, and inherited estimation. This, finally, processes the amount of pattern of virtual machines from the cloud applications. They proposed a solid method to restrict movement cost of VM and also hold tight to the SLA (service level agreement) which guarantees a QoS. Through this, the need is apportioned to VM to extend the response period of the system and to achieve better weight changing. Authors in [47] proposed a procedure that assists the starvation in work change. To vanquish this trouble, they used innate computation with the logarithmic least square strategy. With this, [48] put forth an improved GA by using the fragmented people decline procedure planned parenthood of the rocky mountains (PPRM). Authors in [49] have audited keen cloud scheduling for load changing and proposed antlion optimizer (ALO) to provide better outcome in changing the cloud storage. After this cycle, GA is implemented to the new populace and observes fitness regard. This gives more huge courses of action, while ALO handles the gigantic issue in space.

The following are the three principle exercises. Fundamental GA has terms called mutation, crossover, and fitness work. Authors in [50] have looked into extraordinary GA for making a response. They have considered a need-based basic evaluation. By this idea, they achieved better ordinary response time and augmentations cloudlets with change encoding. It helps with decreasing time in waiting. In [51], they have proposed a cloud-based approach for the most part of the storage and dynamic multimedia load balancing (CSdynMLB) technique to change the stack for specialists. They have introduced job unit vector (JUV) and processing unit vector (PUV) terms to get the fitness of individuals. A similar need is applied to every one of the requesting and ensures better QoS, high interoperability, and flexibility. Authors in [52] have proposed cross variety genetic computation like genetic ant colony algorithm-virtual machine placement (GA-GEL) estimation for VM load balance process in the cloud. In [53], they put forth the genetic ant colony algorithm-virtual machine placement (GACA-VMP) method for managing settling VMP issues using further developed ACA. The outcome showed up with the Cloud Analyst proliferation gadget that fluctuates with a different number of server ranches. Through this procedure, they have picked a feasible way in two phases. This is gained to successfully pick the genuine specialist and assemble the resources [54–59]. Yet it has been discussed in the literature that there are still various areas that need tending to, and this work proposed here aimed to settle these issues.

4. Problem Formulation

This research addresses the issue of task scheduling in the IoMT-cloud which is a widely distributed and heterogeneous environment. Here, the sets of processors and tasks are considered as P_m and T_n, respectively. Let us say the available P_m processors for some set of tasks T_n, with no sharing during execution. Let ECT be the expected completion time, which contains estimated time for execution of a particular task on each resource, and the estimated completion time of a resource. The goal here is to reduce the total completion time of task execution. To increase resources utilization and minimize the time, the tasks have to be efficiently scheduled or mapped appropriately on the resources available. The depiction is shown in Figure 3.

$$\begin{aligned} T_n &= (T_1, T_2, \cdots, T_n), \\ P_m &= (P_1, P_2, \cdots, P_m), \end{aligned} \tag{1}$$

where T_n is given by set of tasks and P_m is set of resources.

The goal here is to map $T_n \longrightarrow P_m$.

5. Proposed Methodology

The IoMT-cloud has different characteristics which gives benefits to the end client. The major features of IoMT-cloud are self-redesigned, adaptability, and customization. The structure intends to work on the display of scheduling in IoMT, while simultaneously diminishing computational costs. The hopeful features of cloud resources are essential to permit organizations that absolutely layout clients' fulfillment. The key objective is to expect the best technique for the scheduling process when required. Certain bodies should be considered while satisfying these destinations like cloud providers and clients of the cloud. To achieve this, we play out a calculated assessment for scheduling in the IoMT-cloud environment and optimize it by utilizing the proposed AI approach which will be the HGA. Furthermore, we separate the essentials and consequences of utilizing quality of service (QoS) with the proposed outcome. The calculation ought to be sufficiently skilled to manage the issues related to scheduling like resource questions, lack of resources, and over-provisioning of resources.

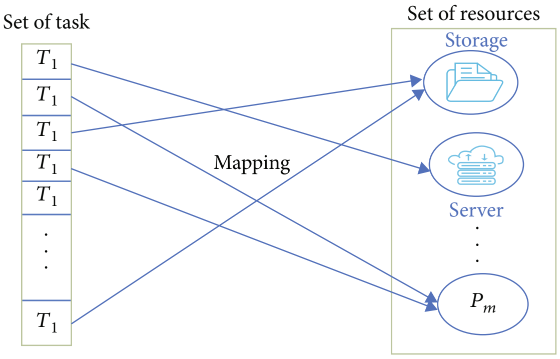


Figure 3: Scheduling task mapping.

The user demands the assets, and the cloud supplier is liable for the task of the expected asset, so the client evades the infringement of the service level agreement (SLA). For the strategies for arranging IoMT-cloud assets, the cloud information service (CIS) is responsible for the properties of each resources and its availability. The cloud scheduler must be efficient to designate different virtual machines (VMs) to various processes. Thus, the scheduling process in the IoMT-cloud is shown in Figure 4. The proposed AI technique will use a hybrid genetic algorithm (HGA) with the blend of a dynamic round robin and a local search, with a variation in step from the authors' previous research, and the depicted outcome will predict the result by recognizing the one with the best result. The outcomes are broken down in light of various related limits (the client and supplier desired) with the best outcome being discussed in the accompanying subsection.

5.1. Hybrid Genetic Algorithm (HGA). GA portrays a general population upgrade technique in light of a progression pattern of nature. In GA, each chromosome addresses a possible response for an issue and is made from a progression of characteristics. Given fitness factors, chromosomes are picked, and mating is performed for a new populace to emerge. The fitness evaluates the idea of each successor. Fitness is described to look at the worth of the chromosome for the general population. The cycle of fitness calculation is

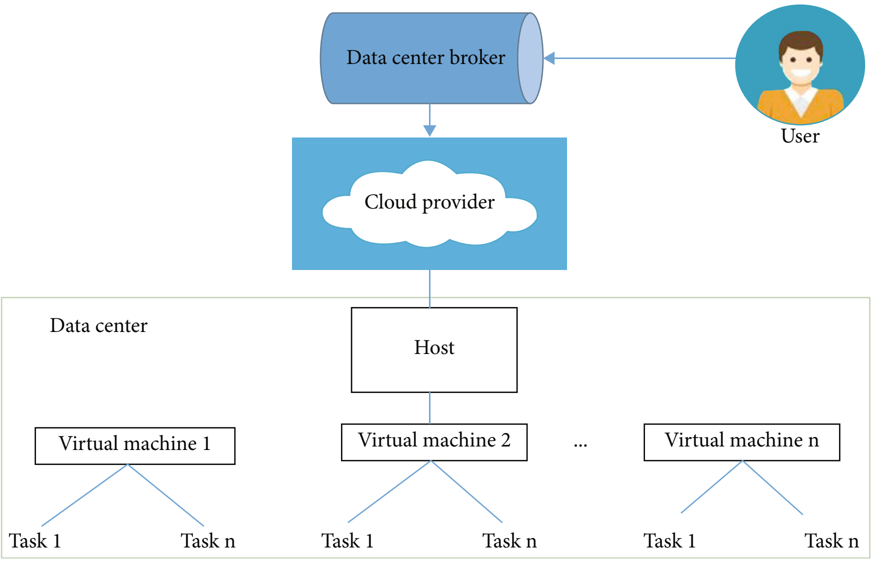


Figure 4: Task scheduling process structure.

repeated until satisfactory successors are made. Here, this approach will have a slight variation from the author's previous research. The proposed HGA will combine dynamic round robin and a local search algorithm known as hill climbing. The flowchart of the HGA in the IoMT-cloud is displayed in Figure 5. The HGA in the IoMT-cloud process is shown as follows:

(i) Initializing the Process

Introductory generation of populace P consisting of chromosomes. In this scheduling problem, we are using the datasets that have been taken from various IoMT devices from users' requests as input. The cumulative time to complete all the operations on all machines will be considered for the IoMT devices. The main objective of the problem is to find a valid schedule that yields the minimum completion time.

For initializing the initial population, the individuals in the population will consist of task and VM ID. This will be embedded together to form a chromosome, and each chromosome is a solution on its own. Each chromosome will have a representation like this: (e.g., VM2: - TS1-TS3-TS6).

(ii) Dynamic Round Robin and Fitness Calculation

In this mode, the round robin will work on a dynamic quantum time. The quantum time will be the median of all the processes. Let us consider the processes (T4, T5, T6, T7, and T8) with their respective completion times of (10, 5, 5, 5, and 10); in this case, the quantum time will be given has the median of these processes. Implementing this will grant task fairness for the task with longer and minima time process. This procedure will continue until all the processes are executed.

Thus, after the dynamic round-robin process, the fitness can be calculated. The completion time for task T_n on R_m is given using

$$\mathrm{TCT} = \max\left(\mathrm{CT}_{n,m}\right), \tag{2}$$

where max $\left(CT_{n,m}\right)$ is the maximum time to complete task T_n on R_m. T_n and R_m are set of tasks and resources, respectively. m and n are the numbers of virtual machines and tasks. TCT is the total completion time.

Then, to minimize the completion time TCT, the execution time of every task for every VM must be calculated. The processing time is to be calculated where P_{nm} is the processing time for task P_n on R_m and C_n computational complexity of task P_n and the processing speed of the virtual machine is PS_m.

$$P_{\mathrm{nm}} = \frac{\mathrm{Cn}}{\mathrm{PSm}}. \tag{3}$$

After getting the processing time, the processing time of every task in the VM has to be calculated using Pj

$$P_m = \sum_{ı=1}^{n} P_{\mathrm{nm}}. \tag{4}$$

(iii) Selection

Once the fitness is calculated for each individual or chromosome, tournament selection is utilized to select the better chromosome from the pool of chromosomes. These selected chromosomes are used to perform crossover and mutation operations. This selected chromosome will be the parents. Chromosomes are selected, and the fitness is compared and then whichever chromosome possesses a lesser completion time is the best chromosome.

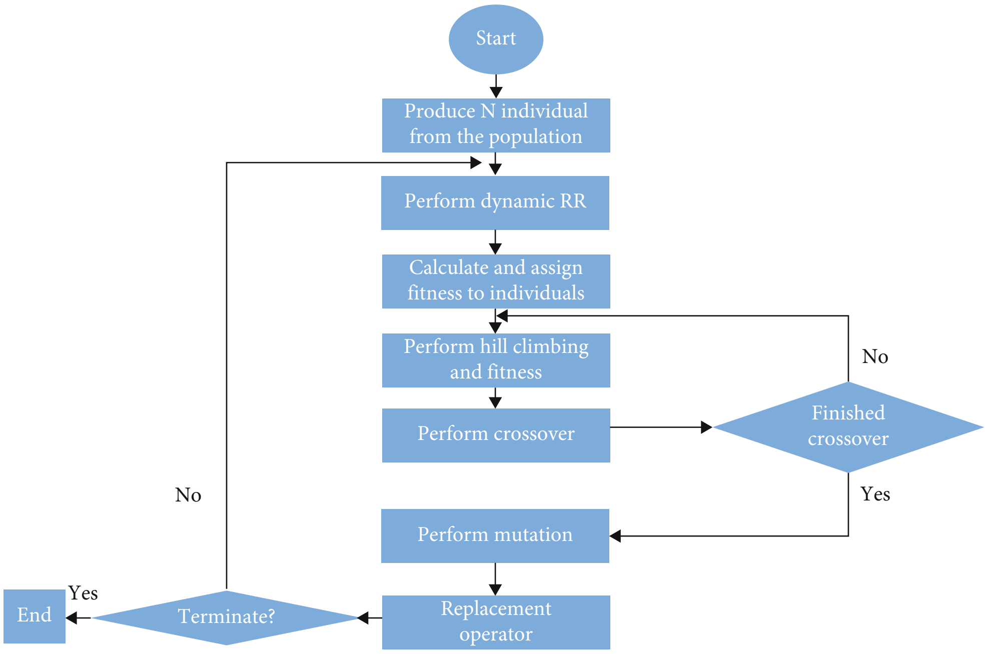


Figure 5: HGA flowchart.

(iv) Hill Climbing Operation

The newly generated parents will be used to perform the hill climbing operation. The hill climbing is going to be a stochastic approach where the initial hill point for the chromosome is chosen at random towards the uphill move. It is an increasing value mode. It generates new solutions on the hill based on its search space. The probability of new solutions might vary due to the steepness of the hill. The hill climbing will consist of two main approaches. A candidate generator is one that maps a solution to a set of possible successors and the evaluation criteria to rank every valid solution. The process will assist to generate a more fit parent that can produce a better offspring.

(v) Crossover and Mutation

This operation is also referred to as the mating pool. The parents get from the selection operation will be used to perform the crossover. Here, uniform crossover is applied. After the crossover, two new chromosomes will be produced. These two new chromosomes will make it four chromosomes in total. From the four new chromosomes, the best of these will be selected as the new offspring, and the latter will be added back into the population for possible selection later on. After this process, the mutation operation will be applied for a fitter value.

(vi) Replacement

Update the populace P. This will replace the populace with better chromosomes from the new generation of offspring. Repeat stages 2 to 6 until stopping criteria are met.

(i) The resulting output will be the best chromosome

(ii) End process

5.2. Experimental Process. Assumptions to be viewed while planning the process in the IoMT-cloud are as follows:

(i) Each task is dispensed to only a solitary VM resource

(ii) The task will be greater than the amount of VMs. This infers that every VM ought to process more than one task

(iii) The task is not obstructed once their executions start

(iv) The lengths of the task will be of various sizes

(v) The available VMs are of prohibitive use and cannot be split among different tasks. It suggests that the available VMs cannot consider various tasks not until the realization of the present task is in progression

(vi) VMs are independent concerning resources and control

5.3. Visualization. The huge motivation driving depiction is portraying the information and graphically speaking with it. This is with the creative aspect that the experimental outcomes are portrayed graphically. The case of information understanding is portrayed as processing and manipulating data, information depiction, and construction attestation, outcome depiction, and finally looking at the information. The yield will be depicted visually in this work for more understanding.

5.4. Computational Environment. Eclipse is an environment for data evaluation and authentic approaches. The assessments were implemented using this IDE. It is an open-source software which implements the use of AI methodologies. CloudSim is used for simulating in the IoMT-cloud stage. Java programming language is likewise the most outstanding language, and it offers various library packages that can handle information science attempts, for example, information assessment, information predealing, and explicitly, working of different techniques. The research is implemented using a PC with Intel i7 processors: 2.3Ghz, GPU: GEFORCE, Disk: 1 TB, RAM: 12GB.

6. Results

Each experimented model and the proposed model will be tested to anticipate which model gets the higher assessment result. To assess the plausibility of our technique, the proposed technique has been contrasted on various optimization and hybrid approaches. The models have been endeavored with various settings to accomplish the most fundamental TWT, TET, TFT, cost, energy efficiency, and resource utilization. This work has done a lot of different assessments with the most reassuring scheduling computations. This work has used traditional optimization and other hybrid algorithms for contrasting with our model to outperform the communicated scheduling issue in IoMT-cloud and accordingly improve it with the proposed model. Likewise, various VMs were used, and various IoMT tasks are used in this evaluation. Each model shows its capability while scheduling. Each model used a relative region of educational collections. Right after the best model is displayed, we see its usefulness with the recently referenced qualities to best predict these outcomes. As follows the eventual outcome of the models is clarified in this part. Eclipse and CloudSim were used which include different libraries for this task.

6.1. Metrics and Parameters. For validating the results of our proposed techniques with other models, the computational metrics below are used for this work. The authors add one more metrics which is energy efficiency in contrast to their previous work, though Table 2 depicts the parameters utilized.

TWT (total waiting time): This is a user-desired criterion. It is the wait time for task execution when a couple of resources compete for a particular resource. This time is the time spent waiting by the cycle or errand on the queue waiting for execution.

TET (total execution time): This is a user-desired criterion. The proportion of time to execute a cycle is a basic part. This time intimates the time between the appearance of a process and the finish time. This is likewise the aggregate sum of time spent by the cycle from coming in the queue until it finishes.

TFT (total finish time): This is a user-desired criterion. It is the distance on schedule from the beginning of an assignment until it wraps up. This is the all-out time at which an undertaking finishes its execution.

Table 2: Used parameter.

Parameters	Value
Task	10-40
Data center	0-3
Population size	120
Iteration	100
Mutation rate	0.05
Crossover rate	0.6
Data center	0-3
Bandwidth (mbps)	500-1000
Ram (Mb)	512-1024
Machine	0-14
Processing elements per Vm (Mips)	500 – 1000

Resource utilization: This is a service provider desired criterion. The utilization of resources is one more parameter that depicts the amplification of assets used. The usage of resources ought to be high in the scheduling framework, though providers need to attain maxima profit by rendering a set number of resources. This parameter is one of the primary meanings in task scheduling. The resource will be kept occupied. Also, reaction time and throughput are huge; however, one more significant boundary for task execution is the utilization of resources.

$$\begin{aligned}&\text{Average resource utilization}\\&\quad=\frac{\text{Time is taken by resource i to finish all the task}}{\text{Makespan}}\times n.\end{aligned}\tag{5}$$

Status/availability: This defines the resources that are available at a given time. This is a huge element in closing how to scatter and apportion the right assets for a given VM. Resource availability is one of the principal parts of scheduling. The accessibility status is a triumph when the right resource is being consigned to the VM.

Throughput: This is a service provider desired criterion. Throughput can be portrayed as the extent of a process being completed per time unit. Thus, throughput is the cycles executed over jobs completed in a unit of time. The schedule should want to extend the quantity of tasks executed per time unit.

Energy efficiency: Energy consumption is the power consumed during the processing of each client's request. To attain energy efficiency, the consumption of power must be reduced drastically. This is one of the major factors to be considered to arrive at a greener environment.

Cost: This is the monetary cost that depicts the total aggregate that ought to be paid by the client for the asset being utilized. This monetary cost will be established on how much time is spent by the client on a particular asset. The equation below portrays how it is calculated where T alludes to the time the resource being utilized and C infers the money-related cost of the resource per time unit. The price of each resource is depicted in Table 3.

$$\text{Cost} = \sum_{\text{i}\in\text{resources}} \{C \times T\}. \quad (6)$$

Table 3: Price unit for each resource.

Number of nodes	2	4	5	3	7	6	8	1
Price unit for each operation	0.8	0.7	0.2	0.9	0.6	0.4	0.4	0.2

6.2. Scheduling Models Performance. Cloud suppliers possess a tremendous number of servers and other handling establishments. An enormous number of virtual machines run inside a server so the resources can be utilized in the best manner. These computations observe the tasks and their needs and attend to them effectively. Optimization scheduling techniques were contrasted like round robin (RR), particle swarm optimization (PSO), first come first serve (FCFS), genetic simulated annealing (GASA), and shortest job first (SJF). To guarantee consistency, the models executed in this work utilized a comparable proportion of tasks with various lengths. The execution of various scheduling computations was finished by using IoMT-cloud tasks. Additionally, when we played out the proposed HGA, the model beats other models concerning the QoS. Also, because of the separation in the technical process, the outcomes were gainful for each model. The authors added one more parameter in contrast to their previous work.

Based on the result, it can be said that the throughput with the HGA is ideal. Table 4 shows the relationship between every one of the models against the embraced parameters. FCFS incorporates little execution time, little fulfillment time, and little holding up time as short cycles hang tight for more expanded ranges. GASA and HGA scheduling give time-sharing limits. With medium holding uptime, for more modest cycles, it is not recommended where fragile traffic is incorporated. SJF is sensible for basically a wide range of circumstances. These outcomes approve our proposed methodology towards getting a proficient model. As in the exploration, it shows that the FCFS is one of the quickest with regard to the execution for the traditional model; however, this standard oddball the waiting time, with this it can prompt terminations of assignments because of the period the patients need to pause. The other AI techniques were separable and indispensable, and they certainly achieved great results, but the best is still being our proposed model. The proposed parameters are truly outstanding in defending how the models will be performed. The planning is executed with the goal that it stops after a time period is achieved. Hence, the proposed model addresses the issue by giving the best waiting time and execution time. In the approaching passage, we will examine the outcomes further with a more pictorial view. Subsequently, we can close by saying our best model being the HGA has an effective QoS.

6.3. Experimental Result Discussion and Comparison. Figure 6 displays a connection of all utilized techniques against the TFT, TWT, and the TET, and these are some of the used validation criteria's to legitimize how effective the proposed technique is. These are profoundly considered while planning to achieve a higher QoS. The outcomes demonstrate that we can achieve the most extreme utilization of assets. In RR, every task gets an identical instance of time, but there are certain circumstances where a typical waiting time may be a problem as depicted in the results. The outcome was examined utilizing similar information to look at the presentation of the calculation. The traditional model has the most waiting time after streamlining, despite their advantage of speed, while the other compared AI models were also efficient but not to the proposed model. Subsequently, our proposed HGA model beats all other techniques which are our standard. In addition, the proposed technique provides the least execution time, and this makes the task execution faster when contrasted with other techniques. Thus, the waiting time should be minima so users can dodge task terminations. This can decide the reasonableness of each task and the technique to use in the ideal opportunity for planning a scheduling process in the IoMT-cloud. The completion time of our model beats different models, conversely, with the way that different models have a higher completion time.

While Figure 7 depicts the throughput relationship between each technique, it shows that the best technique is the proposed HGA model with the best throughput. After a set of several tasks, endeavors were carried out to amplify the throughput. The throughput is certain to be one of the most significant criteria to depict the presence of a cycle for each time unit. The throughput outcome shows how effective the proposed model is. Thus, each assignment was divided in their tenth to depict the exhibition. During this process, our model outperforms other models in this event, during the split. Although the optimization techniques were linearly separable, they showed their efficiency. Regardless, the proposed model was the best. The throughout is the biggest amount of errands that can be finished per time unit; with this, we can conclude that the proposed model outflanks other models and fits this description well.

Figure 8 shows the relationship of usage of resources for the scheduling techniques. Moreover, the HGA utilizes the resources that are free in the run time and pick another request. In this way, the inactive waiting time is diminished in the proposed HGA calculation contrasting other techniques. Likewise, asset usage is improved separately. Nonetheless, when different assets can be used, then others can become ideal. The resource used is looked at under various total number of the makespan. The techniques have an increase in the resource utilized and thereby staying in a normal state. Regarding resource size, or amount of task increments, there is a normal increase in the normal waiting time. Thus, it can be reasoned that the HGA is most effective as opposed to the other contrasted techniques. From the figure, we can derive the effectiveness of various techniques as opposed to the proposed technique, the normal asset utilized by various techniques remains practically comparable, and

TABLE 4: Cumulative results of all models.

Traits	SJF	FCFS	RR	PSO	GASA	HGA
Total execution time	55.36	54.68	54.31	40.21	36.22	32.57
Throughput	0.72	0.73	0.74	1.01	0.99	1.21
Total waiting time	42.21	41.72	40.92	40.80	40.30	40.16
Total finish time	101.67	100.18	99.31	80.10	79.4	76.6
Availability	Success/40	Success/40	Success/40	Success/40	Success/40	Success/40
Cost	0.27	0.27	0.27	0.20	0.20	0.17
Resource utilization	0.42	0.42	0.4	0.61	0.63	0.69
Energy efficiency	0.60	0.62	0.55	0.35	—	0.30

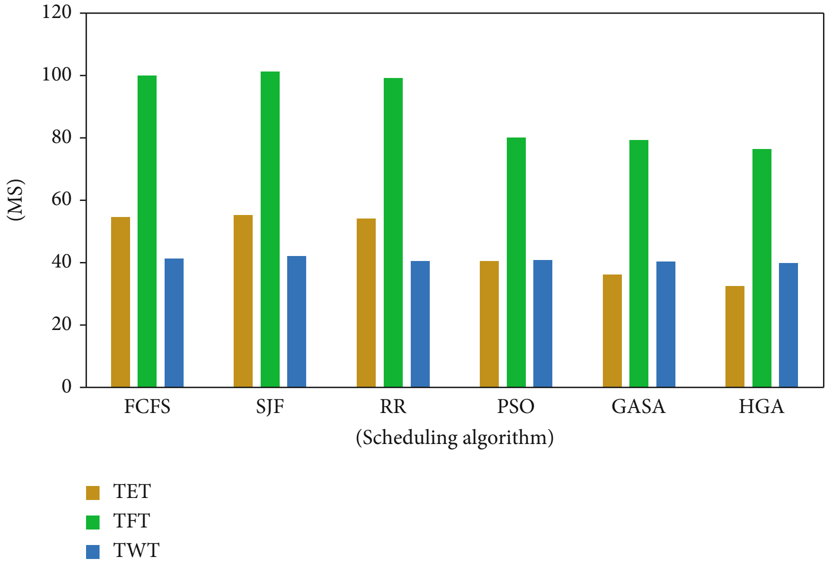


FIGURE 6: TET, TWT, and TFT of the scheduling model.

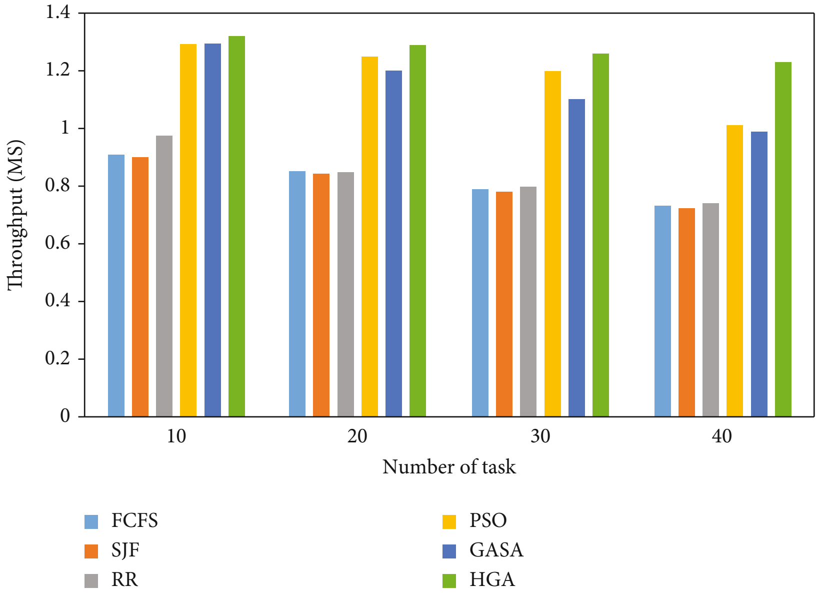


FIGURE 7: Result of the throughput.

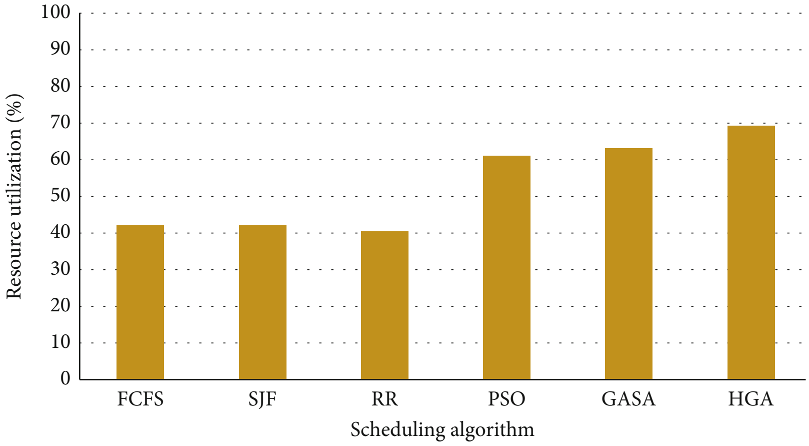


FIGURE 8: Scheduling models vs resource utilization.

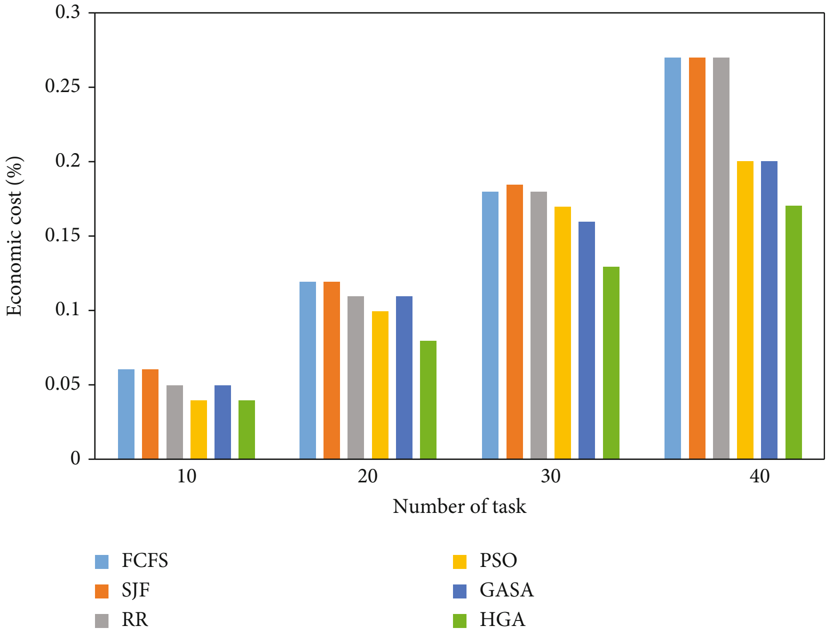


FIGURE 9: Scheduling models vs economic cost.

that signifies that it is affected by the quantity of accessible resources.

Figure 9 shows the financial expense factor. The result gotten shows the HGA model per task as a lesser cost factor. Cost depicts the impact of the charged rate over the used resource for each task. It is an evaluating factor for each center point in the IoMT-cloud environment. This impeding benefit makes it more interesting for users without the sensation of fear regarding being cheated. The HGA technique depicts a significant advantage where the rate was on a similar worth per each interaction. It is set at a level rate for each amount of resources, where making it to a reasonably higher worth will decrease the chances of the resource being picked for an undertaking. Thus, the outcome tells the best way to tackle this issue with the proposed HGA model that is to limit the expense massively. By and by, this will not suit the client's models as the usage in the clinical environment will need a lot of useful time and resources which will expand the expense separately. We can conclude by expressing that the HGA technique outperforms other techniques and as a base conservative expense differentiation to other contrasted models.

Figure 10 depicts the efficiency of energy consumption. Initially, the energy was calculated in KWh and later transformed to percentage to analyze its efficiency thoroughly. The parameter aims to reduce the consumption of energy. The figure shows that the HGA outperformed other models with a 30% reduced rate of energy consumption. The proposed model is compared to other metaheuristics, and it showed how efficient the model is. The PSO and GA seemed a bit fair in contrast, but the model is still lagging with regard to the efficiency of the energy consumption. This parameter,

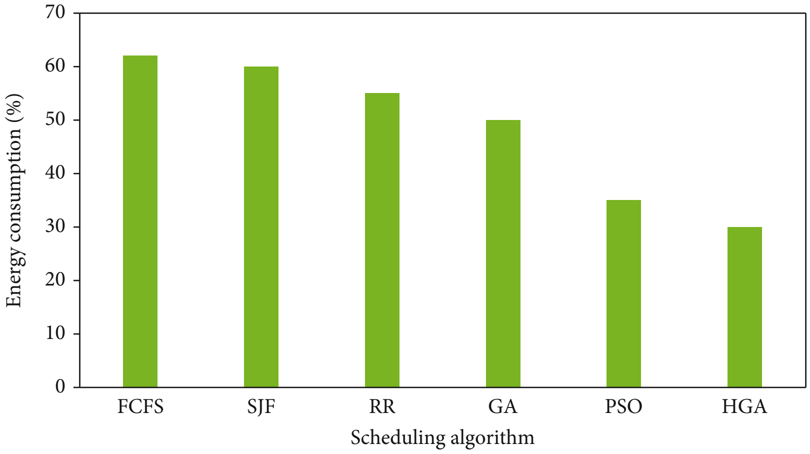


Figure 10: Efficiency of energy consumption comparison with HGA.

being one of the relevant parameters, will increase machine performance and at the same time will create a greener environment. The experimented known customary technique is known for its adequacy but could not outperform the proposed HGA.

Moreover, how the tested model will assume an urgent part in the clinical field where assets are utilized continuously is an eminent concern. This shows that cloud providers are expected to accomplish maxima income while thinking about QoS and solicitations from the clients. The health-care framework can be digitalized to achieve proficient association of medical care assets and administrations. With this, clinical information can be gathered, investigated, and observed. In this manner, the preliminaries show that the HGA beats different models and can be an effective method of planning for IoMT-cloud in the clinical field. Cloud clients or patients can have answer approaching solicitations without the apprehension about a task being dissolved or terminated.

7. Conclusion

The goal of this work is to style a model with a different sequency in contrast to the authors' previous work to solve task scheduling issues while at the same time-sharing resources to reach a productive QoS. The proposed work puts forth the significance of task scheduling computations and the application of AI in the IoMT-cloud environment. As we likely know, the IoMT-cloud is perhaps verifiably the most invigorating point for researchers, industry, and public zone. Thus, this theory targets developing a fast, sharp, and particular structure for task scheduling for IoMT-cloud. This evaluation put forth relies upon utilizing the present-day developments to further develop research on IoMT-cloud. This work in like manner presents an overall report between the scheduling techniques in IoMT-cloud, like the SJF, FCFS, RR, PSO, and GASA, and the proposed model being the HGA. Several parameters were used and added in contrast to the authors previous works like the TWT, TFT, TFT, resource utilization, throughput, and efficiency of energy consumption and cost. This work was reauthorizing the proposed assessment with various scheduling techniques to display the ampleness of the HGA. This study gives a potential guide for clients and experts in understanding task scheduling in the cloud. From the diagrams and calculations, it was exhibited that the HGA beat other different models concerning execution time, resource utilization, throughput, and cost. The process and experiment were executed on CloudSim, which is used for showing the different scheduling process in cloud computations. The proposed HGA had an execution pace of 32.57 ms and a throughput of 1.21 ms. These two parameters are one of the most significant parameters as it satisfies both client and provider's desires against the QoS. The charts and results portray that the HGA is far better than other optimization models even with the change in sequence when veered from the cases of TWT, TET, and TFT. HGA technique can be used in IoMT-cloud as significant task response time gets reduced reasonably. Has IoMT requires high execution speed and is time-dependent. However, future examination is to be considered like completing the computation for other progression factors like speedup and stream time. In future works, it can also lessen the cost and increase the throughput with the computations to get more smoothed out results. Finally, we will update the work using several other characteristics too and will bring the outcomes as they will appear, experimenting more AI models like bee algorithm and Ant algorithm. It is acknowledged that this endeavor will help specialists and researchers whenever considered.

Data Availability

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

Conflicts of Interest

The author does not have any possible conflicts of interest.

References

[1] M. Armbrust, A. Fox, R. Griffith et al., “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. de Rose, and R. Buyya, “CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms,” *Software: Practice and Experience*, vol. 41, no. 1, pp. 23–50, 2011.

[3] Z. Zongbin and D. Zhongjun, “Improved GA-based task scheduling algorithm in cloud computing,” *Computer Engineering and Applications*, vol. 49, no. 5, pp. 77–80, 2013.

[4] Z. Lijuan and W. Chunying, “Cloud computing resource scheduling in mobile internet based on particle swarm optimization algorithm,” *Computer Science*, vol. 42, no. 6, pp. 279–292, 2015.

[5] X. Jie, Z. Jian-chen, and L. Ke, “Task scheduling algorithm based on dual fitness genetic annealing algorithm in cloud computing environment,” *Journal of University of Electronic Science and Technology of China*, vol. 42, no. 6, pp. 900–904, 2013.

[6] W. Fang, L. Mei'an, and D. Weijun, “Cloud computing task scheduling based on dynamically adaptive ant colony algorithm,” *Journal of Computer Applications*, vol. 33, no. 11, pp. 3160–3162, 2013.

[7] F. Al-Turjman and D. Deebak, “Privacy-aware energy-efficient framework using the internet of medical things for COVID-19,” *IEEE Internet of Things Magazine*, vol. 3, no. 3, pp. 64–68, 2020.

[8] H. Chen, X. Zhu, G. Liu, and W. Pedrycz, “Uncertainty-aware online scheduling for real-time workflows in cloud service environment,” *IEEE Transactions on Services Computing*, vol. 14, no. 4, pp. 1167–1178, 2021.

[9] H. Chen, X. Zhu, H. Guo, J. Zhu, X. Qin, and W. Jianhong, “Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment,” *Journal of Systems and Software*, vol. 99, pp. 20–35, 2015.

[10] R. G. Reynolds, Z. Michalewicz, and M. Cavaretta, “Using cultural algorithms for constraint handling in GENOCOP,” in *Procceding of the 4th Annual Conference on Evolutionary Programming*, pp. 298–305, MIT Press, Cambrige, MA, USA, 1995.

[11] A. A. Hussain, O. Bouachir, F. Al-Turjman, and M. Aloqaily, “AI techniques for COVID-19,” *IEEE Access*, vol. 8, pp. 128776–128795, 2020.

[12] A. A. Hussain and F. Al-Turjman, “Resource allocation in volunteered cloud computing and battling COVID-19,” in *AI-Powered IoT for COVID-19*, pp. 39–76, CRC Press, 2020.

[13] A. Al-maamari and F. Omara, “Task scheduling using PSO algorithm in cloud computing environments,” *International Journal of Grid Distribution Computing.*, vol. 8, no. 5, pp. 245–256, 2015.

[14] F. Al-Turjman, A. A. Hussain, S. Alturjman, and C. Altrjman, “Vehicle Price Classification and Prediction Using Machine Learning in the IoT Smart Manufacturing Era,” *Sustainability*, vol. 14, no. 15, p. 9147, 2022.

[15] M. Mezmaz, N. Melab, Y. Kessaci et al., “A parallel bi-objective hybrid metaheuristic for energy-aware scheduling for cloud computing systems,” *Journal of Parallel and Distributed Computing*, vol. 71, no. 11, pp. 1497–1508, 2011.

[16] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (iot): a vision, architectural elements, and future directions,” *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[17] S. Rose, “The IoT trends that no one has spoken about-read this now,” 2019, https://towardsdatascience.com/top-14-iot-trends-to-expect-in-2020-fa81a56e8653.

[18] Y. Kotb, I. Al Ridhawi, M. Aloqaily, T. Baker, Y. Jararweh, and H. Tawfik, “Cloud-based multi-agent cooperation for IoT devices using workflow-nets,” *Journal of Grid Computing*, vol. 17, no. 4, pp. 625–650, 2019.

[19] P. Sangwan, M. Sharma, and A. Kumar, “Improved round robin scheduling in cloud computing,” *Advances in Computational Sciences and Technology*, vol. 10, pp. 639–644, 2017.

[20] S. Oueida, Y. Kotb, M. Aloqaily, Y. Jararweh, and T. Baker, “An edge computing based smart healthcare framework for resource management,” *Sensors*, vol. 18, no. 12, p. 4307, 2018.

[21] M. Al-Khafajiy, T. Baker, H. Al-Libawy, Z. Maamar, M. Aloqaily, and Y. Jararweh, “Improving fog computing performance via fog-2-fog collaboration,” *Future Generation Computer Systems*, vol. 100, pp. 266–280, 2019.

[22] J. Li, T. Ma, M. Tang, W. Shen, and Y. Jin, “Improved FIFO scheduling algorithm based on fuzzy clustering in cloud computing,” *Information*, vol. 8, no. 1, p. 25, 2017.

[23] S. Elmougy, S. Sarhan, and M. Joundy, “A novel hybrid of shortest job first and round robin with dynamic variable quantum time task scheduling technique,” *Journal of Cloud Computing*, vol. 6, no. 1, p. 12, 2017.

[24] D. I. Arkhipov, W. Di, W. Tao, and A. C. Regan, “A parallel genetic algorithm framework for transportation planning and logistics management,” *Access IEEE*, vol. 8, pp. 106506–106515, 2020.

[25] Y. Gan, C. Yin, Q. Fan, and A. Li, “Improved T-matrix method for simultaneous reconstruction of dielectric and perfectly conducting scatterers,” *Access IEEE*, vol. 8, pp. 143622–143631, 2020.

[26] J.-T. Tsai, J.-C. Fang, and J.-H. Chou, “Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm,” *Computers and Operations Research*, vol. 40, no. 12, pp. 3045–3055, 2013.

[27] S. T. Maguluri and R. Srikant, “Scheduling jobs with unknown duration in clouds,” *IEEE/ACM Transactions On Networking*, vol. 22, no. 6, pp. 1938–1951, 2014.

[28] C. Cheng, J. Li, and Y. Wang, “An energy-saving task scheduling strategy based on vacation queuing theory in cloud computing,” *Tsinghua Science and Technology*, vol. 20, no. 1, pp. 28–39, 2015.

[29] W. Lin, C. Liang, J. Z. Wang, and R. Buyya, “Bandwidth-aware divisible task scheduling for cloud computing,” *Software: Practice and Experience*, vol. 44, no. 2, pp. 163–174, 2014.

[30] D. Ergu, G. Kou, Y. Peng, Y. Shi, and Y. Shi, “The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment,” *The Journal of Supercomputing*, vol. 64, no. 3, pp. 835–848, 2013.

[31] X. Zhu, L. T. Yang, H. Chen, J. Wang, S. Yin, and X. Liu, “Real-time tasks oriented energy-aware scheduling in virtualized clouds,” *IEEE Transactions on Cloud Computing*, vol. 2, no. 2, pp. 168–180, 2014.

[32] X. Liu, Y. Zha, Q. Yin, Y. Peng, and L. Qin, "Scheduling parallel jobs with tentative runs and consolidation in the cloud," *Journal of Systems and Software*, vol. 104, pp. 141–151, 2015.

[33] G. Shamsollah and M. Othman, "Priority based job scheduling algorithm in cloud computing," *Procedia Engineering*, vol. 50, pp. 778–785, 2012.

[34] M. A. Rodriguez and R. Buyya, "Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds," *IEEE Transactions on Cloud Computing*, vol. 2, no. 2, pp. 222–235, 2014.

[35] M. Polverini, A. Cianfrani, S. Ren, and A. V. Vasilakos, "Thermal aware scheduling of batch jobs in geographically distributed data centers," *IEEE Transactions on Cloud Computing*, vol. 2, no. 1, pp. 71–84, 2014.

[36] A. E. Keshk, A. B. El-Sisi, and M. A. Tawfeek, "Cloud task scheduling for load balancing based on intelligent strategy," *International Journal of Intelligent Systems & Applications*, vol. 6, no. 5, p. 25, 2014.

[37] S. Ghanbari, M. Othman, W. J. Leong, and M. R. A. Bakar, "Multi-criteria-based algorithm for scheduling divisible load," in *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*, pp. 547–554, Singapore, 2014.

[38] H. Goudarzi, M. Ghasemazar, and M. Pedram, "Sla-based optimization of power and migration cost in cloud computing," in *2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012)*, pp. 172–179, Ottawa, ON, Canada, 2012.

[39] F. Al-Turjman and S. Alturjman, "5G/IoT-enabled UAVs for multimedia delivery in industry-oriented applications," *Multimedia Tools and Applications*, vol. 79, no. 13-14, pp. 8627–8648, 2020.

[40] S. A. Alabady, F. Al-Turjman, and S. Din, "A novel security model for cooperative virtual networks in the IoT era," *International Journal of Parallel Programming*, vol. 48, no. 2, pp. 280–295, 2020.

[41] F. Al-Turjma, "Intelligence and security in big 5G-oriented IoNT: an overview," *Future Generation Computer Systems, Volume*, vol. 102, pp. 357–368, 2020.

[42] S. Ghanbari, M. Othman, M. R. A. Bakar, and W. J. Leong, "Priority-based divisible load scheduling using analytical hierarchy process," *Applied Mathematics & Information Sciences*, vol. 9, no. 5, pp. 25–41, 2015.

[43] B. Radojevic and M. Zagar, "Analysis of issues with load balancing algorithms in hosted (cloud) environments," in *2011 Proceedings of the 34th International Convention MIPRO*, pp. 416–420, Opatija, Croatia, 2011.

[44] S. Ghanbari, M. Othman, M. R. A. Bakar, and W. J. Leong, "Multi-objective method for divisible load scheduling in multi-level tree network," *Future Generation Computer Systems*, vol. 54, pp. 132–143, 2016.

[45] S. Goswami and A. Das, "Optimization of workload scheduling in computational grid," in *Proceedings of the 5th international conference on Frontiers in intelligent computing: theory and applications*, pp. 417–424, Odisa, india, 2017.

[46] K. Kaur and A. Kaur, "A hybrid approach of load balancing through VMs using ACO, MinMax and genetic algorithm," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pp. 615–620, Dehradun, India, 2016.

[47] M. S. Pilavare and A. Desai, "A novel approach towards improving performance of load balancing using genetic algorithm in cloud computing," in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–4, Coimbatore, India, 2015.

[48] R. R. Patel, S. J. Patel, D. S. Patel, and T. T. Desai, "Improved GA using population reduction for load balancing in cloud computing," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2372–2374, Jaipur, India, 2016.

[49] A. A. S. Farrag, S. A. Mahmoud, and E. S. M. El-Horbaty, "Intelligent cloud algorithms for load balancing problems: A survey," in *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 210–216, Cairo, Egypt, 2015.

[50] H. A. Makasarwala and P. Hazari, "Using genetic algorithm for load balancing in cloud computing," in *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1–6, Ploiesti, Romania, 2016.

[51] K. V. Kavitha and V. V. Suthan, "Dynamic load balancing in cloud-based multimedia system with genetic algorithm," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, pp. 1–4, Coimbatore, India, 2016.

[52] S. Dam, G. Mandal, K. Dasgupta, and P. Dutta, "Genetic algorithm and gravitational emulation-based hybrid load balancing strategy in cloud computing," in *Proceedings of the 2015 Third International Conference on Computer, Communication, Control and Information Technology (C3IT)*, pp. 1–7, Hooghly, India, 2015.

[53] L. Hong and G. Yufei, "GACA-VMP: virtual machine placement scheduling in cloud computing based on genetic ant Colony algorithm approach," in *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom)*, pp. 1008–1015, Beijing, China, 2015.

[54] A. A. Hussain and F. Al-Turjman, "Artificial intelligence and blockchain: a review," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 9, 2021.

[55] A. A. Hussain, B. A. Dawood, and F. Al-Turjman, "Application of AI techniques for COVID-19 in IoT and big data era: a survey," in *Artificial intelligence and machine learning for COVID-19*, pp. 175–211, Springer, 2021.

[56] A. A. Hussain, F. Al-Turjman, and M. Sah, "Semantic web and business intelligence in big-data and cloud computing era," in *The Proceedings of the Third International Conference on Smart City Applications*, pp. 1418–1432, Turkey, 2021.

[57] A. A. Hussain, F. Al-Turjman, E. Gemikonakli, and Y. K. Ever, "Design of a navigation system for the blind/visually impaired," in *International Conference on Forthcoming Networks and Sustainability in the IoT Era*, pp. 25–45, Cyberspace, 2021.

[58] A. A. Hussain and K. Dimililer, "Student grade prediction using machine learning in Iot era," in *International Conference on Forthcoming Networks and Sustainability in the IoT Era*, pp. 65–81, Cyberspace, 2021.

[59] A. A. Hussain, B. A. Dawood, and F. Al-Turjman, "IoT and AI for COVID-19 in scalable smart cities," in *International Summit Smart City 360°*, pp. 3–19, Springer, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5167639, 9 pages
https://doi.org/10.1155/2022/5167639

Research Article

Low-Power Communication Signal Enhancement Method of Internet of Things Based on Nonlocal Mean Denoising

Mingchuan Tian[1] **and Jizheng Liu**[2]

[1]*Nanyang Technological University, 50 Nanyang Ave, Singapore 639798*
[2]*Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Jizheng Liu; liujz_ev@bit.edu.cn

Received 29 April 2022; Accepted 28 June 2022; Published 30 July 2022

Academic Editor: Maode Ma

In order to improve the transmission effect of low-power communication signal of Internet of Things and compress the enhancement time of low-power communication signal, this paper designs a low-power communication signal enhancement method of Internet of Things based on nonlocal mean denoising. Firstly, the residual of one-dimensional communication layer is preprocessed by convolution core to obtain the residual of one-dimensional communication layer. Then, according to the two classification recognition methods, the noise reduction signal feature recognition of the low-power communication signal of the Internet of Things is realized, the nonlocal mean noise reduction algorithm is used to remove the low-power communication signal of the Internet of Things, and the weight value between similar blocks is calculated according to the European distance method. Finally, the low-power communication signal enhancement of the Internet of Things is realized by the nonlocal mean value denoising method. The experimental results show that the communication signal enhancement time overhead of this method is low, which is always less than 2.6 s. The lowest bit error rate after signal enhancement is about 1%, and the signal-to-noise ratio is up to 18 dB, which shows that this method can achieve signal enhancement.

1. Introduction

In recent years, cognitive radio signal communication technology has shown great vitality and market potential in a large range of the world. However, in the process of signal transmission, due to the multipath effect in the channel, the limitation of channel transmission bandwidth, and the imperfection of channel transmission characteristics, the signal will be seriously affected by all kinds of noise and electromagnetic interference, which will directly degrade the quality of communication and affect the reception and analysis of the signal at the receiver [1, 2]. For example, reduce spectrum sensing, and increase demodulation difficulty. Only by obtaining high-quality signals, further analysis and processing are meaningful. Therefore, removing noise and interference in communication signals, that is, signal enhancement, is an important technology to promote the development of wireless communication field. Therefore, researchers have proposed many signal enhancement methods, which can be divided into linear method and nonlinear method [3]. The linear signal enhancement method is relatively simple, but it still has low performance for nonlinear signals and cannot find the global optimal solution for noise elimination. In addition, because these methods are based on the assumption that the signal is stationary, their effectiveness is generally acceptable, and the actual signal usually has nonstationary statistical characteristics [4–6]. Nonlinear methods have become a research hotspot in recent two decades because they can clarify the spectrum and time information in the signal at the same time. However, the traditional signal enhancement methods use the prior information of noise or interference to map to the separable transform domain for separation, such as band-pass filtering. However, due to the random characteristics of noise and interference, the artificial construction of the corresponding separable transform domain often has a strong a priori to noise and interference. Therefore, at present, there is no more complete method to adaptively enhance the signal of time-varying system. For this reason, relevant scholars have conducted comparative research and made some progress.

Nan et al. proposed a satellite navigation signal enhancement method in the tunnel based on virtual satellite [7]. By establishing the signal propagation model, simulate the navigation signal of low elevation satellite located in the extension direction of the tunnel at both ends of the tunnel and receive the solution in the tunnel after sending it to realize positioning. At the same time, the signal delay control method is used to precompensate the pseudorange error. Through the experimental analysis of the hardware system, its positioning ability in the tunnel is verified. The simulation results show that the positioning accuracy of the system meets the needs of most tunnels. Pengyu et al. proposed an LPI radar signal enhancement method based on dae-gan network [8], combined with the advantages of noise reduction self-encoder and generation of countermeasure network, constructed a noise enhancement network and signal enhancement network for countermeasure training. The noise enhancement network doped more complex noise components into the noisy signal, and the signal enhancement network reduced the noise components in the noisy signal as much as possible. In this paper, dae-gan network is used to realize complex high-dimensional noise reduction. This method can effectively improve the effect of signal enhancement, but the signal enhancement takes a long time.

To solve the above problems, this paper designs a low-power communication signal enhancement method of the Internet of Things based on nonlocal mean denoising. The low-power communication signal of the Internet of Things is preprocessed through the deep residual network in deep learning, and the low-power communication signal enhancement of the Internet of Things is realized according to the nonlocal mean denoising method.

2. Low-Power Communication Signal Preprocessing of Internet of Things Based on Deep Learning

2.1. Depth Residual Network Analysis. Deep neural network always has the problem of network degradation. Generally speaking, with the increase of the depth of the network, the classification performance of the network should be stronger and stronger. However, the current situation is that the performance of the network gradually tends to saturate with the rise of the network depth but will decline rapidly after the depth rises to a certain node. This phenomenon is called network degradation [9]. For the network whose accuracy is close to saturation, if the identity mapping layer is added in the network structure to make its input equal to the output, it will not increase the error after propagation while continuing to increase the network depth, that is, deepening the network through this method will not increase the training error. The design method of residual network comes from this [10–13]. The residual network is proposed mainly to reduce the network degradation caused by the rise of network depth. The solution is to construct a "residual element." The structure of residual element is shown in Figure 1.

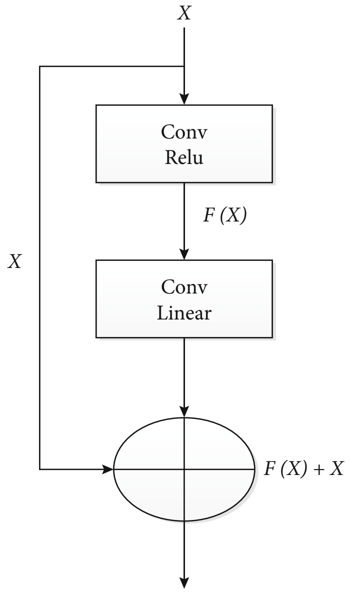
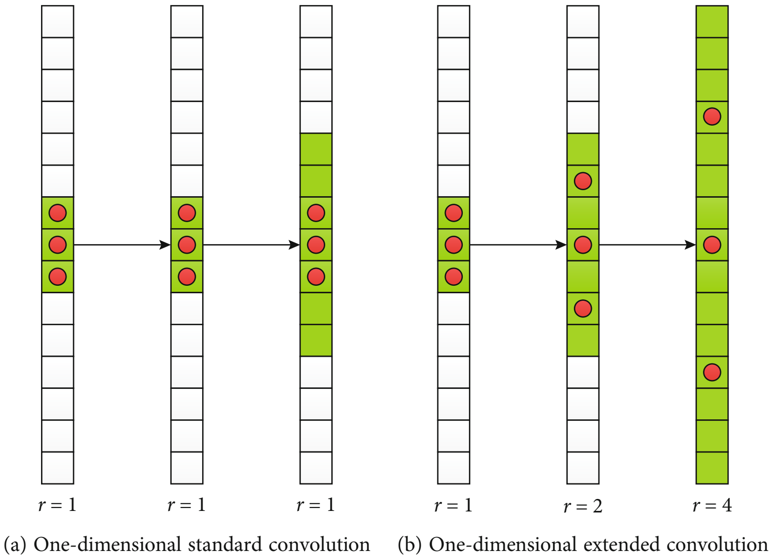


Figure 1: Residual element.

For a neural network, suppose that the input of a certain section in the network is X and its corresponding expected output is $Z(X)$, that is, $Z(X)$ is the expected potential mapping. Generally, when the network is deepened, the training difficulty will increase [13]. In the residual network structure diagram in the figure above, a path from input to output is added on the basis of network mapping, the input X is directly transferred to the output as the initial result, and the output result is $Z(X) = F(X) + X$. When $F(X) = O$, there is $Z(X) = X$, that is, the identity mapping that will not increase the error mentioned earlier [14]. RESNET changes the learning strategy. For an input X, it no longer learns its expected mapping $Z(X)$ directly through the convolution layer [15] but learns the expected residual mapping through the network, that is, $Z(X) - X$. Compared with the expected mapping, the residual mapping is easier to optimize and deepens the network depth on the premise of avoiding network degradation.

2.2. Impulse Noise Preprocessing. The INP used in this paper can be regarded as the neutralization of truncation and zeroing in the threshold suppression method. The difference lies only in the processing of the part whose amplitude is higher than the threshold. Simulation experiments show that it has better pulse suppression effect than the two methods. The main task of INP is to suppress the nonlinear part of the received signal $y(t)$ whose amplitude is greater than the threshold τ_r. The output signal can be expressed as follows:

$$y_{\text{non}}(n) = \begin{cases} y(n), & |y(n)| \le \tau_r, \\ y(n)\left(\dfrac{\tau_r}{|y(n)|}\right)^2 & |y(n)| > \tau_r, \end{cases} \tag{1}$$



(a) One-dimensional standard convolution (b) One-dimensional extended convolution

FIGURE 2: Comparison of receptive field between one-dimensional standard convolution layer and expanded convolution layer.

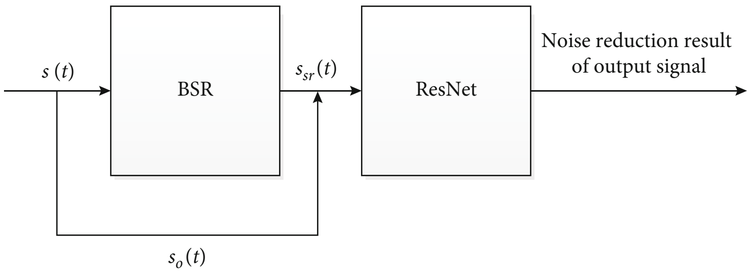


FIGURE 3: Signal denoising algorithm processing flow.

wherein τ_r can be obtained from the following formula:

$$\tau_r = (1 + 2\tau_0)\tau_Q, \quad (2)$$

where τ_0 is the constant coefficient, set to 1.5, and τ_Q is the second quartile value of the received signal $y(n)$ modulus $|y(n)|$ [16]. After pulse suppression, the signal needs to be further normalized to obtain the final output signal of INP, that is, the input of RCGAN noise reduction network:

$$y_p(n) = \frac{y_{\text{non}}(n)}{\max\left(|y_{\text{non}}(n)|\right)}. \quad (3)$$

In order to improve the ability of the network to retain the signal details in the process of noise reduction, the middle layer of the generator adopts the extended convolution structure widely used in the field of image semantic segmentation [17]. This structure expands the receptive field of the convolution kernel by inserting zeros in the convolution kernel. Compared with the receptive field expansion methods such as downsampling used in the standard convolution structure, this method has stronger retention ability of detailed information. Generally, the expansion rate is used to represent the interval of adjacent elements in the convolution kernel, and the expansion rate of the standard convolution kernel is 1. The expansion rate r of three one-dimensional expansion convolution layers in RCGAN generator is taken as 1, 2, and 4, respectively, and the length of convolution core is 3. The perceived field of view of convolution cores in different layers can be expressed as follows:

$$F_r = 2^{r+1} - 1. \quad (4)$$

Based on this, it can be calculated that the convolution kernel receptive field sizes of the three one-dimensional expanded convolution layers are 3, 7, and 15, respectively.

Figure 2 compares the change process of receptive visual field of three one-dimensional standard convolution layers and three one-dimensional expanded convolution layers when the convolution kernel length is 3.

The solid circle in Figure 2 represents the position of nonzero value in the convolution core, and the solid square represents the receptive field of view of the convolution core. The receptive field of vision showed a linear growth trend, while the receptive field of vision showed a linear growth trend. The research shows that expanded convolution achieves better retention of small-scale information features in semantic segmentation by virtue of this information lossless sensory field expansion method. Therefore, we use this structure in the middle layer of the generator to improve the network's ability to extract signal details.

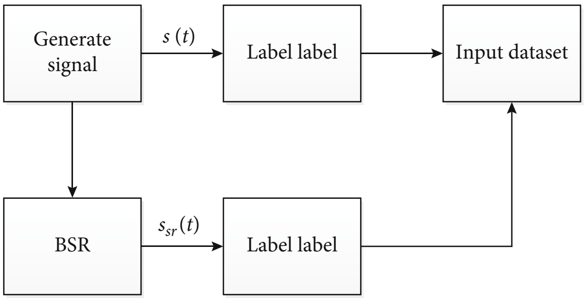


Figure 4: Data set construction method.

2.3. Signal Generation Method. The flow chart of signal noise reduction algorithm generated by low-power communication of Internet of Things used in this paper is shown in Figure 3.

The first part is the bistable stochastic resonance (BSR) system. After the input signal $s_{sr}(t)$ passes through the system, the output signal $s_{sr}(t)$ is obtained. Then, the data set is written together with the nonresonant signal $s_o(t)$ and input into the depth residual network (RESNET), and finally, the noise reduction result is output. After the modulated signal is generated, the bistable coefficient is determined, and then, the signal through stochastic resonance is output through the BSR system. After each signal sample is generated, in order to make the signal waveform correspond one by one, after each original sample is generated, the sample is generated through the BSR system, and then, the samples before and after resonance are stored in the data set at the same time. And mark whether the signal resonates. The construction process is shown in Figure 4.

After the signal is generated, label the two types of signals, and then, input RESNET for noise reduction. After the signal noise reduction is completed, this paper continues to build a binary classification network C based on CNN to automatically extract the detection features used to characterize the presence or absence of the signal in the noise reduction signal and complete the binary classification recognition. With the characteristics of weight sharing and local perception field, CNN has obvious advantages in local feature extraction while greatly reducing the amount of model parameters. It has been widely used in many data classification problems. After INP preprocessing and effective noise reduction of Rcgan network, the noise components in the received signal are significantly suppressed, while the useful signal components will be well preserved. Then, input the signal waveform after noise reduction into the CNN classifier. Due to the significant improvement of signal-to-noise ratio, the difficulty of signal detection after noise reduction will be greatly reduced, and the false alarm rate will be significantly reduced.

In the structure setting of classifier C, this paper similarly adopts the step convolution structure to compress the feature dimension. Its overall structure is similar to D. The convolution layer has six layers, including 16, 32, 64, 128, 256, and one convolution core, respectively. At the end of the convolution layer, it is connected with the softmax classifier through a full connection layer to output the probability of communication signal and pure noise in the received signal, respectively. The setting of the activation function is the same as that of the decider. When (y_s, y_p, y_L) is used to represent the transmission data in a training sample, the INP preprocessed data and the real tag value after one hot coding, and the predicted tag probability vector output by the softmax classifier $y_L{}^\wedge$ can be expressed as follows:

$$\hat{y}_L = \begin{bmatrix} P\left(L_p = 0|\mathrm{Cout}\right) \\ P\left(L_p = 1|\mathrm{Cout}\right) \end{bmatrix} = \frac{1}{\sum_{i=0}^{1}\exp(\mathrm{Cout}_i)} \begin{bmatrix} \exp\left(\mathrm{Cout}_0\right) \\ \exp\left(\mathrm{Cout}_1\right) \end{bmatrix}, \tag{5}$$

$$Cout = f_{\mathrm{CNN}}\left(G\left(y_p\right)\right), \tag{6}$$

where L_p represents the predicted tag value, Cout represents the input vector of softmax, and f_{CNN} represents the nonlinear function formed by the network before softmax layer in classifier C.

3. Nonlocal Mean Denoising and Enhancement Method of Low-Power Communication Signal in Internet of Things

Using a large number of redundant information with similar structure in the natural image, make full use of the redundant information on these images to reduce the noise of the image [18]. When processing each pixel in the noisy image, the distribution around the pixel will be evaluated and compared, and the difference similarity of the distribution will be used to calculate the weight ω. The performance of nonlocal mean denoising algorithm is better than other traditional image denoising algorithms, and the denoising effect is better [19–21].

For a noisy image $Y(i) = X(i) + n(i)$, $X(i)$ is the original image and $n(i)$ is noise. Take the noise reduction of a pixel on a noisy image as an example. First, take this pixel as the center point, and then, create a neighborhood window [22]. Next, traverse the whole noisy image to find similar blocks with similar neighborhood window structure. Then, the weighted re equalization of these similar blocks is calculated, and the calculated pixels are the pixels after noise reduction [23–25]. The formula of nonlocal mean noise reduction algorithm is shown in formula (7):

$$NLM\left[\bar{X}\right](i) = \sum_{j\in I} w(i,j)Y(j). \tag{7}$$

Suppose that the noisy signal Y can be composed of pure signal s and additive interference d:

$$y = s + d. \tag{8}$$

Under the set conditions, y obtains the estimated value $\widehat{s}$ of s, which is the main principle of signal enhancement. In the form of short-time Fourier transform (STFT), $Y_{n,k}$ exp

$(j\alpha_n, k)$, $S_{n,k}$ exp $(j\varphi_n, k)$, and $\widehat{S}_{n,k}$ exp $(j\widehat{\varphi}_n, k)$ represent y, s, and $\widehat{s}$ in the nth frame, respectively, and the frequency sequence number is represented by $K = 1, 2, \cdots, K$. Without considering the phase information, the main purpose of signal enhancement task is to minimize its error function:

$$E_r = \sum_{k=1}^{K} \left(\widehat{S}_{n,k} - S_{n,k}\right)^2. \tag{9}$$

Let the amplitude spectrum vector and estimated value of the pure signal on the nth frame be represented by S_n and $\widehat{S}_n$, respectively. At this time, the function error can be expressed as follows:

$$E_r = \left|\left|\widehat{S}_n - S_n\right|\right|_2^2. \tag{10}$$

The research on the signal enhancement method of deep neural network hospital wireless communication network can be understood as follows: using the training parameter set θ to build a nonlinear function f_θ, which requires that the function f_θ must have cumbersome characteristics, so it is used to ensure the new error function:

$$E_r = ||f_\theta(X_n) - S_n||_2^2. \tag{11}$$

Minimum to obtain the target output

$$\widehat{S}_n = f_\theta(X_n), \tag{12}$$

where $X_n = [Y_{n-N}, Y_{n-N+1}, \cdots, Y_n, \cdots, Y_{n+N-1}, Y_{n+N}]$ is the training feature of the nth frame, which is composed of half $[(2n+1)$ frame$]$ amplitude spectrum vector of the nth frame, and $(2n+1)$ is the input length.

When looking for the pixel block of the neighborhood window, we usually determine the weight value $w(i, j)$ by calculating the Euclidean distance between similar blocks, and the Euclidean distance $d(i, j)$ between similar blocks is shown in the following formula:

$$d(i, j) = \left\| Y(N_i) - Y\left(N_j\right) \right\|_{2,a}^2. \tag{13}$$

The weight value $w(i, j)$ is calculated as shown in formula (14):

$$w(i, j) = \frac{1}{Y(i)} e^{-d(i,j)^2/h^2}. \tag{14}$$

h is the smoothing control parameter of nonlocal mean noise reduction, which has a very important impact on the final noise reduction effect. $\bar{X}$ is the image after nonlocal mean noise reduction. The one-dimensional off diagonal slice $c_{4x}(m)$ of the fourth-order cumulant of IoT low-power communication signal $x(n)$ can simultaneously resist Gaussian noise and maintain the basic framework of available signal $s(n)$. Therefore, based on the weak signal $x(n)$, the estimated data $\widehat{c}_{4x}(m)$ based on the one-dimensional off diagonal slice of the fourth-order cumulant $x(n)$ can be calculated, and then, $\widehat{c}_{4x}(m)$ pairs of fir (finite impulse response) filters can be used to create and collect the signal from the noise. The impulse response of FIR filter is defined as the following formula:

$$h(m) = \begin{cases} \widehat{c}_{4x}(L - m), m = 0, 1, \cdots, L, \\ \widehat{c}_{4x}(m - L), m = L + 1, L + 2, \cdots, 2L. \end{cases} \tag{15}$$

Among them, the maximum lag number is expressed in L, and the length of impulse response of FIR filter is expressed in $2L + 1$. Because $\widehat{c}_{4x}(m)$ is similar to $s(n)$ to some extent, it can be concluded that the FIR filter defined in the above formula is the relative matched filter of the available signal $s(n)$, and its output signal-to-noise ratio tends to be idealized. The output formula of the filter is as follows:

$$y(n) = \gamma \sum_{m=0}^{2L} h(m) x(n - m). \tag{16}$$

In the above formula, the gain factor is expressed as γ, which has the control function of filter output, and the value is usually the reciprocal of kurtosis coefficient [26, 27]. At this time, the output result of the filter is the enhanced low-power communication signal of the Internet of Things, which completes the enhancement of the low-power communication signal of the Internet of Things.

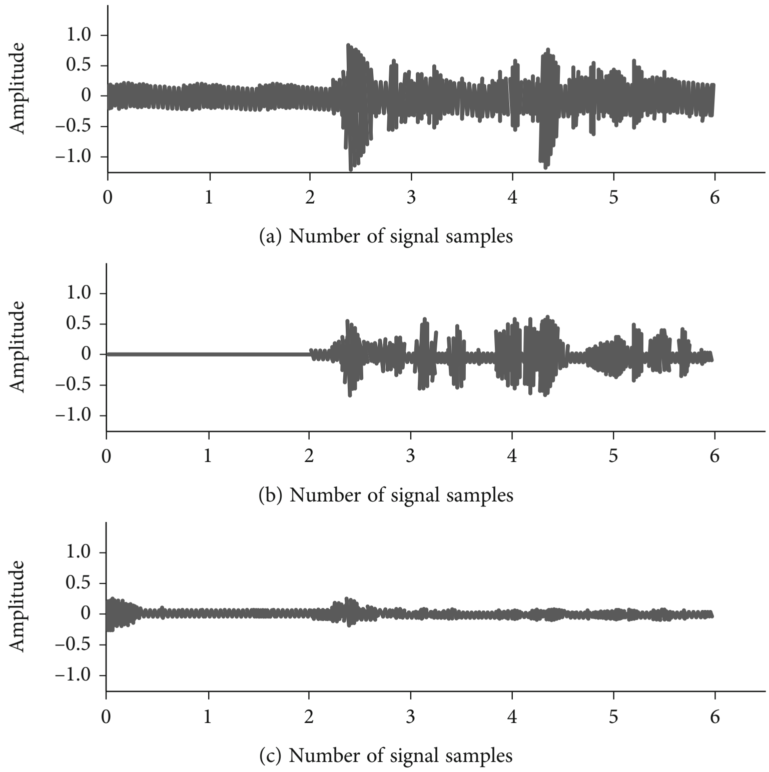
4. Experiment

4.1. Experimental Design. In reality, most of the environmental background noise is unstable signals. In order to ensure the enhancement effect of low-power communication signals of the Internet of Things, the noise needs to be adjusted in time. In this paper, MSP430 series single chip microcomputer with hardware multiplier is selected, so that it cannot be disturbed by CPU in the calculation process. The A/D converter adopts a 12 bits, 60 K ROM. The noise environment background of the simulation experiment is ≥−5 dB.

4.2. Analysis of Experimental Results

4.2.1. Time Domain Waveform Analysis of Enhancement Effect. Through the time domain waveform, the simulation results of this method and reference [7] method are analyzed and compared. The low-power communication signal acquisition frequency of the experimental Internet of Things is set to 8 kHz. In the low-power communication signal enhancement method of the Internet of Things in this paper, the filter order of the first two stages is 32, and the input signal of the first stage is divided into three output channels by the filter according to the frequency factor. The experimental results are shown in the following time domain waveform diagram, as shown in Figure 5.

The y-axis is the amplitude of the dimensionless processed signal, and the x-axis is the number of samples of the signal, about 60000. Figure 5(a) shows a signal with noise, and the interference intensity of the noise is very high. After calculation, the signal-to-noise ratio of the signal to the noise



(a) Number of signal samples

(b) Number of signal samples

(c) Number of signal samples

Figure 5: Time domain waveform.

Table 1: Time costs of signal enhancement by different methods.

Number of signals/piece	Paper method	Reference [7] method	Reference [8] method
1000	0.2	9.2	6.9
1500	0.8	16.2	15.2
2000	0.9	19.9	19.5
2500	1.2	22.6	23.1
3000	1.8	26.9	28.6
3500	2.0	32.1	33.0
4000	2.2	38.9	38.6
4500	2.3	42.9	40.8
5000	2.6	46.8	45.9

is 1.121 dB. Therefore, the weak signal of the original communication is very unclear and is submerged by the noise in a large range. Figure 5(b) shows the time-domain waveform of the method of reference [7]. Although the algorithm can cancel the output signal, the output signal-to-noise ratio is greatly reduced to 0.544 dB. It can be seen that this method is immune to noise and low-power communication signals of the Internet of Things, resulting in signal distortion. It cannot reduce noise or enhance the signal alone. Figure 5(c) shows the second stage output of the method in this paper. The signal-to-noise ratio of the signal is increased to about 10 times of the original signal, and its value is about 13.465 dB, which greatly enhances the low-power communication signal of the Internet of Things. At the same time, the noise is well suppressed and removed.

4.2.2. Signal Enhancement Time. The experiment further analyzes the time cost of low-power communication signal enhancement of the Internet of Things by the method in this paper, the method in reference [7], and the method in reference [8]. The results are shown in Table 1.

By analyzing the experimental data in Table 1, it can be seen that the time cost of IoT low-power communication signal enhancement has changed with the number of IOT low-power communication signals enhanced by the methods in this paper, reference [7], and reference [8]. Among them, the time cost of low-power communication signal enhancement of the Internet of Things in this method is low and always less than 2.6 s. The time cost of low-power communication signal enhancement of the Internet of Things in reference [7] method is low in the early stage, but with the increase of the number, the iteration time continues to increase. The time cost of low-power communication signal enhancement of the Internet of Things in reference [8] method is always higher than the first two methods. It can be seen that this method has shorter enhancement time and certain work efficiency.

4.2.3. Signal Enhancement Effect. In the experiment, firstly, the bit error rate after weak signal enhancement in wireless communication network under electromagnetic interference environment is analyzed, and the bit error rate after signal enhancement using this method, reference [5], method, and reference [6] method is compared. The results are shown in Figure 6.

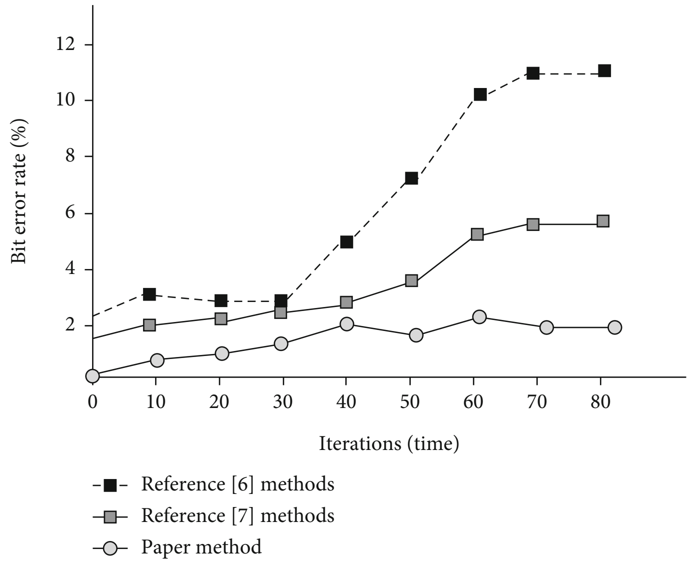


Figure 6: Bit error rate analysis of weak signal enhanced by different methods.

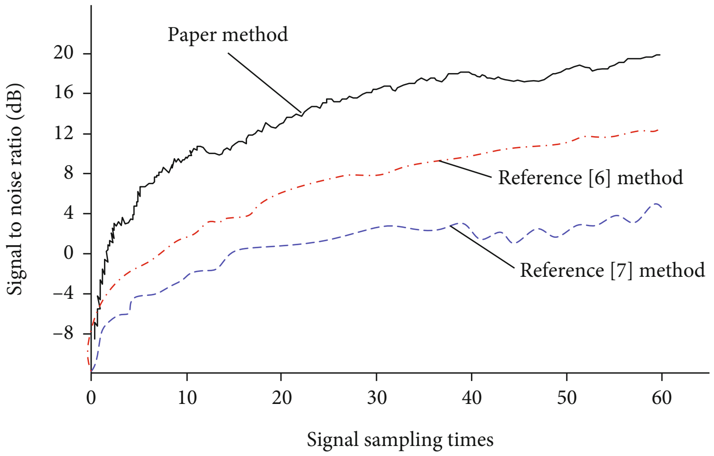


Figure 7: Comparison of signal-to-noise ratio under different methods.

By analyzing the experimental results in Figure 6, it can be seen that there are some differences in the bit error rate after signal enhancement using the methods in this paper, reference [7], and reference [8]. Among them, the bit error rate of this method after signal enhancement is lower than that of reference [8] and reference [7], and the fluctuation trend of its curve is small, the lowest is about 1%, while the experimental curve of bit error rate of the other two methods after signal enhancement fluctuates greatly and is always higher than that of this method, so it can be seen that the bit error rate of this method is lower. This is because this method takes into account the interference of electromagnetic wave during signal enhancement in communication, calculates the autocorrelation function between random signals, maintains the stability of the signal through the average of sample sequence time, adjusts the frequency offset and initial phase through low-pass filter by amplitude modulation coefficient, and estimates the amplitude of weak signal and the interference degree of signal by maximum likelihood. The signal enhancement is realized.

4.2.4. Signal to Noise Ratio. For the reference [7] method with good effectiveness and accuracy, the weak signal enhanced by this method has better effectiveness, lower distortion, and stronger noise reduction ability.

After the method in reference [7] is coherently averaged N times, the optimization degree of its signal-to-noise ratio will be increased by $\sqrt{N}$ times. This method cannot only idealize the noise reduction of the environmental background but also improve the optimization progress of signal-to-noise ratio, highlight the available signals, and enhance them adaptively. Figure 7 shows the trend change of signal-to-noise ratio in the enhancement stage, of which $N = 150$, $\mu = 5e - 6$.

According to the analysis of Figure 7, the signal-to-noise ratio is different under different methods. When the signal is sampled 10 times, the signal-to-noise ratio of the method in reference [7] is 2 dB, the signal-to-noise ratio of the method in reference [8] is 1 dB, and the signal-to-noise ratio of the method in this paper is 11 dB. When the signal is sampled

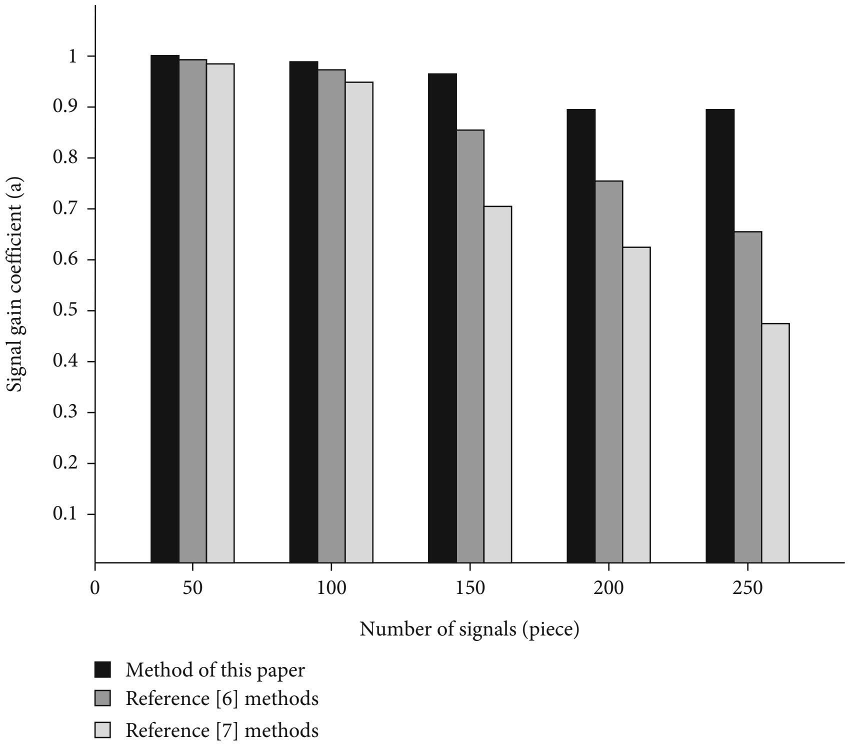


Figure 8: Test results of signal gain coefficient.

50 times, the signal-to-noise ratio of reference [7] method is 10 dB, the signal-to-noise ratio of reference [8] method is 2 dB, and the signal-to-noise ratio of this method is 18 dB. It can be seen from the above figure that the signal-to-noise ratio of this method continues to grow, and its curve trend has been higher than the other two traditional methods, indicating that this method has very significant advantages.

4.2.5. Signal Enhancement Stability Test. After the signal is enhanced, the signal gain coefficient can reflect the enhancement effect of the signal enhancement method. Set the gain coefficient of the signal as α and the optimal gain interval as [0,1]. The higher the test result of the signal gain coefficient in this interval, the better the enhancement effect of the signal, and vice versa. The methods proposed in this paper, reference [6], and reference [7] are used for signal enhancement, and the signal gain coefficients of the three methods are tested. The test results are shown in Figure 8.

As can be seen from Figure 8, the increase in the number of signals will reduce the gain coefficient after signal enhancement. It can be seen from the analysis of Figure 5 that the method proposed in this paper will reduce the test results with the increase of signals, but when the number of signals increases to a certain range, the method proposed in this paper can stabilize the test results within a certain coefficient. The test results of reference [7] method and reference [8] method are similar at the initial stage of the test, but with the progress of the test, the gap between them continues to widen. Finally, the test results of reference [7] method are much higher than those of reference [8]. It can be seen that the gain coefficient measured by the method proposed in this paper is high after signal enhancement, which shows that the enhancement effect of this method is good.

5. Conclusion

This paper designs a low-power communication signal enhancement method for the Internet of Things based on nonlocal mean denoising. Preprocess the low-power communication signal noise of the Internet of Things through the deep residual network, realize the noise reduction signal feature recognition of the low-power communication signal of the Internet of Things according to the two classification recognition method, calculate the weight value between similar blocks according to the Euclidean distance method, and enhance the low-power communication signal of the Internet of Things through the nonlocal mean denoising method. The experimental results show that the communication signal enhancement time overhead of this method is low, always less than 2.6 s, the lowest bit error rate after signal enhancement is about 1%, and the highest signal-to-noise ratio is 18 dB, which shows that this method has very significant advantages in signal enhancement.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] T. Minglei, Z. Wenpeng, J. Weidong, and G. Xunzhang, "Micro motion signal enhancement method based on multi-resolution saliency filtering," *Systems engineering and electronic technology*, vol. 44, no. 4, pp. 1148–1157, 2022.

[2] J. Liu, Z. Wang, and L. Zhang, "Integrated Vehicle-Following Control for Four-Wheel-Independent-Drive Electric Vehicles Against Non-Ideal V2X Communication," *in IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 3648–3659, 2022.

[3] Y. Ma, M. Gao, L. Wang, Y. Sha, W. Shao, and G. Shen, "Accuracy enhancement of moments-based OSNR monitoring in QAM coherent optical communication," *IEEE Communications Letters*, vol. 24, no. 4, pp. 821–824, 2020.

[4] D. Shukla, A. Prakash, and R. Tripathi, "Adaptive modulation and coding for performance enhancement of vehicular communication," *Wireless Personal Communications*, vol. 26, no. 7, pp. 126–131, 2021.

[5] H. B. Kim, J. Morris, K. Miyashiro et al., "Astrocytes promote ethanol induced enhancement of intracellular Ca 2+ signals through intercellular communication with neurons," *iScience*, vol. 24, no. 5, pp. 102436–1024310, 2021.

[6] M. M. Hassan, K. Ahmed, B. K. Paul, M. N. Hossain, and F. A. Al Zahrani, "Anomalous birefringence and nonlinearity enhancement of As2S3 and As2S5 filled D-shape fiber for optical communication," *Physica Scripta*, vol. 96, no. 11, pp. 115501–115527, 2021.

[7] Z. Nan, S. Maozhong, and L. Haokai, "Satellite navigation signal enhancement method in tunnel based on virtual satellite," *Telecommunication technology*, vol. 60, no. 5, pp. 511–516, 2020.

[8] C. Pengyu, Y. Chengzhi, S. Limeng, and W. Hongchao, "LPI radar signal enhancement based on dae-gan network," *Systems engineering and electronic technology*, vol. 43, no. 9, pp. 2493–2500, 2021.

[9] D. Zou and S. Ma, "Satellite navigation and communication integration based on correlation domain indefinite pulse position modulation signal," *Wireless Communications and Mobile Computing*, vol. 2021, no. 9, p. 11, 2021.

[10] J. M. Moualeu and T. Ngatched, "Physical-layer security enhancement via relay-aided D2D communications underlaying cellular networks," *IEEE Open Journal of the Communications Society*, vol. 15, no. 9, pp. 413–427, 2020.

[11] O. I. Younus, N. B. Hassan, Z. Ghassemlooy et al., "Data rate enhancement in optical camera communications using an artificial neural network equaliser," *IEEE Access*, vol. 8, no. 36, pp. 42656–42665, 2020.

[12] X. Du and D. Wu, "Visual inspection system for trackside communication and signal infrastructure," *Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit*, vol. 235, no. 1, pp. 409–416, 2020.

[13] K. Ramadan, M. I. Dessouky, and F. El-Samie, "Equalization and co-carrier frequency offsets compensations for UWA-OFDM communication systems," *Wireless Personal Communications*, vol. 124, no. 3, pp. 2229–2245, 2022.

[14] K. Ramadan, M. I. Dessouky, and F. El-Samie, "A modified OFDM configuration with equalization and CFO compensation for performance enhancement of OFDM communication systems," *AEU-International Journal of Electronics and Communications*, vol. 126, no. 12, article 153247, 2020.

[15] M. M. Eid, V. Sorathiya, S. Lavadiya, E. Shehata, and A. N. Rashed, "Free space and wired optics communication systems performance improvement for short- range applications with the signal power optimization," *Journal of Optical Communications*, vol. 16, no. 15, pp. 139–146, 2019.

[16] Z. Qichao, M. Xin, and M. Shexiang, "Combined estimation of AIS mixed signal separation and detection based on PSP. Computer," *Simulation*, vol. 38, no. 1, pp. 464–470, 2021.

[17] M. Towliat, M. Rajabzadeh, and S. Tabatabaee, "On the noise enhancement of GFDM," *IEEE Wireless Communication Letters*, vol. 9, no. 8, pp. 1160–1163, 2020.

[18] R. Kamimura, "SOM-based information maximization to improve and interpret multi-layered neural networks: from information reduction to information augmentation approach to create new information," *Expert Systems with Application*, vol. 125, no. 5, pp. 397–411, 2019.

[19] L. Bo, J. Lv, X. Luo, H. Wang, and S. Wang, "A novel and fast nonlocal means denoising algorithm using a structure tensor," *Journal of Supercomputing*, vol. 75, no. 2, pp. 770–782, 2019.

[20] B. Amutha, A. C. J. Malar, K. Nanmaran, D. M. Hussain, V. Jeyakrishnan, and M. Karthikeyan, "Enhanced development of communication between the network and the end user by eliminating the interference signals in MIMO channel," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 12, pp. 77–83, 2020.

[21] C. Liu, J. He, C. Zhang, M. Cao, X. Song, and X. Zhu, "Pattern recognition method for time-domain waveform images of GIS partial discharge," *Proceedings of the CSU-EPSA*, vol. 43, no. 2, pp. 171–178, 2019.

[22] L. Cui, J. Yang, L. Wang, and H. Liu, "Theory and application of weak signal detection based on stochastic resonance mechanism," *Security and Communication Networks*, vol. 2021, no. 2, p. 9, 2021.

[23] F. M. Yu, K. C. Lee, K. W. Jwo, R. S. Chang, and J. Y. Lin, "Low distortion of noise filter realization with 6.34 V/μs fast slew rate and 120 mVp-p output noise signal," *Sensors*, vol. 21, no. 3, pp. 1008–1011, 2021.

[24] D. Pantke, F. Mueller, S. Reinartz et al., "Frequency-selective signal enhancement by a passive dual coil resonator for magnetic particle imaging," *Physics in Medicine & Biology*, vol. 67, no. 11, pp. 115004–115113, 2022.

[25] X. Zhou, Z. Sun, and H. Wu, "Wireless signal enhancement based on generative adversarial networks," *Ad Hoc Networks*, vol. 103, no. 8, p. 102151, 2020.

[26] C. C. Chuang, C. C. Lee, C. H. Yeng, E. C. So, B. S. Lin, and Y. J. Chen, "Convolutional denoising autoencoder based SSVEP signal enhancement to SSVEP-based BCIs," *Microsystem Technologies*, vol. 32, no. 10, pp. 321–325, 2019.

[27] L. I. Jun-Xia and S. F. Yuan, "Signal enhancement algorithm of OFDM network based on frequency domain narrowband noise cancellation," *Journal of Southwest China Normal University (Natural Science Edition)*, vol. 52, no. 3, pp. 83–89, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 6942940, 14 pages
https://doi.org/10.1155/2022/6942940

Research Article

A Novel SVM Network Using HOG Feature for Prohibition Traffic Sign Recognition

Yang Liu[1] **and Wei Zhong**[2]

[1]*School of Information Science and Engineeriing, Chongqing Jiaotong University, Chongqing 40074, China*
[2]*Department of Logistics Command, Army Logistics University of PLA, Chongqing 40041, China*

Correspondence should be addressed to Wei Zhong; 44526081@qq.com

Received 25 April 2022; Accepted 27 June 2022; Published 20 July 2022

Academic Editor: Maode Ma

To recognize prohibition traffic sign, this paper proposes a novel method that is trained by a small number of samples and uses the feature of histogram of oriented gradient (HOG) and support vector machine (SVM) network. The recognition method is mainly divided into three stages. The first stage is image preprocessing, which includes image interception based on ellipse detection, image resizing, and Gamma correction. In the part of image interception, a new ellipse detection method called RHT_MCN is proposed based on RHT, which uses the maximum coincidence number (MCN) of image edge points and detected ellipse edge to choose the final ellipse for image interception. The second stage is the feature extraction of HOG. The third stage is the prohibition traffic sign recognition (PTSR) based on SVM network. In the design and implementation of the PTSR model, a new single-layer SVM network is proposed. The ascending spiral training method of the recognition model is introduced in detail. Finally, the data from GTSRB is used to test and analyze the prohibition traffic sign recognition method. The method is proven to have good applicability.

1. Introduction

With the development of intelligent vehicles and Internet of Vehicles, traffic sign recognition (TSR) is becoming more and more practical and popular. TSR plays an important role in advanced driver assistance systems (ADAS), intelligent vehicles, and intelligent transportation and plays an important role in vehicle traffic safety and pedestrian safety. More specially, some reported TSR applications are as follows [1]: driver assistance systems, autonomous vehicles, maintenance of traffic signs, engineering measurements, Vehicle-to-X (V2X) communication, reducing fuel consumption, and so on.

In the implementation of TSR, the TSR approach is accompanied by the development of pattern recognition methods and classification methods. In the research [2], Ruta et al. summarized that the more popular TSR methods were feature-based approaches and the pixel-based cross-correlation template matching was a baseline approach. With the development of machine learning algorithms such as deep learning (DL) [3] and support vector machine (SVM) [4], the intelligent learning algorithm is applied to establish the model of traffic sign recognition. In the research [5], Wang et al. proposed that the TSR methods can be divided into two categories: traditional (non-DL) machine learning methods and deep learning methods. From the time perspective, Badue et al. [6] summarized that most of the earlier approaches for traffic sign detection and recognition were model-based which used simple features, and learning-based approaches started leveraging simple features but evolved into using more complex ones. In general, the TSR method is feature-based. The difference in feature extraction is the difference between hand-crafted features and self-extraction features. And the difference in classifiers is the difference between the rule-based classification model and the machine learning classification model.

Here, we briefly summarize the main application scenarios and implementation methods of traffic sign recognition. The rest of this paper is composed of six parts. In Section 2, the typical methods of TSR are introduced. In Section 3, the composition characteristics and color statistical characteristics of prohibition traffic sign (PTS) are discussed. In Section 4, a novel method of prohibition traffic sign recognition is proposed, which is based on histogram of oriented gradient (HOG) feature and support vector machine (SVM) network. In Section 5, the training method of the recognition model is introduced in detail. In Section 6, the self-built data set and GTSRB are used for the verification and analysis. In Section 7, conclusions and the future work of this work are drawn.

2. Related Work

Deep learning model has made a great success in ImageNet contest which is a challenging image classification task with 1000 classes and 1.2 million high-resolution images [7]. TSR is an image classification problem. More and more deep learning models were successfully utilized to recognize the traffic sign. Convolutional neural network (CNN) and its variant are used for the deep learning model for TSR since CNN has witnessed great success in the task of image classification. Especially, the multicolumn deep neural network (MCDNN) was used to win the championship of the 2012 German traffic sign recognition competition, and its recognition rate exceeded the human recognition rate [8]. In the work of Zhang et al. [9], they used a shallow network architecture based on convolutional neural networks (CNNs) for TSR and reached a high accuracy (99.84%) based on the full GTSRB [10, 11] data set. The number of samples is large for deep learning models. Here, we list the number of signs of some typical deep learning training data sets for TSR.

Table 1: The list table of the number of signs.

Data set	Number of classes	Number of signs
GTSRB	43	50000+
TT100K [12]	45	30000
STS [13]	7	20000
BTSC [14]	62	7125
ETSD [15]	164	82476

As shown in Table 1, the number of each TSR data set is at least 7125, and it takes time, manpower, and equipment to collect sample signs. However, there are not so many samples for the model training in some applications and the traffic sign recognition model should be trained by using a small number of samples. In this case, the deep learning model is in the dilemma of insufficient samples. It is necessary for finding a new model which only need a small number of samples.

Either the machine learning (non-DL) model or the deep learning model, the thing which is used for the recognition is the features. The features can be extracted from the artificially designed network model, but the feature may not be visualized well and the features are some black boxes. Also, the features can be manually extracted by using some algorithm and the features are white boxes for researchers. In addition, the selected features can be taken to account for the global and local characteristics. For the above reasons, the HOG feature is selected. HOG feature was first proposed by Dalal and Triggs and combined with SVM to realize pedestrian detection [16]. Based on the extraction of HOG feature, the single hidden layer feed-forward network trained by extreme learning machine (ELM) is used to realize the efficient recognition of traffic signs [17]. Based on a multitask convolution neural network, using an amount of data for training, Luo et al. realized the traffic sign recognition [18]. Based on HOG features extended to its color space combined with local self-similar descriptors, the random forest method is used to classify and recognize traffic signs [19].

SVM is a classical algorithm of machine learning, which has a good performance in binary classification and multiclassification. Some classification models based on HOG feature and SVM were used for traffic sign recognition. In the work of Yao et al. [20], a traffic sign recognition method using histogram of oriented gradient support vector machine and grid search was proposed, and the grid search technique was applied to optimize the parameters of the support vector machine, and traffic signs were extracted from the different condition images by using HSI color space and normalization. In the work of Junges et al. [21], the red color segmentation and the Hough transform were used to find circular regions for traffic sign detection, and SVM and the HOG feature were used for TSR. In the research of Tun and Lwin [22], Real-time Myanmar Traffic Sign Recognition System (RMTSRS) was proposed, and each incoming frame was segmented using the color threshold method for traffic sign detection, and the HOG feature was extracted and RMTSRS classified traffic sign types using SVM. In the work of Tang et al. [23], the traffic sign was located with Hough transformation based on the spatial characteristics of the image, and the SVM classifier was used to get the training model with HOG features of traffic signs. Tang et al. also pointed out that the first thing to recognize traffic signs is to segment the image, to reduce the interference of the image outside the sign area. In the work of Cotovanu et al. [24], the traffic sign detection which was based on color information and certain object properties used the image processing techniques to identify regions of interest (ROIs) in an image, and a linear SVM binary classifier trained with HOG features was used for TSR.

From the above work, it is widely considered to be a good way to divide TSR into three steps:

(i) Segment the traffic sign image

(ii) Extract the features of the traffic sign

(iii) Classify the traffic sign by trained model

However, it must also be mentioned that their work remains some uncertainties and problems:

(1) The common method of the segmentation of traffic sign image is usually based on color space. But the color of traffic signs can be easily disturbed by environmental factors and the color is usually not the standard color

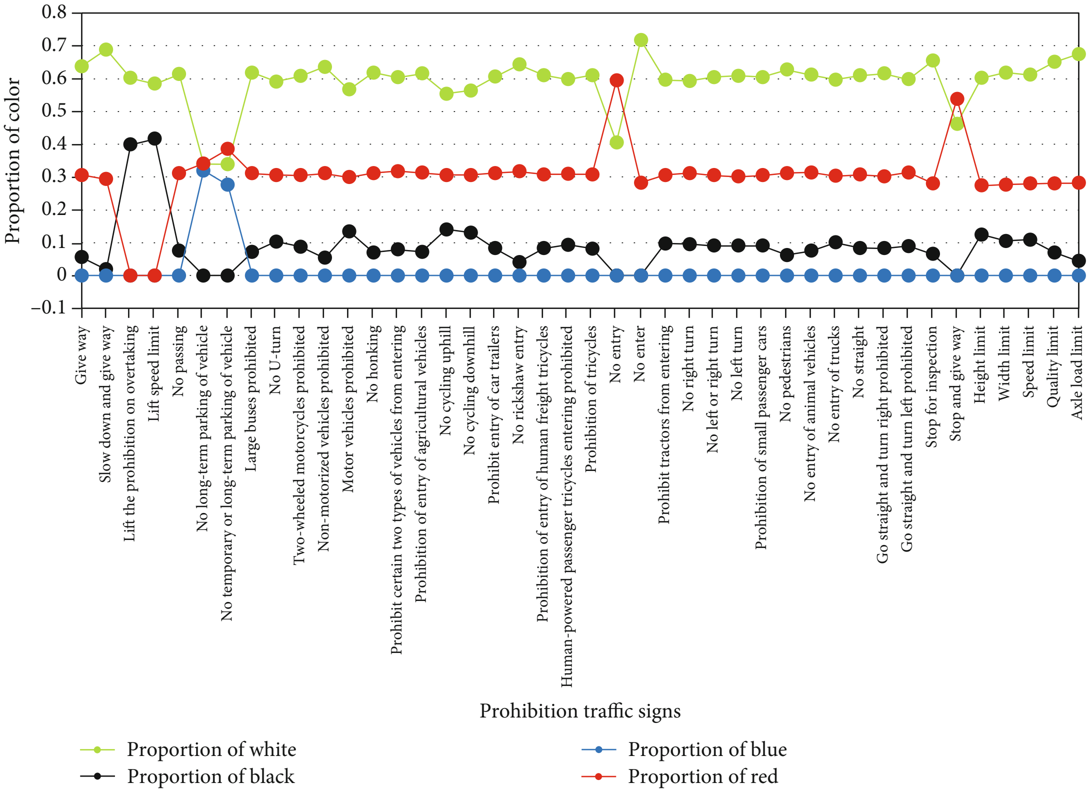


Figure 1: Color proportion of standard prohibition signs.

(2) Although lots of studies have been conducted to utilize SVM for TSR, little attention has been done to utilizing SVM to construct SVM network. And one of the challenges in SVM model is the optimization of the parameters of SVM model

In this paper, based on the extraction of HOG feature, a SVM network is used to realize the recognition of prohibition traffic signs and the method of optimizing the parameters of SVM model is introduced.

3. Analysis of Prohibition Traffic Sign

In this section, the necessity of prohibition traffic sign recognition (PTSR) will be expounded. The color composition of prohibition traffic sign will be analyzed, and the reason of select ellipse detection for traffic sign segmentation will be introduced too.

Traffic signs are mainly divided into seven categories: warning traffic sign, prohibition traffic sign, indication traffic sign, guide traffic sign, tourist area sign, road construction sign, and auxiliary sign [25]. Prohibition traffic sign is one of the commonly used signs. According to the national standard of the People's Republic of China (gb5768.2-2009), there are 42 kinds of prohibition traffic signs. Some standard prohibition traffic signs are shown in Figure S1.

From the composition and frequency of use, it is necessary to research PTSR. From the analysis of the composition color of the prohibition traffic sign image, the main composition colors are red, white, black, and blue. Based on color standardization [26, 27] and image capture of the sign image, the number of pixels of four colors involved in the prohibition traffic sign image is counted, and their proportion in the total pixels of the image is calculated, respectively; then, each sign can obtain four characteristic values of color proportion. The statistical results are shown in Figure 1.

As shown in Figure 1, there are obvious color differences in some of the prohibition signs. For example, the color composition of the long-term parking prohibition sign and the temporary or long-term parking prohibition sign is red and blue, regardless of the white background of the sign image. Considering the proportions of white, black, blue, and red of all prohibition signs, only the rough classification of prohibition signs can be realized based on the color proportion information, and the better classification of prohibition signs should be realized based on the texture, shape, and other local or global features of prohibition signs. HOG feature describes the local detail features through the directional gradient data and describes the global features of the image through the histogram statistical data of the directional gradient. Therefore, the HOG feature is used to describe the prohibition traffic signs and will be used for the recognition model which is based on HOG feature and SVM.

4. Prohibition Traffic Sign Recognition Method Based on HOG-SVM

In this section, the overall framework of the prohibition traffic sign recognition method will be proposed and introduced stage by stage.

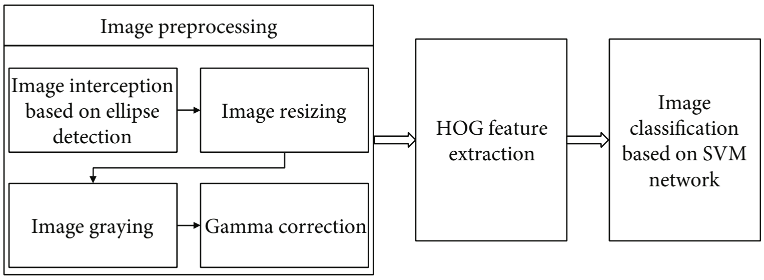


Image preprocessing
Image interception based on ellipse detection
Image resizing
Image graying
Gamma correction
HOG feature extraction
Image classification based on SVM network

Figure 2: Flow chart of prohibition traffic sign recognition method.

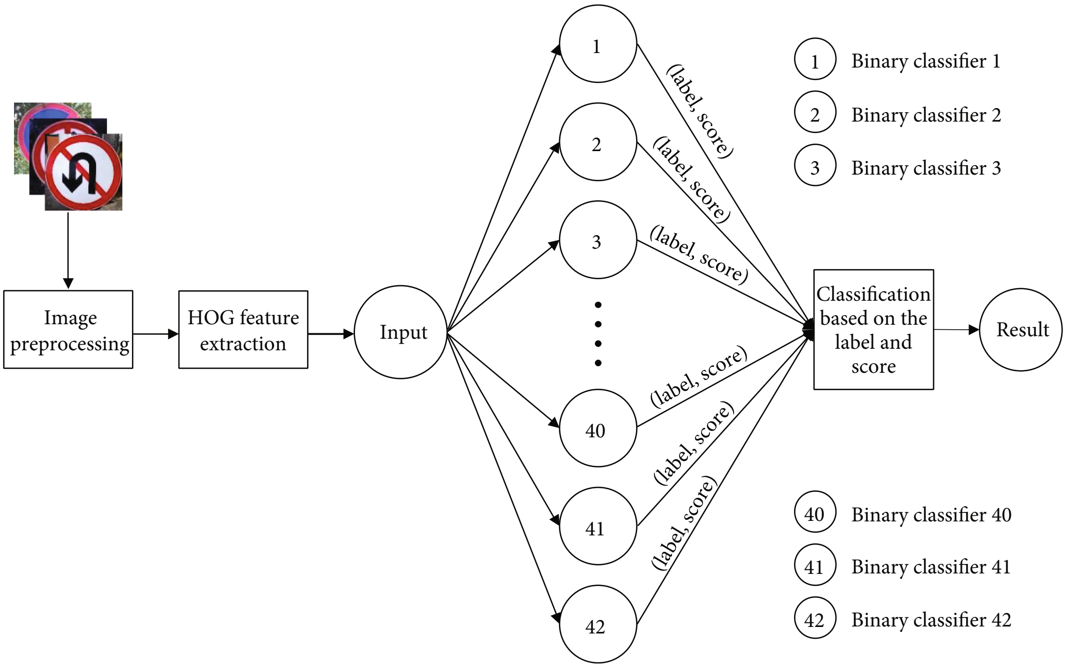


Figure 3: Schematic diagram of prohibition traffic sign recognition network based on SVM which is 42 in this paper.

As shown in Figure 2, the method of prohibition traffic sign recognition is mainly divided into the following three steps:

(i) Image preprocessing

(ii) HOG feature extraction

(iii) Image classification based on SVM model

Especially, the image classification based on SVM model is constructed by SVM network. As shown in Figure 3, a novel single-layer SVM network is proposed in detail. Each node of the single-layer SVM network is a trained SVM. Each SVM node is a binary classifier. The number of the SVM node is depending on the number of types of prohibition traffic signs.

4.1. Image Preprocessing. In this part, image preprocessing includes image interception based on ellipse detection, image resizing, image graying, and Gamma correction.

4.1.1. Image Interception Based on Ellipse Detection. In the actual scene, there are other objects in the background. To get the image of the traffic sign, the interception of the prohibition traffic sign is needed. As is shown in Figure S1, the shape of the prohibition traffic sign is a circle. There are 40 kinds of prohibition traffic signs which is circular. Due to the influence of various factors, most of the shapes of prohibition traffic sign images are an ellipse in the actual scene. And it is necessary to get a method of the detect ellipse in the image.

Many researchers have worked on the detection of the ellipse. Hough transformation was proposed by Hough in 1962 [28]. But the traditional ellipse detection method based on Hough transformation has high computational complexity, which is not conducive to the fast interception of traffic sign image. To improve the computational performance, randomized Hough transformation (RHT) was proposed in 1996 [29]. RHT has a better performance in the simple image than the complex image. In this paper, an ellipse detection method is proposed based on RHT, which uses the maximum coincidence number (MCN) of image edge points and detected ellipse edge to choose the final ellipse. The method is called RHT_MCN.

The steps of RHT_MCN are as follows:

(1) Get the grayscale image of the prohibition traffic sign and get the binary image of the prohibition traffic sign image

(2) Use the convolution to realize edge extraction of the binary image; the convolution kernel is as follows:

$$K = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{1}$$

(3) Get the alternative ellipse by using the method of RHT

(4) Calculate the coincidence number of the image edge and each alternative ellipse, and choose the ellipse of maximum coincidence number for the final detected ellipse

The general equation of the ellipse is as follows:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0. \tag{2}$$

A, B, C, D, E, and F are the coefficients of elliptic general equations and x and y are the coordinates of points on the ellipse.

The pixel of the image is discrete. And there is a set S_P and an error d_error.

$$\text{s.t.} Ax^2 + Bxy + Cy^2 + Dx + Ey + F \leq d\_\text{error}, \quad \forall (x, y) \in S_P. \tag{3}$$

The x and y are the coordinates of image pixels. And the image edge points included in the set S_P are counted, and the coincidence number of the image edge and each alternative ellipse is obtained. Based on ellipse detection, the image is intercepted, to realize the ellipse interception of the prohibition traffic sign image.

As is shown in the figure, Figures 4(a), 4(c), and 4(e) are the original image before the interception. Figures 4(b), 4(d), and 4(f) are the resulting image after interception based on ellipse detection using RHT_MCN. In Figures 4(b), 4(d), and 4(f), the green ellipse line is the detected ellipse.

As shown in the figure, Figure 5(a) is the ideal result of common ellipse detection and there are 4 ellipses detected finally. Typically, more than four ellipses are detected for the discontinuity of image edge point coordinates in the traffic signs. If there are multiple ellipses detected finally, it will be necessary to find a suitable ellipse for the image segmentation. Figure 5(b) is the result of ellipse detection using RHT_MCN, and there is only one ellipse detected which is more suitable for the segmentation of traffic sign image.

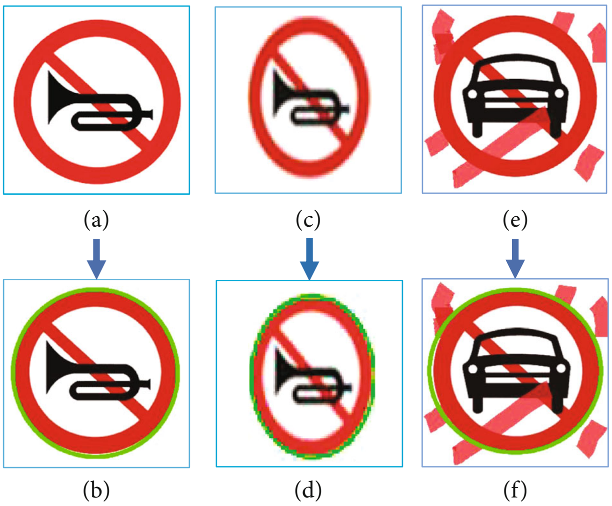


Figure 4: Interception of prohibition traffic sign based on ellipse detection.

4.1.2. Image Resizing and Gamma Correction. As a result of the interception, the shape of the prohibition traffic sign is an ellipse in most cases. The shape change of the image will affect the feature extraction and recognition of the prohibition traffic sign. The image size should be resized. The aspect ratio of an ellipse circumscribed rectangle is not 1 : 1, while that of a circle circumscribed rectangle is 1 : 1. The method of image size adjustment is as follows:

Taking the circumscribed rectangle of the intercepted image as the benchmark, the aspect ratio of the circumscribed rectangle is adjusted to 1 : 1 to realize the transformation from rectangle to square. Considering the subsequent HOG feature extraction, the size of the image also needs to be enlarged or reduced. The final image size of this paper is 24 × 24. The schematic diagram of image adjustment is shown in Figure 6.

Gamma correction is mainly used to process the brightness of the image and weaken the influence of light and shadow on the image.

4.2. HOG Feature Extraction. As shown in Figure 7, the extraction of HOG feature can be generally divided into 5 steps. And it is important to set the following parameters: spatial/orientation bins, cell, block, and sliding step size. These parameters are related to the comprehensiveness of hog features to global and local features and the dimension of hog features. While extracting HOG features, the corresponding spatial/orientation bins, cell, block, and sliding step size are shown in Table 2. The image size and the corresponding HOG feature dimension are shown in Table 3. The image size used in this paper is 24 × 24. Therefore, the dimension of HOG feature is 144.

4.3. Image Classification Based on SVM Model. In this part, firstly, based on the training data set, support vector machine is used to generate the initial classifier of prohibition traffic signs. The initial classifier is constructed by 42

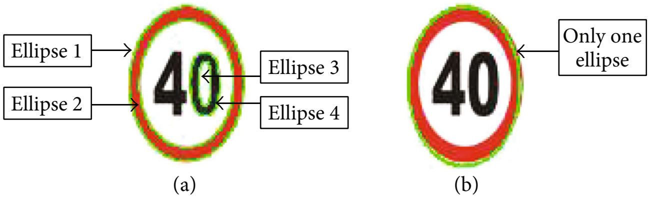


Figure 5: Comparison of different ellipse detection methods.

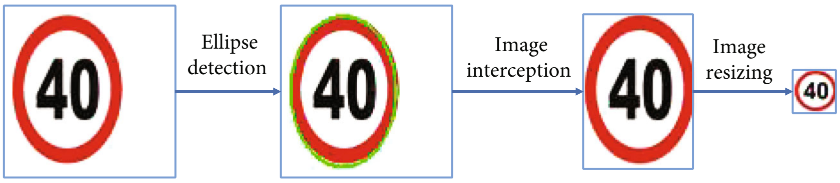


Figure 6: The schematic diagram of image adjustment.

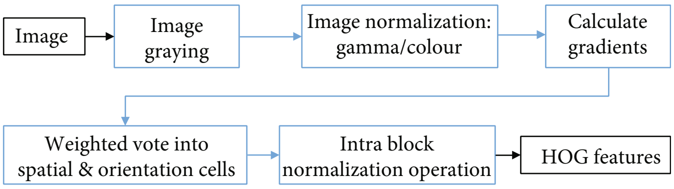


Image
Image graying
Image normalization: gamma/colour
Calculate gradients
Weighted vote into spatial & orientation cells
Intra block normalization operation
HOG features

Figure 7: The chain of HOG feature extraction.

Table 2: Parameter table of HOG feature extraction.

Spatial/orientation bins	0°~360°/9 bins
Cell	8 × 8
Block	2 × 2
Sliding step	1

Table 3: Image size and dimensions of HOG features.

Image size	Dimensions of HOG features
16 × 16	36
24 × 24	144
32 × 32	324

binary classifiers. Then, the initial classification result is used to get the final classification result.

As shown in Figure 8, the prohibition traffic sign recognition network based on SVM binary classification is mainly divided into the following two steps:

(1) Initial classification based on SVM binary classifier

In this part, the HOG feature data extracted after image preprocessing is input into 42 trained binary classifiers one by one for prediction, and the result [*label, score*] is obtained. The *label* is the classification and prediction result of the input data image by the binary classifiers, which is used to indicate which category the image belongs to, and the *score* is the credit score of each category of the image classification results. In this paper, the label values are "1" and "2," where "1" indicates that the image belongs to the corresponding prohibition traffic sign image class of the classifier and "2" indicates that the image does not belong to the corresponding prohibition traffic sign image class of the classifier.

(2) Classification based on the predicted result

In this part, based on the prediction result [*label, score*] in step (1), we can get the prediction result label data set LABEL and the prediction result score data set SCORE. The following are the steps:

(i) Analyze the data set LABEL. If the prediction result of one binary classifier N_1 is $label = 1$, the prohibition sign image is considered as the corresponding traffic sign of the binary classifiers N_1; otherwise, enter the next step and analyze the data set SCORE

(ii) Analyze the data set SCORE. If the result score of the binary classifier N_2 is the maximum value of the data set SCORE, the prohibition traffic sign image is considered as the corresponding traffic sign of the binary classifier N_2

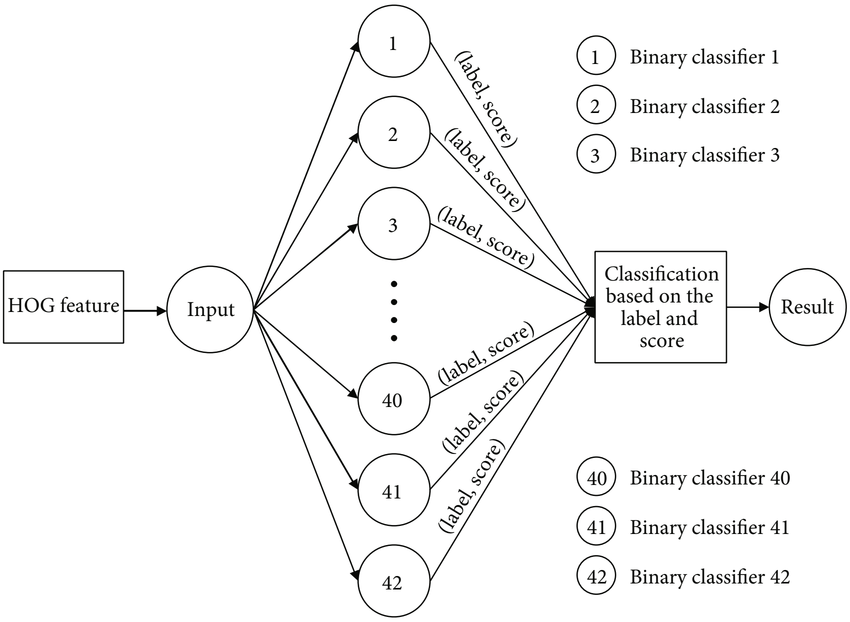


Figure 8: Schematic diagram of prohibition traffic sign recognition network based on SVM binary classification.

5. Generating SVM Classifier Based on HOG Feature

In this section, the method of training every single binary classifier will be introduced. Then, the training method of the model will be presented too.

5.1. Training of Single Binary Classifier. A single binary classifier is a classifier that uses support vector machine to distinguish one kind of prohibition traffic sign from other prohibition traffic signs. It can be seen from Figure 1 that the prohibition traffic signs contain at least 42 types of signs, so at least 42 binary classifiers are generated. Let X_i be the HOG feature vector set corresponding to the i th prohibition traffic sign and F_i be the binary classifier of X_i $(i = 1, 2, 3, \cdots, 42)$.

$$Y_i = F_i(x) = \begin{cases} 1, \forall x \in X_i, \\ 0, \forall x \in \overline{X_i}, \\ i = 1, 2, 3, \cdots, 42. \end{cases} \tag{4}$$

42 classifiers need to be trained one by one. It should be noted that if $\forall x$ can be accurately classified according to formula (4), then any prohibition traffic sign can be recognized only by judging all binary classifiers one by one. However, if

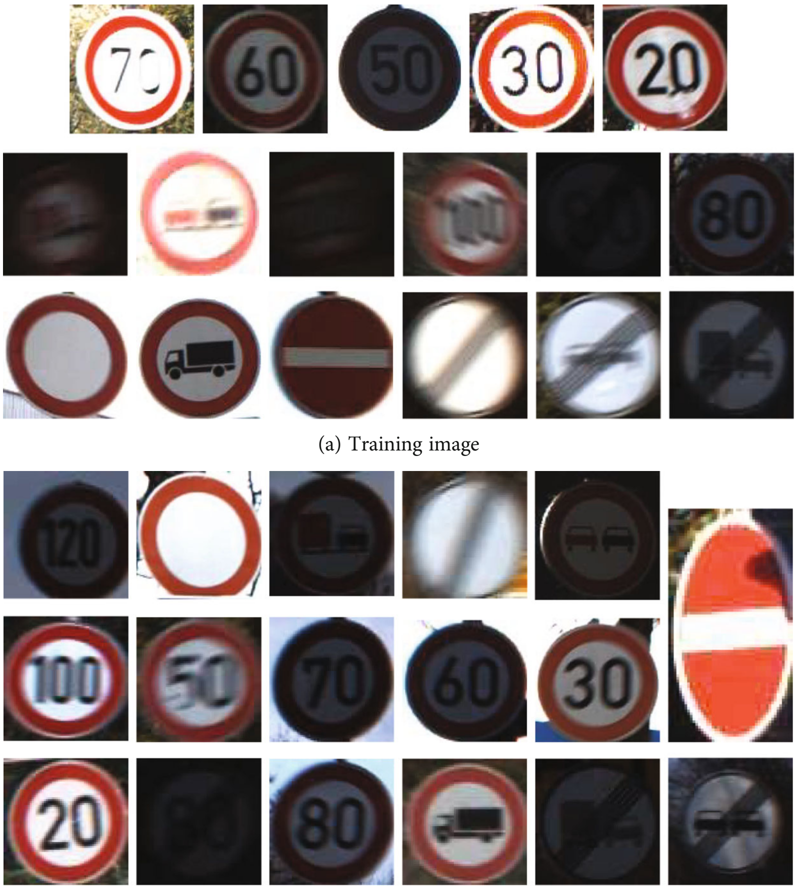
$$\begin{gathered} \text{s.t.} F_i\left(x'\right) = 1, \exists x' \in \bar{X}_i, \\ \text{or s.t.} F_i\left(x''\right) = 0, \exists x'' \in X_i. \end{gathered} \tag{5}$$

Table 4: Experimental environment configuration.

Experimental environment	Configuration
Operating system	Windows 10 Home Edition (64)
Software platform	Matlab 2017
CPU	11th Gen Intel(R) Core(TM) i7-1165G7 2.80 GHz
RAM	16 GB

Table 5: The data of test result.

Class ID	Traffic sign (TS)	Number of TS	Correct rate
TS1	No long parking	40	100.00%
TS2	No right turn	40	100.00%
TS3	Height limit	54	98.15%
TS4	No left turn	38	94.74%
TS5	No entry	36	94.44%
TS6	No pedestrian access	32	93.75%
TS7	No U-turn	36	91.67%
TS8	No motor vehicles	40	90.00%
TS9	No temporary or long-term parking	44	88.64%
TS10	No honking	33	87.88%
TS11	Speed limit 60	50	86.00%
TS12	No overtaking	58	82.76%
TS13	No entry of trucks	33	81.82%
TS14	No entry of nonmotor vehicles	38	81.58%
TS15	Speed limit 40	38	81.58%
Total		610	90.20%



(a) Training image

(b) Test image

Figure 9: Sample of training image and test image.

When the classifier is trained, there is

$$\begin{aligned} &\text{s.t.} F_i(x) = 1, \forall x \in X_i, \\ &\text{s.t.} F_i\left(x'\right) = 1, x' \in \bar{X}_i. \end{aligned} \tag{6}$$

The purpose is to avoid any unrecognizable traffic signs.

5.2. Ascending Spiral Training Method of Traffic Sign Recognition Model. The PTSR model is constructed by multiple SVM binary classifiers and a classifier based on the predicted result. The key part is the training of multiple SVM binary classifiers. The training method of each SVM binary classifier has been introduced above. The final model should not be obtained at one time. And the final model can be obtained through iterative training.

The number of the SVM binary classifiers is the number of prohibition traffic sign types. If there are N types of traffic signs needed to be recognized, the number of SVM binary classifiers should be N. There is a constraint parameter P for each SVM binary classifier. The default value of P is 1. The value of P is relative to the margin of SVM. The greater the value of P, the smaller margin, and the fewer points lie within the margin. The smaller the value of P, the wider margin, and the more points lie within the margin. On the basis of having selected training samples and kernel function, the setting of parameter P can affect the classification effect. In this paper, the initial kernel function is radial basis function (RBF). The process of setting the constraint parameter P is necessary and is as follows:

(1) The training data set is divided into two parts. One part is the sample data of SVM, and the other part is used for testing the traffic sign recognition model which is called training data used for testing (TDUT)

(2) Set the value of parameter P by default initially, and there is

$$P_i = 1 \; (i = 1, 2, 3, \cdots, N) \tag{7}$$

(3) Train the SVM binary classifiers using the selected sample data and obtain the N SVM binary classifiers

(4) Use the SVM binary classifiers of step 3 to construct the traffic sign recognition model. Then, use the TDUT to test the model, and get the total recognition rate R_a and the recognition rate of each type of traffic sign $R_i (i = 1, 2, 3, \cdots, N)$

Table 6: The test result based on the data of GTSRB.

Class ID	Number of samples	Number of samples recognized	Correct rate	Final value of constraint parameter	Index of constraint parameter
00000	13	11	84.62%	0.9	1
00001	135	113	83.70%	4.1	2
00002	118	97	82.20%	3.0	3
00003	75	68	90.67%	1.0	4
00004	120	110	91.67%	8.6	5
00005	100	84	84.00%	8.3	6
00006	28	26	92.86%	0.6	7
00007	90	84	93.33%	4.3	8
00008	66	54	81.82%	5.3	9
00009	109	105	96.33%	0.8	10
00010	156	147	94.23%	4.3	11
00015	45	44	97.78%	0.2	12
00016	41	41	100.00%	0.2	13
00017	93	93	100.00%	0.2	14
00032	15	14	93.33%	2.1	15
00041	11	10	90.91%	1.3	16
00042	24	23	95.83%	0.5	17
Total	1239	1124	90.72%		

(5) Find out the minimum value of R_i from the noniterated SVM binary classifier in this round of iterative processes. If the recognition rate of the jth type of traffic sign is equal to the minimum value, then let the P_j varies from 0.1 to 10 in steps of 0.1 and the other parameter P is invariant at this time

(6) Find out the best total recognition rate R_{a_temp} ($R_{a\_\text{temp}} \geq R_a$), and set the P_j to the first number which varies from 0.1 to 10 in steps of 0.1 and makes the total recognition rate best

(7) If the parameter P_j is the last reset in this iteration and the total recognition rate has never been increased in this iteration, it should be the end of traffic sign recognition model training. Otherwise, go to step (3).

6. Verification and Analysis

In this section, the proposed prohibition traffic sign recognition method will be tested by using self-built data set. And the method will be verified based on GTSRB for comparative analysis in the case of using a different SVM kernel function.

The experimental environment configuration is shown in Table 4.

6.1. Source of Verification Data. Constructing a good benign data set is crucial to our model's performance. To collect our benign data set, first, we downloaded traffic sample pictures from China traffic sign detection data set (CCTSDB) [30] and Tsinghua Tencent 100K (TT100K) data set to ensure a diversity of types of benign files. Then, we obtain the prohibition traffic sign that we need from those downloaded sample pictures by image matting. To ensure that the label of the prohibition traffic sign is sufficiently detailed, we relabelled the obtained prohibition traffic signs and classified them into 15 categories which were classified by the national standard of the People's Republic of China (gb5768.2-2009). However, even after that, we found the data set still seemed not complete (e.g., missing import prohibition traffic sign). We further photoed those missing samples manually and labelled them as an important supplement to our data set. To improve the quality of our data set, we only accepted benign samples. In total, after filtering, we obtained 610 unique test samples. An example of test data is shown in Figure S2 in the supplemental files.

6.2. Test Results. As shown in Table 5, the test results of 15 kinds of prohibition traffic signs show that the correct recognition rate of 8 kinds of prohibition traffic signs (traffic sign of no long parking, traffic sign of no right turn, traffic sign of height limit, traffic sign of no left turn, traffic sign of no entry, traffic sign of no pedestrian access, traffic sign of no U-turn, and traffic sign of no motor vehicles) is greater than or equal to 90%. The correct recognition rate of the other 7 kinds of prohibition traffic signs is within the range of (0.81, 0.9), and the total correct recognition rate of 15 kinds of prohibition traffic signs is 90.2%. Overall, the proposed classification model achieves the classification and recognition of traffic signs.

6.3. Result Analysis. By comparing the spatial detail complexity O of the traffic sign TS1~TS8 and the traffic sign TS9~TS15, there is

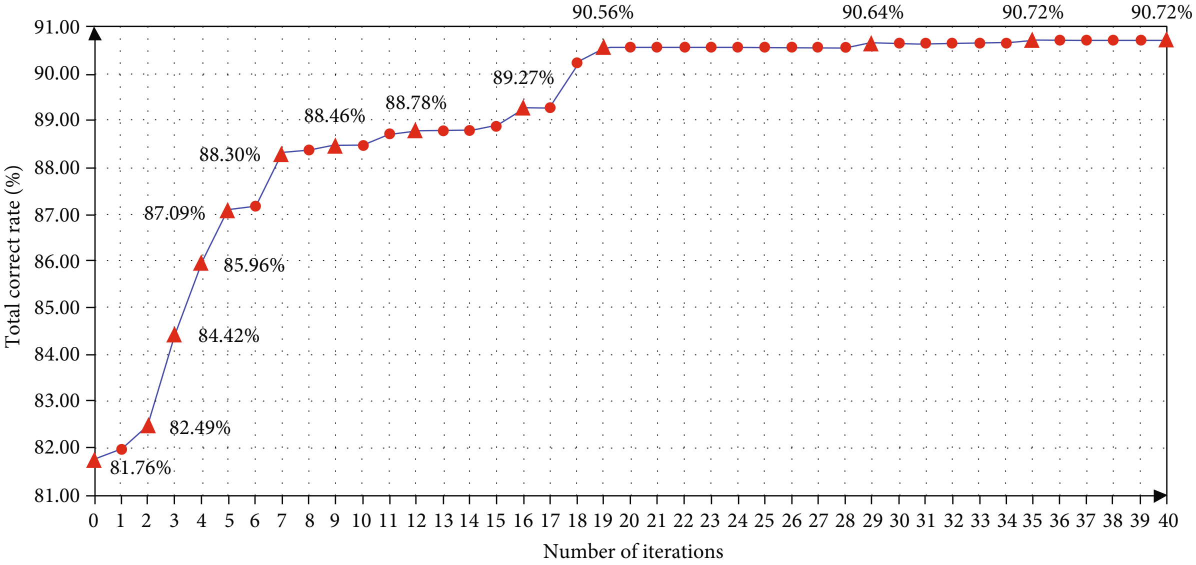


Figure 10: Schematic diagram of the change of total correct rate (TCR) with the number of iterations.

Table 7: The correspondence table between constraint parameters and TCR during iteration.

Number of iterations	Constraint parameter	TCR	Number of iterations	Constraint parameter	TCR
1	P_6	82.00%	21	P_2	90.56%
2	P_9	82.49%	22	P_1	90.56%
3	P_3	84.42%	23	P_4	90.56%
4	P_5	85.96%	24	P_{16}	90.56%
5	P_8	87.09%	25	P_5	90.56%
6	P_{16}	87.17%	26	P_7	90.56%
7	P_2	88.30%	27	P_8	90.56%
8	P_1	88.38%	28	P_{15}	90.56%
9	P_{15}	88.46%	29	P_{11}	90.64%
10	P_4	88.46%	30	P_{17}	90.64%
11	P_{11}	88.70%	31	P_{10}	90.64%
12	P_7	88.78%	32	P_{12}	90.64%
13	P_{10}	88.78%	33	P_{13}	90.64%
14	P_{17}	88.78%	34	P_{14}	90.64%
15	P_{12}	88.86%	35	P_9	90.72%
16	P_{13}	89.27%	36	P_3	90.72%
17	P_{14}	89.27%	37	P_2	90.72%
18	P_9	90.23%	38	P_6	90.72%
19	P_6	90.56%	39	P_1	90.72%
20	P_3	90.56%	40	P_4	90.72%

$$O_{\mathrm{TS1\sim TS8}} < O_{\mathrm{TS9\sim TS15}}. \tag{8}$$

When the image size is reduced, the spatial detail information of the image is lost; thus, the HOG feature data of the image is lost too. Besides if the adverse factors such as image rotation and illumination are superimposed, the change of HOG feature of image exceeds the tolerance limit of SVM binary classifier, which leads to a decrease in classification accuracy. The correct rate of recognizing the prohibition traffic sign TS1~TS8 is greater than the correct rate of recognizing the prohibition traffic sign TS9~TS15.

In addition, insufficient training sample is one of the reasons for the low classification accuracy. In this paper, the training samples consider the linear deformation of the image such as translation, horizontal and vertical unequal ratio deformation completely, and the nonlinear deformation caused by rotation and lens motion is not considered enough, which reduces the classification accuracy of the classifier for nonlinear deformation image.

Table 8: New correspondence table between constraint parameters and TCR during iteration.

Number of iterations	Constraint parameter	TCR
1	P_6	82.00%
2	P_9	82.49%
3	P_3	84.42%
4	P_5	85.96%
5	P_8	87.09%
6	P_{16}	87.17%
7	P_2	88.30%
8	P_1	88.38%
9	P_{15}	88.46%
11	P_{11}	88.70%
12	P_7	88.78%
15	P_{12}	88.86%
16	P_{13}	89.27%
18	P_9	90.23%
19	P_6	90.56%
29	P_{11}	90.64%
35	P_9	90.72%

6.4. Verify the HOG-SVM Model Based on the Data of GTSRB. To analyze the applicability of the SVM network, the model is trained by the data of GTSRB. Although the training data was chosen, the training data is mainly the prohibition traffic sign image. There are 17 types of prohibition traffic signs in the GTSRB. To train the HOG-SVM model, 595 sample pictures are used for the sample data of SVM, and the number of each prohibition traffic sign image used is 35. And all the pictures are chosen from the training data set randomly. There are 1239 pictures used to test the model finally. The sample training image and test image are shown in Figure 9.

6.4.1. Use the RBF Kernel Function for the Test. In this part, the RBF kernel function for SVM is used firstly for the test. And the test result is shown in Table 6.

As shown in Table 6, the highest correct rate is 100% and the lowest correct rate is 80.30%. The total correct rate (TCR) is 90.72%. The applicability of the recognition method is confirmed. The constraint parameter of class 0015~0017 is the lowest (0.2) and the correct rate is almost the highest. That is to say, the margin of the three SVM is wider than others and the SVM has the stronger generalization ability. The value of the constraint parameter is relative to the difference in the HOG feature of the traffic sign. The HOG feature of class 0015~0017 is more different from other classes of the prohibition traffic sign. Also, with the iterative training, the changing trend of the total recognition correct rate is shown in Figure 10.

From the first iteration to the 17th iteration, the total correct rate improved from 81.76% to 89.27%. Then, the second round of iteration begins. Finally, the total correct rate is 90.72%. Judging from the growth trend of the total correct rate, the first few iterations of each round of iteration make the total correct rate improve more greatly. This shows that the traffic sign classes of the lowest correct rate part provide more growth on the total correct rate and the model training method is confirmed too.

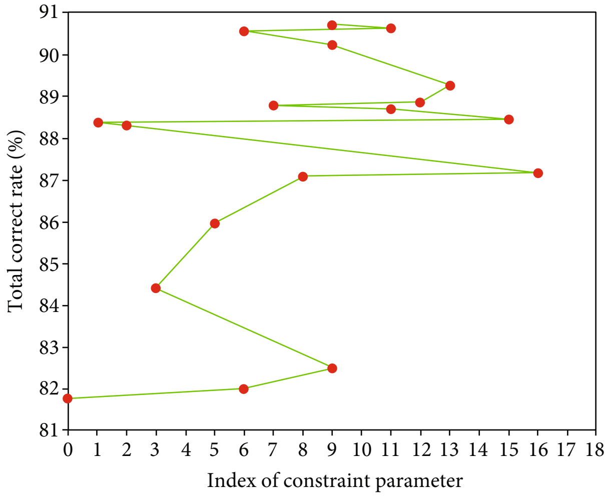


Figure 11: The 2D line chart of correspondence between constraint parameters and TCR during iteration.

Table 9: The result of parameter granularity refinement.

Constraint parameter	Final value of constraint parameter	Classification correct rate
P_1	0.88	84.62%
P_2	3.99	83.70%
P_3	3.00	82.20%
P_4	1.00	90.67%
P_5	8.54	91.67%
P_6	8.30	84.00%
P_7	0.57	92.86%
P_8	4.29	93.33%
P_9	5.26	81.82%
P_{10}	0.75	97.25%
P_{11}	4.30	94.23%
P_{12}	0.18	97.78%
P_{13}	0.12	100.00%
P_{14}	0.13	100.00%
P_{15}	2.07	93.33%
P_{16}	1.21	90.91%
P_{17}	0.50	95.83%
Total		90.80%

As shown in Table 7, the TCR is not improved while some constraint parameters are changing. This shows that some constraint parameters have little effect on TCR. To more clearly observe the process of TCR rising with iteration, the iteration data by which TCR is not changed is deleted. We can get a new correspondence table between constraint parameters and TCR during iteration.

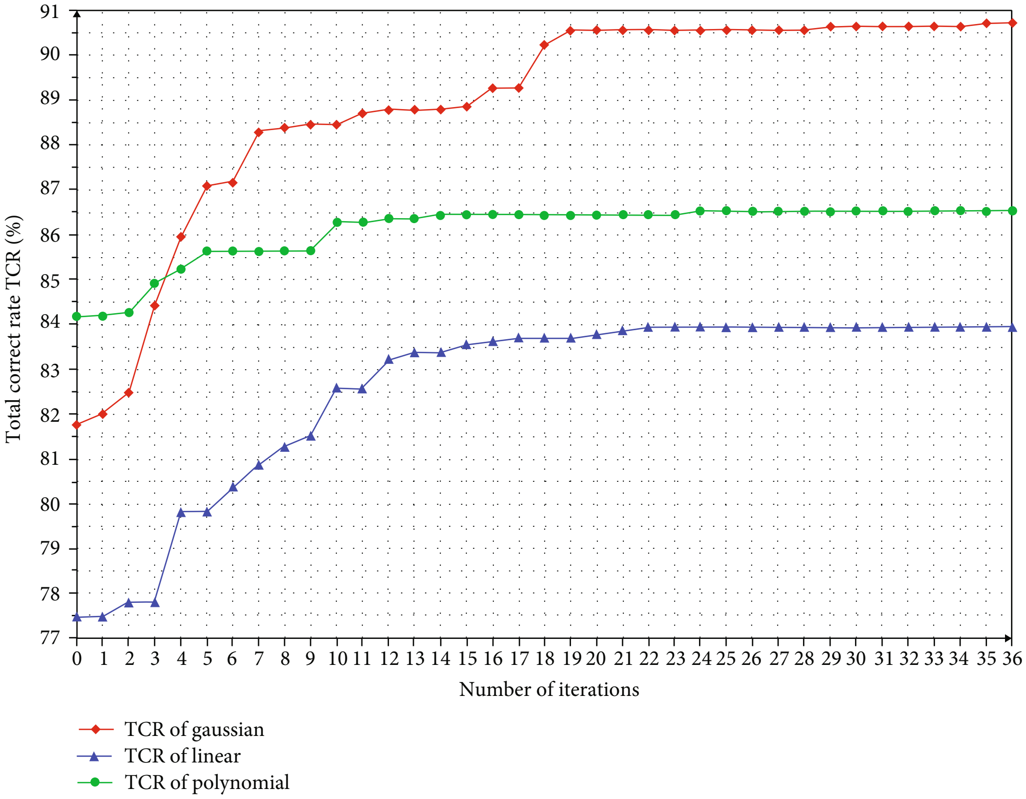


Figure 12: Schematic diagram of the change of TCR with the number of iterations under different kernel functions.

As shown in Table 8, the TCR is increased with the iteration process. And the increasing rate of TCR is faster at the beginning and slower in the follow-up. It can be seen that more and more test flag pictures are correctly classified with the change of constraint parameters and changing the constraint parameters is more effective at the beginning. From the data of Table 7, the 2D line chart can be obtained.

As shown in Figure 11, the increasing rate is different with different constraint parameters of SVM binary classifier. And the TCR is rising alternately with the change of constraint parameter. In this way, the TCR is increasing gradually, and the unstable oscillation during TCR lifting is avoided.

6.4.2. Effect Analysis of Parameter Granularity Refinement. In this part, the granularity of the constraint parameter will be smaller and will be changed from 0.01 to 0.001. To try to find more optimized constraint parameters for a higher classification correct rate.

The search range is limited to the range of 0.1 above and below the final value which is shown in Table 5. In this way, the calculation time can be saved, and there is no need to search the interval [0.001, 10] with the granularity of 0.001. As shown in Table 5, the final value of the constraint parameter for 00001 SVM binary classifier is 4.1, so the search range is in the interval [4.0, 4.2] with a granularity of 0.01. And the search order of constraint parameters is consistent with the previous method which has been elaborated in Ascending Spiral Training Method of Traffic Sign Recognition Model.

As shown in Table 9, the TCR is 90.80%, and it is increased little by refining the constraint parameter granularity. This shows that the change of constraint parameter granularity is less effective and the early used granularity (0.1) is satisfied the application scenario.

6.4.3. Compare the Classification Correct Rate by Using Different Kernel Function. In this part, the Gaussian kernel function, linear kernel function, and polynomial kernel function are used for comparing the classification effect.

As is shown in Figure 12, the final TCR varies from different kernel functions. It is clear that the final TCR using Gaussian kernel function is equal to the final TCR using RBF kernel function, and the final TCR using polynomial kernel function is second, and the final TCR using linear kernel function is the lowest. Judging from the rising speed of TCR, the model using Gaussian kernel function is better than the other two methods which means that we can take less time to train the network. Also, it shows that the RBF kernel function and the Gaussian kernel function are more suitable for the classification method proposed in this paper. On the other hand, the final TCR shows that it performs well enough in the case of only a single-layer SVM network.

6.4.4. Compare with the Methods of GTSRB. We can get the results for IJCNN 2011 competition (1st stage) from the INI Benchmark Website [31]. There are 190 types of results submitted for the final GTSRB data set. For comparative analysis, we make statistics on the methods used by all participating teams.

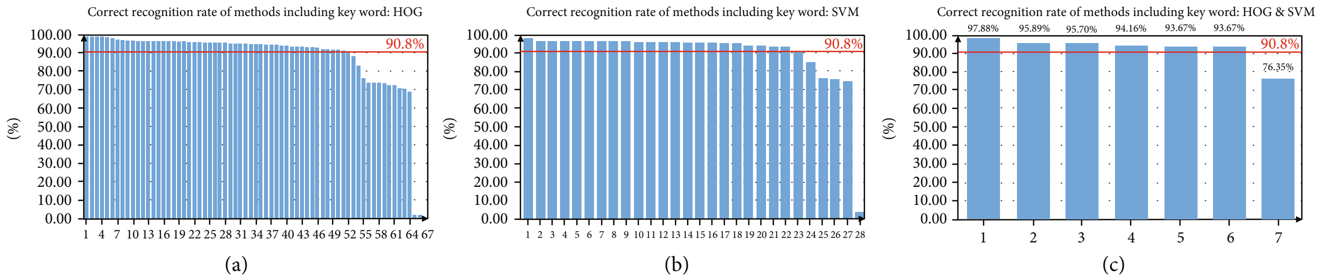


Figure 13: The correct recognition rate of methods including HOG, SVM, and HOG&SVM.

As shown in the figure, Figure 13(a) shows the correct recognition rate of 67 methods including the keyword HOG. Figure 13(b) shows the correct recognition rate of 28 methods including the keyword SVM. Figure 13(c) shows the correct recognition rate of 7 methods including the keyword HOG&SVM. The red line shows the correct recognition rate of the method proposed in this paper. Although the correct recognition rate of multiple methods is better, these methods used a multilayer network model, not a single-layer network. And it will take more time to use more complex methods to train the recognition model. The PTSR model proposed in this paper which takes less time to be trained has a good performance in the subset of GTSRB data set. Especially, one method shown in Figure 13(c) is called HOG_SVM, and its correct recognition rate is 76.35% in the subset of the GTSRB data set. And the correspondence relations between numbers and methods in Figure 13 could be found in Table S1, Table S2, and Table S3 in the supplemental file.

7. Conclusion and Future Work

In conclusion, for recognizing prohibition traffic signs, a classification model based on HOG feature and SVM network is proposed. The classification model consists of three parts, which are the image preprocessing part, the initial binary classification part, and the final classification part based on the predicted result. The classification model is validated. Analyzing the test result data, the validity of the model is confirmed.

In future work, on the one hand, we will pay attention to the completeness of training samples. Let the training samples cover the changes of prohibition traffic sign image in deformation, illumination, and occlusion as much as possible, and improve the classification performance of SVM binary classifiers, and enhance the accuracy of a single SVM binary classifier. On the other hand, focusing on the appropriateness of image size and the optimal image size will be considered, so that the HOG features can fully describe the details of the image and will not greatly increase the dimension of HOG features because the image size is larger, so as to improve the training efficiency of SVM classifier as much as possible. Also, the granularity of constraint parameters can be smaller (such as 0.01) which is mainly 0.1 in this paper, and the search range of constraint parameters can be wider which is in the interval [0,10] in this paper. Also, the kernel function of SVM binary classifier should not be limited to the RBF, Gaussian, linear, and polynomial kernel function, and other kernel functions can be chosen, and a different kernel function could be used for each SVM binary classifier that has different types of kernel functions. Of course, fusing other traffic sign classification and recognition methods to improve the accuracy of classification is also an important aspect of future research work.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yang Liu and Wei Zhong contributed equally to this work.

Acknowledgments

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K20201290).

Supplementary Materials

Figure S1: schematic diagram of prohibition traffic signs. Displaying 36 standard prohibition traffic sign pictures to help get a better understanding of prohibition traffic signs. Figure S2: samples of verification data. Displaying 15 different prohibition traffic sign samples to help get an impression of the verification data set. Table S1: corresponding table of numbers and methods in Figure 13(a). Explaining the correspondence relations between different numbers and HOG-based methods in Figure 13(a). Table S2: corresponding table of numbers and methods in Figure 13(b). Explaining the correspondence relations between different numbers and SVM-based methods in Figure 13(b). Table S3: corresponding table of numbers and methods in Figure 13(c). Explaining the correspondence relations between different numbers and SVM&HOG-based methods in Figure 13(c). *(Supplementary Materials)*

References

[1] C. Liu, S. Li, F. Chang, and Y. Wang, “Machine vision based traffic sign detection methods: review, analyses and perspectives,” *IEEE Access*, vol. 7, pp. 86578–86596, 2019.

[2] A. Ruta, Y. M. Li, and X. H. Liu, “Robust class similarity measure for traffic sign recognition,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 846–855, 2010.

[3] S. Dong, P. Wang, and K. Abbas, “A survey on deep learning and its applications,” *Computer Science Review*, vol. 40, p. 100379, 2021.

[4] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[5] Y. Wang, D. X. Zhang, Y. Liu, B. Dai, and L. H. Lee, “Enhancing transportation systems via deep learning: a survey,” *Transportation Research*, vol. 99, pp. 144–163, 2019.

[6] C. Badue, R. Guidolini, R. V. Carneiro et al., “Self-driving cars: a survey,” *Expert Systems with Applications*, vol. 165, article 113816, 2021.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[8] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural Networks*, vol. 32, pp. 333–338, 2012.

[9] J. M. Zhang, Q. Q. Huang, H. L. Wu, and Y. K. Liu, “A shallow network with combined pooling for fast traffic sign recognition,” *Information*, vol. 8, no. 2, p. 45, 2017.

[10] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “The German traffic sign recognition benchmark: a multi-class classification competition,” in *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 1453–1460, San Jose, CA, USA, 2011.

[11] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “Man vs. computer: benchmarking machine learning algorithms for traffic sign recognition,” *Neural Networks*, vol. 32, pp. 323–332, 2012.

[12] Z. Zhu, D. Liang, S. H. Zhang, X. Huang, B. Li, and S. Hu, “Traffic-sign detection and classification in the wild,” in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2110–2118, Las Vegas, USA, 2016.

[13] F. Larsson and M. Felsberg, “Using Fourier descriptors and spatial models for traffic sign recognition,” in *Scandinavian Conference on Image Analysis*, vol. 6688 of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), , pp. 238–249, Springer, Berlin, Heidelberg, 2011.

[14] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, “Traffic sign recognition — how far are we from the solution?,” in *Proceedings of International Joint Conference on Neural Networks*, pp. 1–8, Dallas, TX, USA, 2013.

[15] C. G. Serna and Y. Ruichek, “Classification of traffic signs: the European dataset,” *IEEE Access*, vol. 6, pp. 78136–78148, 2018.

[16] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition*, pp. 886–893, San Diego, CA, USA, 2005.

[17] Z. Y. Huang, Y. L. Yu, J. Gu, and H. P. Liu, “An efficient method for traffic sign recognition based on extreme learning machine,” *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 920–933, 2017.

[18] H. L. Luo, Y. Yang, B. Tong, F. C. Guo, and B. Fan, “Traffic sign recognition using a multi-task convolutional neural network,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1100–1111, 2018.

[19] A. Ellahyani, M. El Ansari, and I. El Jaafari, “Traffic sign detection and recognition based on random forests,” *Applied Soft Computing*, vol. 46, pp. 805–815, 2016.

[20] C. Yao, F. Wu, and H. J. Chen, “Traffic sign recognition using HOG-SVM and grid search,” in *International Conference on Signal Processing Proceedings*, pp. 962–965, Hangzhou, China, 2015.

[21] R. Z. Junges, M. Paula, and M. Aguiar, “Brazilian traffic signs detection and recognition in videos using CLAHE, HOG feature extraction and SVM cascade classifier with temporal coherence,” in *In Mexican International Conference on Artificial Intelligence*, Springer, Cham, 2019.

[22] M. Tun and T. Lwin, “Real time Myanmar traffic sign recognition system using HOG and SVM,” *International Journal of Trend in Scientific Research and Development*, vol. 3, no. 5, pp. 2367–2371, 2019.

[23] J. L. Tang, Q. L. Su, and C. Y. Lin, “Traffic sign recognition based on HOG feature and SVM,” in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*, pp. 534–538, New York, 2020.

[24] D. Cotovanu, C. Fosalau, C. Zet, and M. Skoczylas, “Detection of traffic signs based on support vector machine classification using HOG features,” in *Proceedings od 2018 International Conference and Exposition on Electrical And Power Engineering (EPE)*, pp. 518–522, Iasi, Romania, 2018.

[25] Standardization Administration of P R C, *Part 1: General, Road Traffic Signs and Markings*, GB 5678.1-2009, China Standard Press, Beijing, China, 2009.

[26] Y. Liu, D. R. Huang, Y. Liu, and W. Zhong, “Traffic sign image color standardization based on multi color space cascade classification,” *Computer Engineering*, vol. 46, no. 9, pp. 233–241, 2020.

[27] Y. Liu, Y. Liu, and W. Zhong, “Research on color normalization of traffic sign image based on surface segmentation and improved K-means clustering algorithm,” in *Proceedings of CSAE 2020: The 4th International Conference on Computer Science and Application Engineering*, New York, 2020.

[28] P. V. C. Hough, “Method and means for recognizing complex patterns,” USA Patent, 3069654, 1962.

[29] L. Xu, E. Oja, and P. Kultanen, “A new curve detection method: randomized Hough transform (RHT),” *Pattern Recognition Letters*, vol. 11, no. 5, pp. 331–338, 1990.

[30] J. M. Zhang, M. T. Huang, X. K. Jin, and X. D. Li, “A real-time Chinese traffic sign detection algorithm based on modified YOLOv2,” *Algorithms*, vol. 10, no. 4, p. 127, 2017.

[31] Institute of Neuro informatics, “Results for IJCNN 2011 competition (1st stage),” 2022, https://benchmark.ini.rub.de/gtsrb_results_ijcnn.html.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 9735887, 16 pages
https://doi.org/10.1155/2022/9735887

Research Article

Moving Object Tracking in Satellite Videos by Kernelized Correlation Filter Based on Color-Name Features and Kalman Prediction

Wenjing Pei and Xuhui Lu

The Seventh Research Division and The Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing 100191, China

Correspondence should be addressed to Xuhui Lu; luxuhui2020@buaa.edu.cn

Received 21 April 2022; Accepted 21 May 2022; Published 9 June 2022

Academic Editor: Maode Ma

This paper studies moving object tracking in satellite videos. For the satellite videos, the object size in the images may be small, the object may be partly occluded, and the image may contain an area resembling dense objects. To handle the above problems, this paper puts forward a kernelized correlation filter based on the color-name feature and Kalman prediction. The original image is mapped to the color-name feature space so that the tracker can process the image with multichannel color features. The Kalman filter is used to predict the moving object position in the tracking process, and the detection area is determined according to the predicted position. The Kalman filter is updated with the detection results to improve the tracking accuracy. The proposed algorithm is tested on Jilin-1 datasets. Compared with the other seven tracking algorithms, the experiment results show that the proposed algorithm has stronger robustness for several complex situations such as rapid target motion and similar object interference. Besides, it is also shown that the proposed algorithm can prevent the problem of tracking failure when the moving object is partially occluded.

1. Introduction

With the development of remote sensing technologies, the earth observation satellites are extensively applied to several fields [1–7]. However, traditional earth observation satellites can only take a single image of a certain area. The valuable and interesting dynamic data in the object area is hard to obtain only based on the static medium and high resolution optical remote sensing data, which may limit the reconnaissance capability of the earth observation satellites in emergencies [8].

On the other hand, video satellites can overcome the limitation of traditional earth observation satellites and can obtain high time-resolution images [9–11]. Video satellites are used to obtain a series of images of fixed object areas on the ground. Therefore, the videos can be formed to obtain the dynamic object information directly. Currently, the continuous monitoring capability of high-resolution satellites in a certain time range has been realized [12]. Due to the above advantages, video satellites are used to observe and track the states of moving objects on the ground [13] and have wide application potential in the fields of vehicle real-time monitoring [8, 14, 15], rapid response to natural disaster emergency [16], major engineering monitoring [17], and so on. In recent years, some representative video satellites are Skysat-1 and Skysat-2 in the United States [18, 19], TUBSAT series satellites by the Technical University of Berlin [20] and Jilin-1 [7, 21] and Tiantuo-2 (TT-2) [22] in China.

Besides, moving object tracking (MOT) in traditional videos has been a research hotspot in computer vision. MOT is widely used in automatic monitoring, automatic driving, human-computer interaction, and so on [23, 24]. The task of MOT is to predict the size and position of the object in subsequent frames based on the size and position of the object in the initial frame. Much research has been done in this aspect, and numerous algorithms have been investigated for accurate tracking in ordinary videos [25–30]. Notice that the commonly used visual object tracking methods can be divided into generation methods [25–27] and discrimination methods [28–30].

On the one hand, the generative methods establish the object model and describe the real-world object, so as to search for the position with the highest similarity with the object template in the next frame [31]. Many breakthroughs have been made and several generative methods have been utilized in MOT, including mean shift, sparse coding, dictionary learning, particle filter, and sliding window [32–37]. On the other hand, compared with generative methods, the discriminative methods regard object tracking as a detection problem, also known as tracking by detection (see [38, 39] and the related references). The discriminative methods generally train the classifier in the first frame to separate the object from its surrounding background. In particular, considering the complex background such as background changes, the discriminative methods establish an online discrimination classifier model to distinguish objects from cluttered backgrounds to provide more effective features and avoid unwanted model drift. Recently, some representative machine learning techniques are adopted into the discriminative methods, such as Boosting, Support Vector Machines (SVM), Multiple Instance Learning (MIL), Random Forests, Semisupervised Learning, and Structured Output Support Vector Machines (SOSVM) [40–42]. A theoretical framework of dense sampling in tracking-by-detection is presented [43]. In [44], a tracking learning detection (TLD) algorithm is also proposed, where learning and detection are introduced into the long-term tracking of the objects in the videos to enhance the tracking accuracy. Hare et al. [45] propose an adaptive tracking-by-detection method called STRUCK, which exploits Gaussian kernels and SVM as a structured output to accurately locate the objects. A cooperative model tracking algorithm based on sparse representation is presented, which is suitable for the situation of object occlusion and blur [46].

Apart from the machine-learning-based discriminative methods, the kernelized correlation filter (KCF) can also accelerate the calculation speed and improve tracking accuracy simultaneously [43, 47, 48]. In particular, the KCF-based method can efficiently handle the object in changing environment [47]. A scale adaptive tracker is proposed based on the separate discriminative correlation filters, where the computational cost is reduced [49]. Galoogahi et al. [50] demonstrate a background-aware correlation filter (BACF) for real-time visual tracking, where the handcrafted features are introduced to effectively describe the changing background. In [48], a KCF and the Normalized Cross-Correlation (NCC) template matching is proposed for long-term target tracking of UAVs to improve the tracking performance.

In addition, note that the moving objects in satellite videos include vehicles, airplanes, rockets, and ships. Compared with moving target tracking in ordinary videos, real-time object tracking in satellite videos should overcome three main challenges as follows [51, 52].

(1) Compared with the ordinary videos on the ground, the object size is small in satellite videos, which may lose their effective features (see Figure 1). For instance, the length of a car is about 4-5 m in real life, while it is about four pixels in satellite videos. The size of these objects in the satellite videos is so small that it may be difficult to track these objects reliably

(2) Since remote sensing images have a relatively large field of vision, there is low contrast between the background and the objects (see Figure 1(a))

(3) The background of the image sequence of the satellite videos may be fuzzy and chaotic (see Figure 1(b)). Besides, there may be more than one moving target in the image sequence, which have the characteristics of high similarity, serious mutual interference, and low resolution. This will result in partial or complete occlusion between moving targets (see Figure 1(c))

Due to these above challenges and the increasing demand for MOT in the field of remote sensing, up until now, several moving object detection/tracking algorithms have been carefully designed for MOT in satellite videos [51–55]. Lei and Guo [53] propose a road masking and Gaussian mixture method to achieve multiple object detection and tracking of the remote sensing video satellite. Meanwhile, the method can improve the reconnaissance capability of the remote sensing satellite for the dynamic mobile small target. In [51], a fusion tracker is introduced where the kernel correlation filter and the three-frame-difference method are synthesized for satellite videos to improve the performance of the tracker. Li and Man [52] put forward an optical-flow-based detection algorithm with video attention saliency for the moving ships in the satellite videos, where the Gabor filter is utilized to extract the texture feature, and the registration of the sea-scene images can be avoided. Liu et al. [54] present a kernelized correlation filter where multifeature fusion and motion trajectory compensation are employed for satellite videos to mitigate the tracking drifts. In [55], an object tracking algorithm is also proposed for the high-resolution multispecial satellite images with multiangular observation capability, where a novel regional operator is constructed and the tracking capability is verified in the WorldView-2 satellite images.

However, the above trackers [51, 52, 54, 55] generally rely on the original pixel information (such as HSV or HOG), which ignores the color information. In fact, several satellite videos contain some color-name (CN) features. Compared with other color features, CN features show better discriminative capability in MOT [56, 57]. In addition, due to undesirable environmental factors, such as shadows, similar backgrounds, and other interferences, it will become more complicated to realize the MOT.

On this foundation, aiming at MOT in satellite videos, this paper proposes the kernelized correlation filters based on color-name features and Kalman prediction (CNK-KCF). The images are mapped to CN feature space so that the tracker can process the image with multichannel color

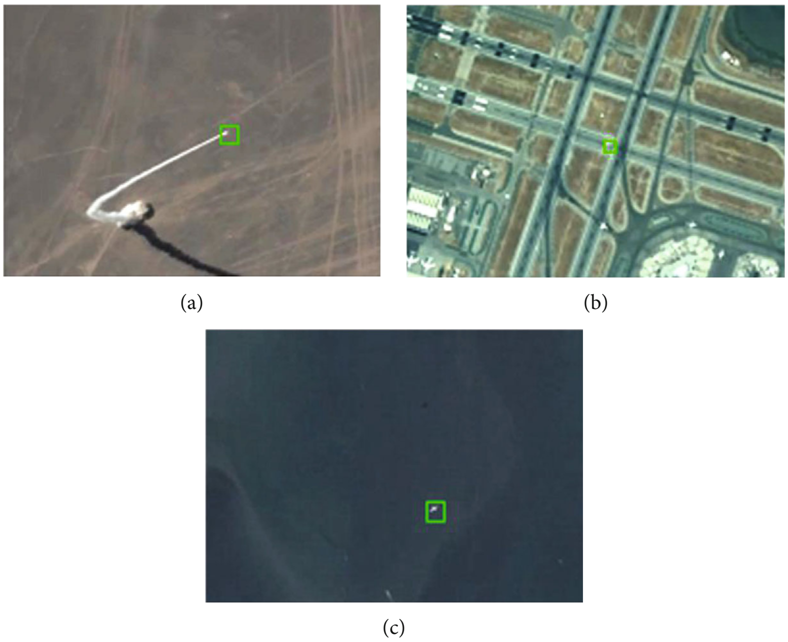


Figure 1: The moving objects in satellite videos. (a) There is low contrast between the background and the object. (b) The background is fuzzy and chaotic. (c) The target is partially occluded.

features. Moreover, the KCF can improve the calculation speed in the Fourier translation based on the cyclic matrix and kernel trick. Meanwhile, the Kalman filter can help correct and update the predicted position of the moving object and, together with the kernelized correlation filters based on the color-name features (CN-KCF), can improve the tracking accuracy. The proposed algorithm is tested on Jilin-1 datasets. Experimental results are analyzed and show that the proposed method is robust to some environmental factors such as partial occlusion, background similarity, and rapid motion.

In summary, the contributions of this paper are twofold.

(1) For MOT in satellite videos, a framework named CNK-KCF is carefully designed based on the CN feature and the Kalman filter. Besides, the proposed CNK-KCF algorithm is stable in complex situations such as rapid target motion, occlusion, and similar object interference, which can solve the problem of tracking failure when a moving object is partially occluded

(2) In the experiment section, the CNK-KCF algorithm is compared with other algorithms, and it is shown that the CNK-KCF algorithm possesses better tracking accuracy and success rate for the airplane, rocket, and ship in Jilin-1 satellite videos. The performance of each type of tracker in satellite videos is analyzed in detail

The rest of this paper is organized as follows. Section 2 presents the design of the CNK-KCF for satellites videos MOT. Section 3 introduces the experiment results and some analysis. Finally, Section 4 concludes this article.

2. Materials and Methods

In this section, to solve the problem of MOT in satellite videos, the CNK-KCF is developed carefully. As shown in Figure 2, the CNK-KCF mainly consists of 3 parts: (1) KCF [58, 59], (2) CN, and 3) KF prediction. In the rest of this section, each part is described in one subsection each.

2.1. Kernelized Correlation Filter (KCF). The KCF has a relatively low computational cost and relatively high tracking accuracy, especially for rapid deformation. KCF is also robust to illumination changes. Hence, KCF is suitable for MOT in satellite videos. The procedures of the KCF tracking algorithm are shown as follows. Firstly, in order to construct the tracking area, the tracking object is selected from the initial frame in the satellite videos. Then, according to the cyclic matrix theory, the tracking area is cyclically shifted. The kernel function is applied to calculate the similarity between the possible region of the target location and the tracking objects. Finally, the area with the largest output response is selected as the new target, and the classifier is trained based on the Fourier transform to reduce the calculation time.

In the KCF, the following regression function

$$f(x) = \omega^T z \tag{1}$$

is trained to obtain weight coefficients $\omega = [\omega^1, \omega^2, \cdots, \omega^n]^T$, where $z = [z^1, z^2, \cdots, z^n]^T$ is an n-dimensional vector. Correspondingly, the cost function can be minimized as

$$f(x) = \min_{\omega} (X\omega - y)^T (X\omega - y) + \lambda \|\omega\|^2, \tag{2}$$

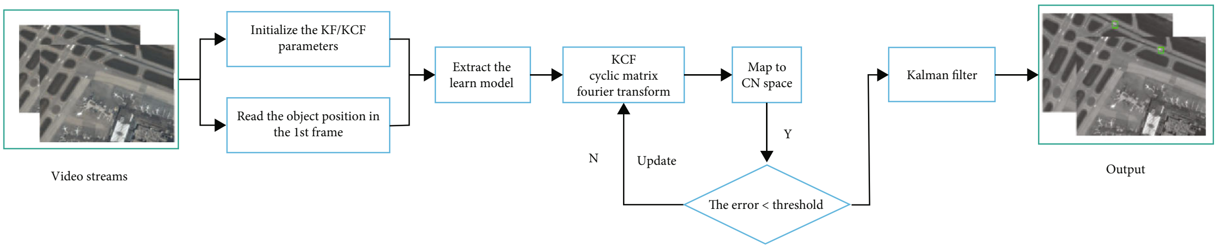

Video streams
Initialize the KF/KCF parameters
Read the object position in the 1st frame
Extract the learn model
KCF cyclic matrix fourier transform
Map to CN space
Y
The error < threshold
N
Update
Kalman filter
Output

Figure 2: The flow charts of our proposed tracker (CNK-KCF).

where X is the data matrix, y is the desired output, and λ is the regularization parameter to prevent overfitting. Based on [64, 65], the extreme value of Equation (2) can be obtained as

$$\omega = \left(X^T X + \lambda I\right)^{-1} X^T y, \tag{3}$$

where I is an identity matrix. In Equation (3), calculating the inverse matrix $((X^H X + \lambda I)^{-1})$ is very time-consuming. Therefore, the calculation is performed in the Fourier domain, and Equation (3) can be rewritten in a complex field as

$$\omega = \left(X^H X + \lambda I\right)^{-1} X^H y, \tag{4}$$

where X^H represents the Hermitian transpose of X. Obviously, if X is a real matrix, Equation (4) can be considered as equivalent to Equation (3).

Besides, in order to accelerate the calculation speed of MOT, the cyclic shift is also introduced. The cyclic shift operator is a permutation matrix, which can be used to simulate the one-dimensional translation of this vector. The permutation matrix can be shown as

$$P = \begin{bmatrix} 0 & 0 & 0 & & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{5}$$

so the cyclic shift of x can be presented as

$$Px = \left[x^n, x^1, \cdots, x^{n-1}\right]^T. \tag{6}$$

Since the product Px shifts x by one element, u shifts are used to chain with the matrix power $P^u x$ and achieve more translations. The same signal x can be obtained periodically every n-time translation based on the cyclic property of the cyclic matrix, which means that all shifted signals can be represented as

$$\{P^u x \mid u = 0, 1, 2, \cdots, n-1\}. \tag{7}$$

Correspondingly, the data matrix X can be the circulant matrix and is denoted as

$$X = C(x) = \begin{bmatrix} x^1 & x^2 & x^3 & & x^n \\ x^n & x^1 & x^2 & \cdots & x^{n-1} \\ x^{n-1} & x^n & x^1 & & x^{n-2} \\ & \vdots & & \ddots & \vdots \\ x^2 & x^3 & x^4 & \cdots & x^1 \end{bmatrix}. \tag{8}$$

One-dimensional vector cyclic displacement is given in Figure 3. All cyclic matrices can be diagonalized through the discrete Fourier transform (DFT), which is independent of the generated vector X. Therefore, X can be diagonalized as

$$X = C(x) = F \operatorname{diag}(\hat{x}) \mathrm{F}^H, \tag{9}$$

where the constant matrix F is known as the discrete DFT that does not depend on x. The Hermitian transpose of F is represented as F^H. The matrix diag $(\bullet)$ is the diagonal matrix. Accordingly, the vector $\hat{x}$ is the DFT of x and is defined as

$$\hat{x} = F(x) = \sqrt{n} F x. \tag{10}$$

In the following section, the DFT of the vector will be represented by the hat (^). Due to the diagonalization property of the matrix diag $(\bullet)$, the matrix X^H can be obtained as

$$\hat{\mathrm{x}} = X^H = (X^*)^T = \left(F^* \operatorname{diag}(\hat{x}^*) F^{H*}\right)^T = F \operatorname{diag}(\hat{x}^*) F^H, \tag{11}$$

where X^* is the complex-conjugate of X and $(\bullet)^*$ can be defined as a conjugate symbol. In addition, $X^H X$ is represented to be a noncentral covariance matrix. From (9) and (11), it follows that

$$X^H X = F \operatorname{diag}(\hat{x}^*) F^H F \operatorname{diag}(\hat{x}) F^H. \tag{12}$$

Because the diagonal matrices are symmetric, the Hermite transpose is used only after complex-conjugation $\hat{x}$. In this way, the factor $F^H F = I$ can be eliminated. Moreover, the operations on diagonal matrices are done by elements,

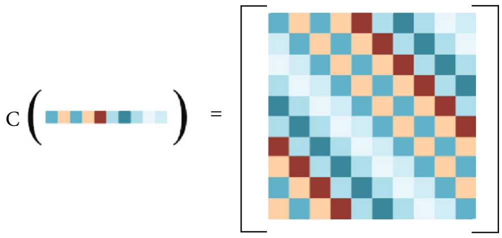


Figure 3: One-dimensional vector cyclic displacement.

so Equation (12) can be rewritten as

$$X^H X = F\ \mathrm{diag}\ (\hat{x}^*)\ \mathrm{diag}\ (\hat{x}) F^H = X^H X = F\ \mathrm{diag}\ (\hat{x}^* \odot \hat{x}) F^H, \tag{13}$$

where $\odot$ is the dot product of two vectors. It can be seen in (13) that the original complex matrix operation is transformed into a simple vector and dot product operation based on the diagonalization property of the cyclic matrix. Based on (4) and (13), the DFT of ω is obtained as

$$\hat{\omega} = \mathrm{diag}\ \left(\frac{\hat{x}^*}{\hat{x}^* \odot \hat{x} + \lambda}\right)\hat{y} = \frac{\hat{x}^* \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda}, \tag{14}$$

where fractions represent the division element. ω can be recovered in the spatial domain on the basis of the inverse DFT.

In addition, a nonlinear mapping function $\varphi(x)$ is employed to make the mapped samples linearized in the new space. Therefore, $f(x)$ can be transformed into

$$f(x) = \omega^T \varphi(x). \tag{15}$$

Furthermore, the kernel technique can be used to map the input of a linear problem to a nonlinear feature space $\varphi(x)$. Correspondingly, the solution ω is rewritten as

$$\omega = \sum_i \alpha_i \varphi(x^i), \tag{16}$$

where is x^i a column vector. Therefore, the parameters of the solution are changed from ω to α. Meanwhile, Equation (2) can be rewritten as

$$\min_{\omega} \sum_i (\varphi(X)\omega - y)^T (\varphi(X)\omega - y) + \lambda \|\omega\|^2. \tag{17}$$

Besides, the dot products can be defined as $\varphi^T(x) \odot \varphi(x')= \kappa(x, x')$, where $\kappa(x, x')$ is the Gaussian kernel expressed as

$$\kappa\left(x, x'\right) = \exp\left(-\frac{1}{\sigma^2}\left(\|x\|^2 + \left\|x'\right\|^2 - 2\mathscr{F}^{-1}\left(\hat{x} \odot \hat{x}'\right)\right)\right), \tag{18}$$

where $\mathscr{F}^{-1}$ is a mapping function in the high dimensional space. Furthermore, the dot products $\varphi^T(x) \odot \varphi(x') = \kappa(x, x')$ between all pairs of samples are stored in a $n \times n$ kernel matrix $K \triangleq [K_{ij}]_{n\times n}$, which can be defined as

$$K_{ij} = \kappa\left(x_i, x_j\right). \tag{19}$$

The complexity of the regression function increases with the increase of the sample size, so that $f(X)$ can be rewritten based on (16) as

$$f(X) = K^T \alpha. \tag{20}$$

Meanwhile, Equation (17) can be rewritten as

$$\min_{\omega} \left(K^T \alpha - y\right)^T \left(K^T \alpha - y\right) + \lambda \alpha^T K \alpha. \tag{21}$$

The optimal solution of (21) is given by

$$\alpha = (K + \lambda I)^{-1} y. \tag{22}$$

If the kernel function satisfies Equation (22), K is circulant for any permutation matrix M as

$$\kappa\left(x, x'\right) = \kappa\left(Mx, Mx'\right). \tag{23}$$

In addition, if the kernels can make K circulant, Equation (22) can be diagonalized as

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda}, \tag{24}$$

where $k^{xx'}$ is the first row of the kernel matrix K defined as

$$k^{xx'} = [k(x_1, x_1), k(x_1, x_2), \cdots, k(x_1, x_n)]. \tag{25}$$

Furthermore, K is a circulant matrix and can be expressed as

$$K = C(k^{xx}), \tag{26}$$

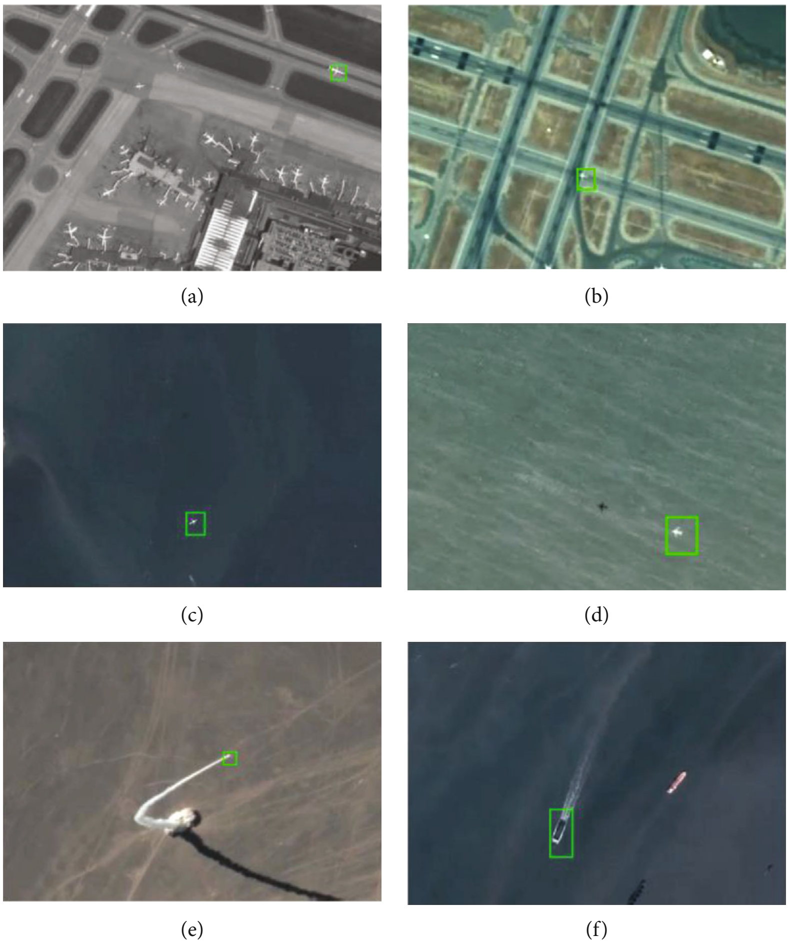
so it further follows that

$$f(X) = C(k^{xx})^T \alpha = \left(F\ \mathrm{diag}\ \left(\hat{k}^{xx}\right) F^H\right)^T \alpha, \tag{27a}$$

$$\hat{f}(X) = \mathrm{diag}\ \left(\hat{k}^{xx}\right) \odot \alpha. \tag{27b}$$

2.2. KCF Tracking Algorithm Based on Color-Name (CN) Feature.

In the traditional KCF algorithm, the original pixel and the directional gradient histogram (HOG) feature are required, and the original pixel features are utilized to convert the images into gray images. The pixel gray value is regarded as the image feature. However, the generation process of the HOG feature descriptor is lengthy, which results in slow speed and poor real-time performance for MOT in satellite videos. In addition, the HOG feature descriptor is relatively sensitive to noise, so it is difficult to deal with occlusion based on the gradient characteristics. Compared



(a)

(b)

(c)

(d)

(e)

(f)

Figure 4: Overview of the MOT in our experiments. There are six scenes of the satellite videos. In every scene, there is only one object. (a) The moving airplane-1 is selected as the object; (b) the moving airplane-2 is selected as the object; (c) the moving airplane-3 is selected as the object; (d) the moving airplane-4 is selected as the object; (e) the moving rocket is selected as the object; (f) the moving ship is selected as the object.

Table 1: The size of the objects in satellite videos.

The object	Airplane-1	Airplane-2	Airplane-3	Airplane-4	Rocket	Ship
Size (pixels)	44 × 44	28 × 28	28 × 28	22 × 22	46 × 46	30 × 30

Table 2: The precision and success rate in comparison with other seven trackers for airplane-1.

Methods	KCF	K-KCF	CN-KCF	CSK	STRUCK	MeanShift	CamShift	CNK-KCF (ours)
Precision	0.929	0.888	0.907	0.943	0.889	0.621	0	0.957
Success rate	0.873	0.816	0.841	0.895	0.827	0.495	0.020	0.917

Table 3: The precision and success rate in comparison with other seven trackers for airplane-2.

Methods	KCF	K-KCF	CN-KCF	CSK	STRUCK	MeanShift	CamShift	CNK-KCF (ours)
Precision	0.732	0.284	0.939	0.955	0.272	0.932	0	0.967
Success rate	0.632	0.260	0.856	0.881	0.224	0.824	0.0003	0.902

Table 4: The precision and success rate in comparison with other seven trackers for airplane-3.

Methods	KCF	K-KCF	CN-KCF	CSK	STRUCK	MeanShift	CamShift	CNK-KCF (ours)
Precision	0.963	0.952	0.959	0.958	0.977	0.220	0	0.981
Success rate	0.889	0.866	0.878	0.873	0.925	0.186	0.010	0.935

Table 5: The precision and success rate in comparison with other seven trackers for airplane-4.

Methods	KCF	K-KCF	CN-KCF	CSK	STRUCK	MeanShift	CamShift	CNK-KCF (ours)
Precision	0.691	0.848	0.973	0.973	0.970	0.936	0	0.974
Success rate	0.506	0.632	0.917	0.916	0.890	0.803	0.0003	0.918

Table 6: The precision and success rate in comparison with other seven trackers for the rocket.

Methods	KCF	K-KCF	CN-KCF	CSK	STRUCK	MeanShift	CamShift	CNK-KCF (ours)
Precision	0.962	0.950	0.980	0.889	0.960	0.050	0	0.980
Success rate	0.903	0.879	0.941	0.822	0.898	0.075	0.048	0.941

Table 7: The precision and success rate in comparison with other seven trackers for the ship.

Methods	KCF	K-KCF	CN-KCF	CSK	STRUCK	MeanShift	CamShift	CNK-KCF (ours)
Precision	0.640	0.698	0.975	0.978	0.962	0.458	0.279	0.979
Success rate	0.563	0.625	0.948	0.954	0.931	0.426	0.251	0.954

Table 8: The basic parameters of the algorithms.

Methods	σ	λ	η
KCF	0.5	0.0001	0.02
K-KCF	0.5	0.0001	0.02
CN-KCF	0.2	0.01	0.075
CSK	0.2	0.01	0.075
STRUCK	0.2	0.01	0.075
MeanShift	0.2	0.01	0.075
CamShift	0.2	0.01	0.075
CNK-KCF (ours)	0.2	0.01	0.075

with the original pixel feature and HOG feature, the CN feature has better stability properties. As a result, the proposed algorithm applies a CN statistical feature to extract the image feature.

CN feature space is a special color space based on a potential probability model. CN space contains 11 color channels: yellow, red, black, blue, gray, pink, white, brown, green, orange, and purple. In the form of mathematical description, a color image $M(x)$ represents the color pixel value at position x, and the image is mapped to CN space so that $M(x)$ can be converted into an 11-dimensional (11D) probability feature vector $f(x)$. Specifically, the RGB value is represented by an 11D color with a total probability sum of 1, so as to realize the low-dimensional extraction of color information. The model can be expressed as follows:

$$M(x_0) = \arg \max_i \left\{ i \mid \sum_{x \in \Omega_c(x_0)} \phi_i(x) \bullet N(x_0, \sigma), i = 1, 2, \cdots, 11 \right\}, \tag{28}$$

where Ω_c presents a region with x_0 as the center and c as the radius. $N(\bullet)$ is a Gaussian function. σ is the standard deviation. However, in the process of tracking, since not all of useful object information can be provided by the 11D color attributes, the 11D color attributes are firstly reduced to 10D. Subsequently, the PCA is employed to reduce 10D to 2D, which can reduce calculation and accelerate the calculation speed of the algorithm.

We supposed that x_{CN}^k is the CN feature extracted from the target region in the kth frame and $\widehat{x}_{CN}^k$ is the Fourier transform of x_{CN}^k. To reduce dimensions, the dimension reduction matrix is given as

$$\tilde{x}_{CN}^k = B_k \widehat{x}_{CN}^k, \tag{29}$$

where $\tilde{x}_{CN}^k$ is the CN feature through the dimension-reduction operation in the kth frame and B_k is a dimension-reduction matrix. The reconstructed minimum cost function as the decision function is given to obtain the

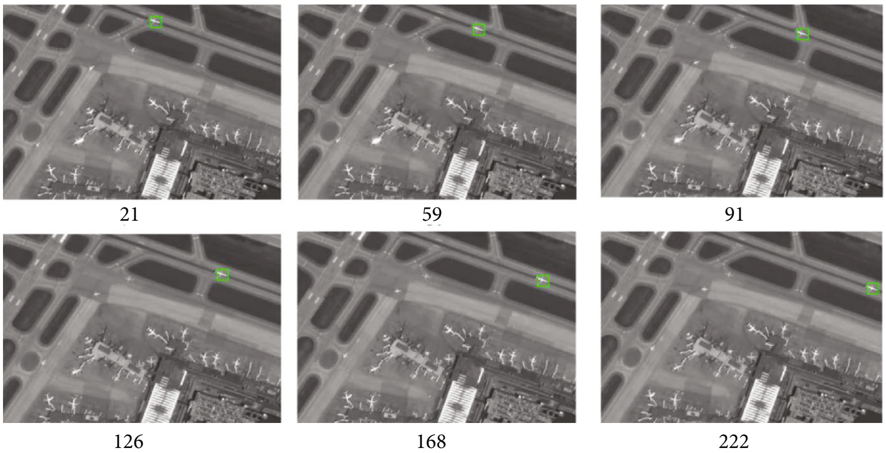


Figure 5: The results of moving airplane-1 tracking by CNK-KCF.

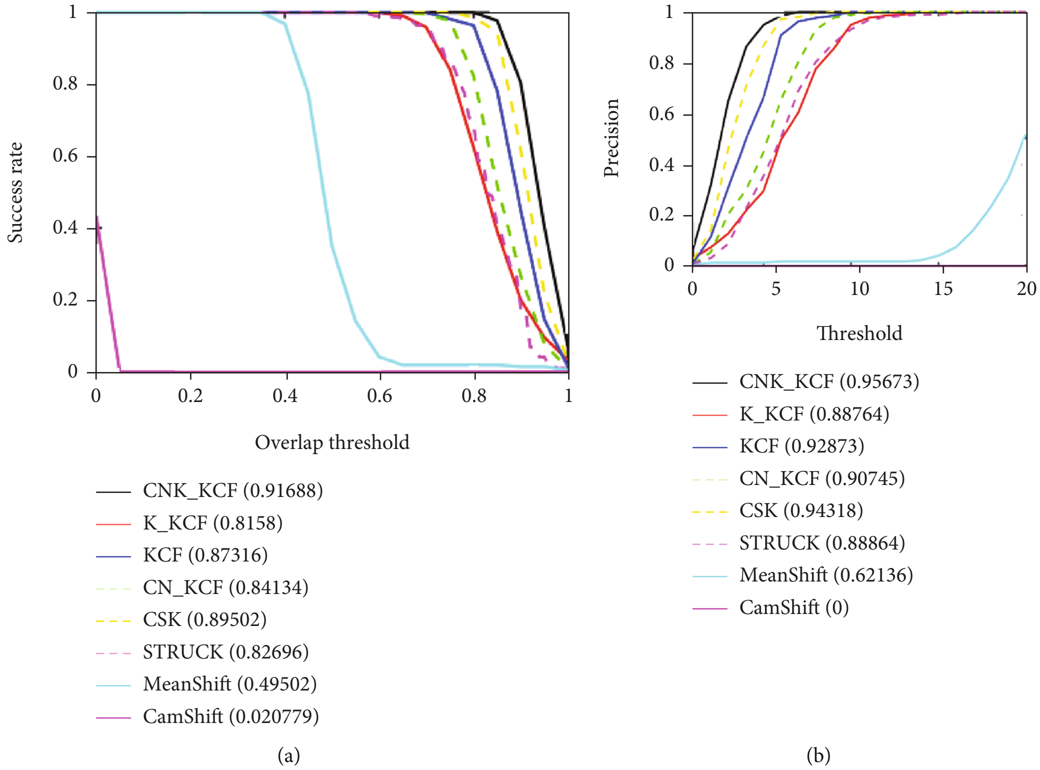


Figure 6: The success and precision plots in comparison with the other seven trackers: (a) the success plot for airplane-1; (b) the precision plot for airplane-1.

dimension-reduction matrix B_k as

$$\tilde{x}_{CN}^{k} = \min_{B_k} \left(\left(\alpha_k \eta_{\mathrm{DATA}} + \sum_{i=1}^{k-1} \alpha_k \eta_{\mathrm{SMOOTH}}^{i} \right) \right), \tag{30}$$

where α_k and α_i are weight coefficients and η_{DATA} can be used for the solution of the dimension reduction matrix B_k in Equation (29). Meanwhile, due to the poor discriminant performance in this process, $\eta_{\mathrm{SMOOTH}}^{i}$ is introduced to increase the robustness of B_k in (30), where the first ith frames are added for training. The forms of η_{DATA} and $\eta_{\mathrm{SMOOTH}}^{i}$ are

$$\eta_{\mathrm{DATA}} = \frac{1}{MN} \sum_{m,n} \left\| x_{CN}^{k}(m,n) - B_k B_k^{T} \hat{x}_k(m,n) \right\|^2, \tag{31a}$$

$$\eta_{\mathrm{SMOOTH}}^{i} = \sum_{p=1}^{s} \lambda_i^{p} \left\| b_i^{p} - B_k B_k^{T} b_i^{p} \right\|^2. \tag{31b}$$

In Equations (31a) and (31b), x_{CN}^{k} represents the CN

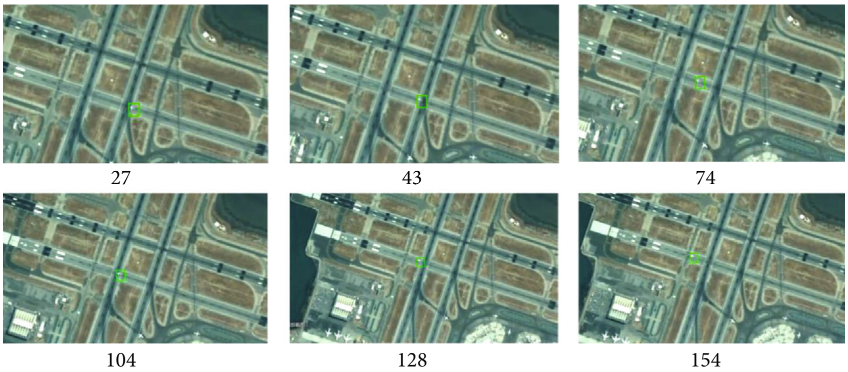
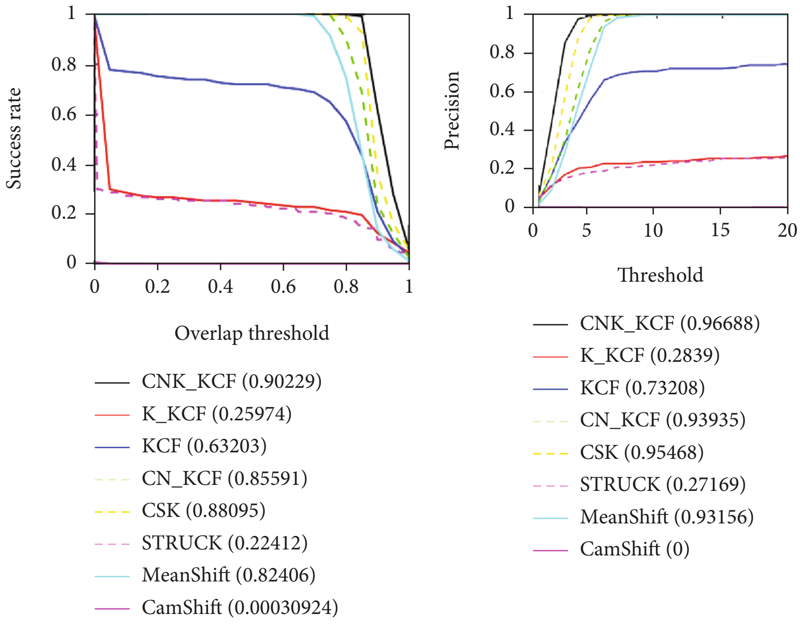


Figure 7: The results of moving airplane-2 tracking by CNK-KCF.



(a)

(b)

Figure 8: The success and precision plots in comparison with the other seven trackers: (a) the success plot for airplane-2; (b) the precision plot for airplane-2.

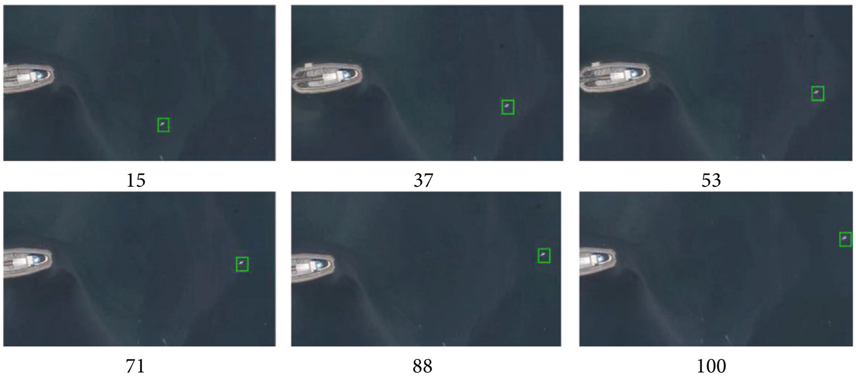
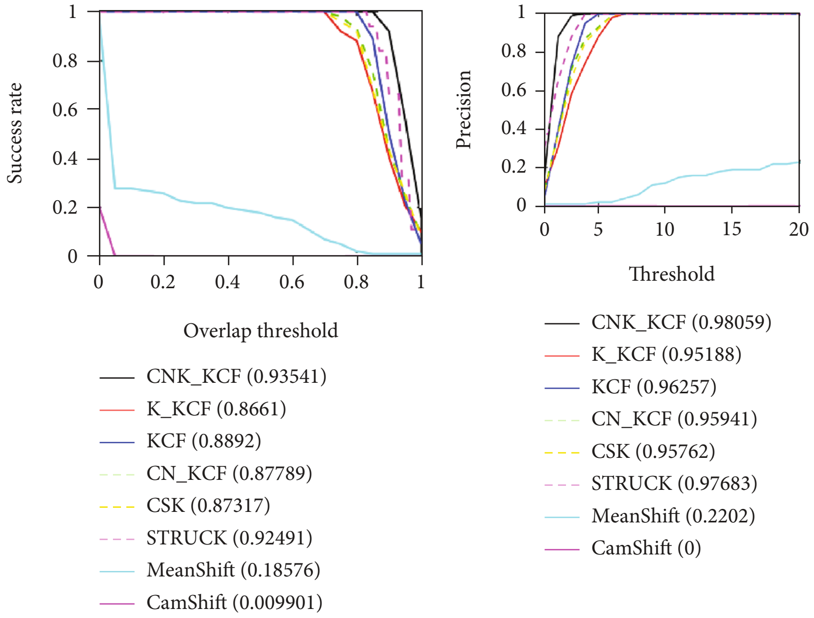


Figure 9: The results of moving airplane-3 tracking by CNK-KCF.



(a) (b)

Figure 10: The success and precision plots in comparison with the other seven trackers: (a) the success plot for airplane-3; (b) the precision plot for airplane-3.

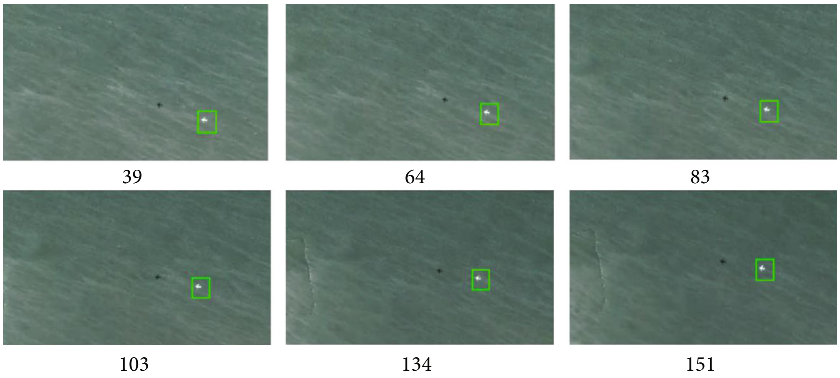


Figure 11: The results of moving airplane-4 tracking by CNK-KCF.

features in the kth frame, s is the total number of basis vectors, and λ_i^p is a nonnegative weight. Accordingly, based on Equations (31a) and (31b), Equation (30) can be rewritten as

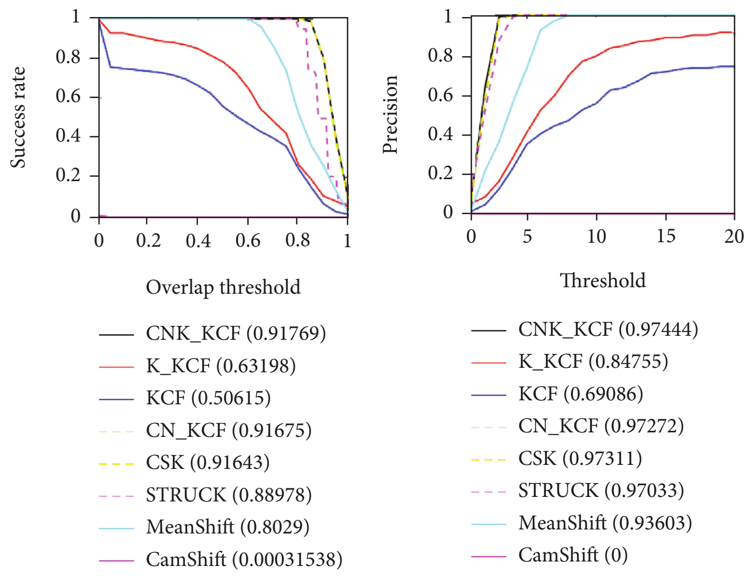
$$\tilde{x}_{CN}^{k} = \min_{B_k} \left(\left(\alpha_k \frac{1}{MN} \sum_{m,n} \left\| \tilde{x}_{CN}^{k}(m,n) - B_k B_k^T \hat{x}_k(m,n) \right\|^2 + \sum_{i=1}^{k-1} \alpha_k \sum_{p=1}^{s} \lambda_i^p \left\| b_i^p - B_k B_k^T b_i^p \right\|^2 \right) \right). \quad (32)$$

Subsequently, based on (30), B_k obtained from (32) and the values $\hat{x}_{CN}^{k}$ and $\tilde{x}_{CN}^{k}$ can be updated as

$$\hat{x}_{CN}^{k} = (1-\gamma)\hat{x}_{CN}^{k-1} + \gamma \hat{x}_{CN}^{k}, \quad (33a)$$

$$\tilde{x}_{CN}^{k} = B_k \hat{x}_{CN}^{k} = B_k \left[(1-\gamma)\hat{x}_{CN}^{k-1} + \gamma \hat{x}_{CN}^{k} \right], \quad (33b)$$

where γ is a learning rate parameter.



(a) (b)

Figure 12: The success and precision plots in comparison with the other seven trackers: (a) the success plot for airplane-4; (b) the precision plot for airplane-4.

2.3. Motion Estimation by Kalman Filter. The Kalman filter is initialized before tracking, and the initial state vector containing the manually labeled real coordinate value of the target center and the velocity component on the coordinate axis is obtained. The state equation and observation equation of the Kalman filtering algorithm are, respectively, shown as follows:

$$U_t = A_{t,t-1} U_{t-1} + W_{t-1}, \tag{34a}$$

$$Z_t = H_t U_t + V_t, \tag{34b}$$

where U_t is the state vector of the system at time t. Z_t is the observation vector at time t. $A_{t,t-1}$ is a state transition matrix defined as

$$A_{t,t-1} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{35}$$

and H_t is an observation matrix. Let the initial state $U_0 = [x_0, y_0, x_0', y_0']$, where (x_0, y_0) is the initial coordinate of the target center point, and (x_0', y_0') is the velocity component at x-axis and y-axis. Both process noise W_{t-1} and observation noise V_t are white noise sequences with the mean value of 0, which are uncorrelated.

The current state matrix and covariance are used to predict the speed and position of the target in the next frame according to the recursive estimation principle. Finally, the prediction equation is obtained:

$$U_{t,t-1} = A_{t,t-1} U_{t-1} + W_{t-1}, \tag{36a}$$

$$P_{t,t-1} = A_{t,t-1} P_{t-1} A_{t,t-1}^T + Q_t, \tag{36b}$$

where $U_{t,t-1}$ is the state prediction vector, $P_{t,t-1}$ is the covariance matrix, Q_t is the covariance matrix of the process noise W_t, and Δt is time interval that is usually taken as 1.

After the predicted coordinates are obtained, the target sampling area is expanded to make the sampling area 3.5 times that of the target image, and the CN features of the model are extracted. Samples are constructed using a cyclic matrix. At the same time, the Fourier transform is performed. The Gaussian kernel is also obtained based on the properties of the cyclic matrix and (18).

Combined with the new actual observations and the prior estimates obtained in the previous step, an a posteriori estimate is obtained using the feedback method. The correction equation is given as

$$P_t = (I - K_t H_t) P_{t,t-1}, \tag{37a}$$

$$K_t = P_{t,t-1} H_t^T (H_t P_{t,t-1}) H_t^T + R_t, \tag{37b}$$

$$U_t = U_{t,t-1} + K_t (Z_t - H_t U_{t,t-1}), \tag{37c}$$

$$H_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{37d}$$

where R_t is the covariance matrix of the noise vector V_t. K_t is the Kalman gain matrix.

In theory, the trajectory of a moving target can be regarded as a smooth curve in a short time. However, on the one hand, there is a certain jitter in the target trajectory curve obtained by KCF. Besides, when there are shadows, similar backgrounds, or other interference, it can easily lead to the failure of MOT. The above two problems can be improved by using KF to correct the tracking results. On

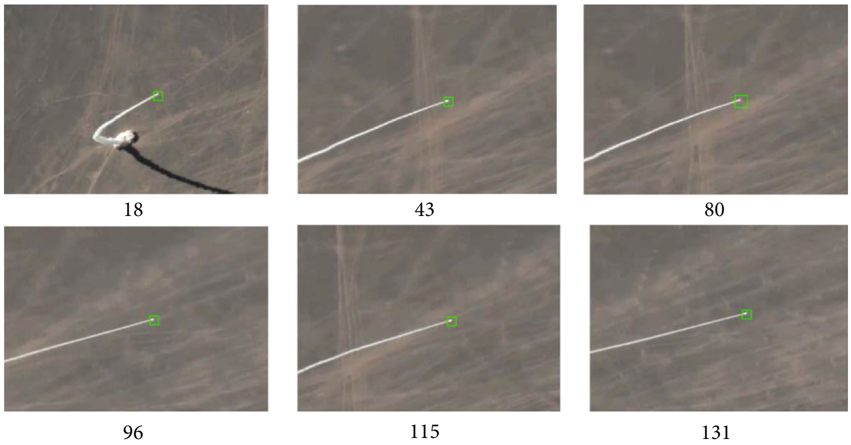
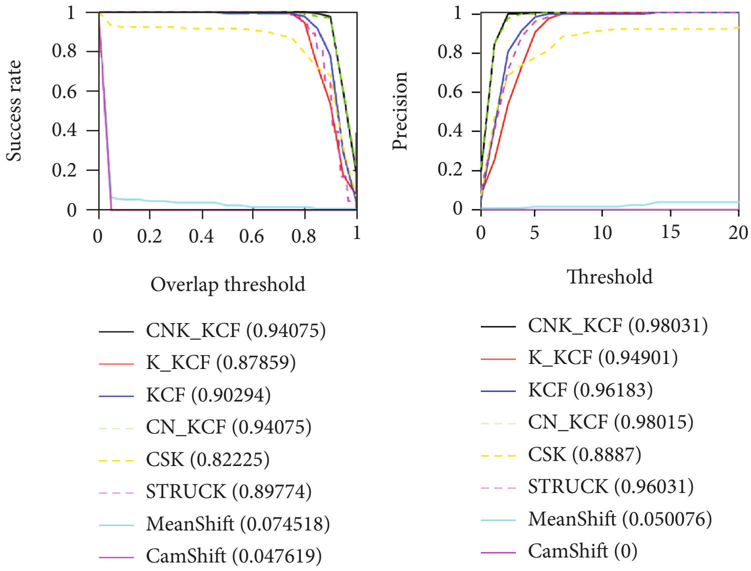


FIGURE 13: The results of moving rocket tracking by CNK-KCF.



(a)

(b)

FIGURE 14: The success and precision plots in comparison with the other seven trackers: (a) the success plot for the rocket; (b) the precision plot for the rocket.

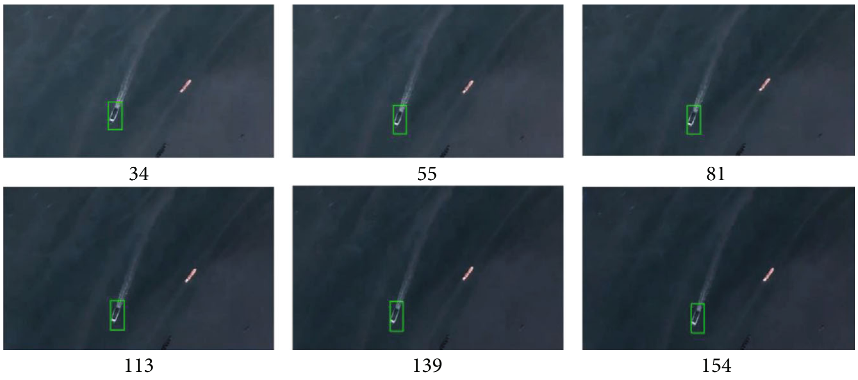
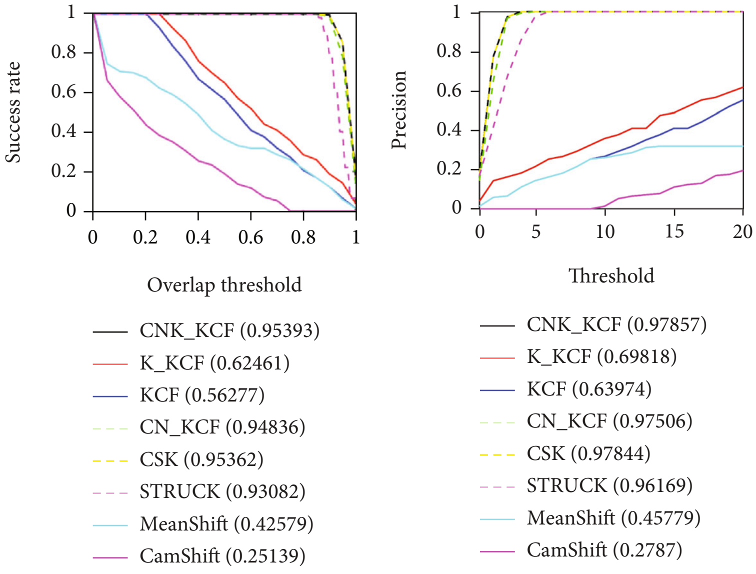


FIGURE 15: The results of moving ship tracking by CNK-KCF.



(a)

(b)

Figure 16: The success and precision plots in comparison with the other seven trackers: (a) the success plot for the ship; (b) the precision plot for the ship.

the other hand, the reason is that KF can predict the possible position in the next frame based on the current frame. Hence, the search area can be relatively reduced. Correspondingly, the tracking accuracy can be relatively improved and is robust to some environmental factors such as partial occlusion, background similarity, and rapid motion.

3. Results and Discussion

3.1. Datasets and Compared Algorithms. In this paper, the experimental datasets are from the Jilin No. 1 satellite constellation developed by the China Changchun Satellite Technology Co., Ltd. There are six videos used in our experiments. Among the videos, six scenes are selected from the satellite videos. Besides, the moving objects are the airplanes in the first four videos, and in the other two videos, the moving objects are the rocket and the ship. The number of the objects is only one in each video. The videos show the part of the process of aircraft flying, ship driving, and rocket launching. Each object is manually labeled through a directional bounding box to describe the position. Figure 4 illustrates the overview of the moving objects in the datasets. In addition, the sizes of the targets in the satellite videos are shown in Table 1. Notice that the sizes of moving objects are small, which may result in tracking failure.

Besides, there are many problems for MOT in satellite videos, which will affect the tracking performance. As a result, to analyze the MOT performance of the proposed method, the authors choose seven trackers, KCF, the kernelized correlation filter-based Kalman filter (K-KCF), CN-KCF, the circulant structure of tracking-by-detection with kernels (CSK), STRUCK, MeanShift, and CamShift, to compare with the proposed CNK-KCF algorithm. In Tables 2–7, the comparison among the above 8 trackers is made in detail.

3.2. Details on the Setting of Parameters. The KCF, K-KCF, CN-KCF, CSK, STRUCK, MeanShift, CamShift, and CNK-KCF are implemented in MATLAB R2018b and NVIDIA GeForce GXT 2080Ti GPU. Therefore, for all tested video datasets, the basic parameters of the algorithms are shown in Table 8, where σ is the standard deviation of the Gaussian kernel, λ is the regularization coefficient, and η is the learning factor. Besides, the parameters of each algorithm are consistent in all video sequences.

3.3. Evaluation Metrics. In this paper, two common evaluation criteria are used, that is, precision plot and success plot [60]. The horizontal axis of the accuracy chart is the center location error (CLE). In a frame image, CLE is described as

$$\mathrm{CLE} = \sqrt{\left(x_{tr} - x_{gr}\right)^2 + \left(y_{tr} - y_{gr}\right)^2}, \tag{38}$$

where CLE represents the average Euclidean distance, (x_{tr}, y_{tr}) is the tracked target center position coordinates, and (x_{gr}, y_{gr}) is the manually marked real coordinates. When CLE is less than this prescribed threshold, it indicates that the tracking target is correct. That means the smaller the prescribed CLE value is, the more accurate the MOT is. The vertical axis of the accuracy map is the percentage of frames in which tracking accuracy is greater than the threshold.

On the other hand, the horizontal axis of the success rate graph is the overlap threshold of the bounding box. The mathematical expression of overlap rate S is as follows:

$$S = \frac{R_t \cap R_a}{R_t \cup R_a}, \tag{39}$$

where R_t is the predicted tracking box obtained from the algorithm, R_a is the real target box marked manually, and S is the ratio of the overlapping area of R_t and R_a to the total area of R_t and R_a. $|\bullet|$ represents the number of pixels in the region. The vertical axis of the success rate graph is the proportion of successful frames to all image frames. In this experiment, the area under the success rate curve (AUC) is used as the performance evaluation criterion of the algorithm. The larger the value is, the better the tracking performance is.

3.4. Experimental Analysis on Moving Airplane Tracking. The moving objects are the airplanes in the first four real satellite videos of this experiment, recorded as airplane-1, airplane-2, airplane-3, and airplane-4. Note that the objects are small, but the scales are large. First, the results of the moving airplane-1 tracking by CNK-KCF are shown in Figure 5. Note that the airplanes at the airport are not completely occluded. It is shown that for the airplane-1 whose shape is clear, CNK-KCF performs best with a precision of 0.957 and a success rate of 0.917. However, the other seven trackers do not possess similar performance as the CNK-KCF in the aspects of success rate and precision (in fact, in Tables 2–7, we can see the CamShift tracking failure in the satellite videos). Figure 6 shows the success rate and precision plots containing the eight trackers for airplane-1.

For the airplane-2 video, Figure 7 shows the results of the moving airplane-2 tracking. Because the environment around the moving target is relatively complex, the K-KCF and STRUCK have a weaker performance than the proposed CNK-KCF, which have a precision of 0.284 and 0.260 and success rate of 0.272 and 0.224, respectively, followed by the KCF. However, the proposed CNK-KCF has better performance with a precision of 0.902 and a success rate of 0.967. Figure 8 shows the success rate and precision plots containing the eight trackers for airplane-2.

For the airplane-3 video, Figure 9 shows the results of the moving airplane-3 tracking. In Table 4, the KCF, CN-KCF, K-KCF, CSK, STRUCK, and CNK-KCF have similar precision performance. However, the success rate of CN-KCF, KCF, K-KCF, and CSK is less than 0.9, compared with that of the proposed CNK-KCF. Figure 10 shows the success rate and precision plots containing the eight trackers for airplane-3.

For the airplane-4 video, Figure 11 shows the results of the moving airplane-4 tracking. In Table 5, the proposed CNK-KCF has better results than the other 7 trackers. The CN-KCF and CSK have slightly weaker results, followed by STRUCK. However, the KCF and K-KCF have relatively weaker results, only with a precision of 0.691 and 0.848 and a success rate of 0.506 and 0.632, respectively. Figure 10 shows the success rate and precision plots containing the eight trackers for airplane-4. In all, in the experimental results, on the MOT of airplanes, it can be seen that the designed CNK-KCF has better tracking success rate and precision, compared with the other seven algorithms. Figure 12 shows the success rate and precision plots containing the eight trackers for airplane-4.

3.5. Experimental Analysis on Moving Rocket Tracking. The results of moving rocket tracking are shown in Figure 13. Because a rocket over the sea will be partly occluded in the satellite videos, it is necessary to consider the occlusion detection in the MOT. Besides, the rocket is small in satellite videos. Hence, the texture features and shape are not clear. In Table 6, CNK-KCF performs best with a precision of 0.980 and a success rate of 0.941, as well as CN-KCF. KCF, K-KCF, STRUCT, and CSK showing weaker performance in the aspect of precision and success rate. Besides, MeanShift fails to track the rocket. Figure 14 shows the success rate and precision plots containing the 8 trackers for the rocket.

3.6. Experimental Analysis on Moving Ship Tracking. The results of moving ship tracking are shown in Figure 15. CNK-KCF performs best with a precision of 0.979 and a success rate of 0.954. KCF and K-KCF show weaker performance in the aspect of precision and success rate. The performance of MeanShift and CamShift is also weaker with a precision of 0.458 and 0.279 and a success rate of 0.426 and 0.251, respectively. Figure 16 shows the success rate and precision plots containing the 8 trackers for the ship.

In all, compared with the other 7 algorithms, the proposed CNK-KCF possesses better tracking performance.

4. Conclusions

In this paper, the authors propose an effective tracker called CNK-KCF based on the framework of the correlation filter for the MOT in satellite videos. Based on the CN feature, the proposed tracker can process the videos in the multichannel color feature. The Kalman filter is also utilized to improve the tracking success rate and accuracy.

The improved algorithm is tested on Jilin-1 datasets many times. Meanwhile, we compared with other seven tracking algorithms. The experimental results are analyzed and it has been determined that the proposed method is robust in several complex situations such as rapid target motion, occlusion, and similar object interference. The proposed algorithm solves the problem of tracking failure when a moving object is partially occluded. In the future work, the above tracker will be combined with the controller rather than working separately.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62133001 and 61520106010) and the National Basic Research Program of China (973 Program) (2012CB821200 and 2012CB821201).

References

[1] P. d'Angelo, G. Máttyus, and P. Reinartz, "Skybox image and video product evaluation," *International Journal of Image and Data Fusion*, vol. 7, no. 1, pp. 3–18, 2016.

[2] A. Hawbani, X. Wang, S. Karmoshi et al., "GLT: grouping based location tracking for object tracking sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, 19 pages, 2017.

[3] J. Zhang, X. Jia, and J. Hu, "Local region proposing for frame-based vehicle detection in satellite videos," *Remote Sensing*, vol. 11, no. 20, p. 2372, 2019.

[4] Z. Tang, "Intelligent target detection and tracking algorithm for martial arts applications," *Wireless Communications and Mobile Computing*, vol. 2022, 10 pages, 2022.

[5] X. Li, L. Zhang, and J. You, "Domain transfer learning for hyperspectral image super-resolution," *Remote Sensing*, vol. 11, no. 6, p. 694, 2019.

[6] F. Nunziata, X. Li, A. Marino et al., "Microwave satellite measurements for coastal area and extreme weather monitoring," *Remote Sensing*, vol. 13, no. 16, p. 3126, 2021.

[7] A. Xiao, Z. Wang, L. Wang, and Y. Ren, "Super-resolution for "Jilin-1" satellite video imagery via a convolutional network," *Sensors*, vol. 18, no. 4, p. 1194, 2018.

[8] G. Kopsiaftis and K. Karantzalos, "Vehicle detection and traffic density monitoring from very high resolution satellite video data," in *Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1881–1884, Milan, Italy, July 2015.

[9] Y. Guo, D. Yang, and Z. Chen, "Object tracking on satellite videos: a correlation filter-based tracking method with trajectory correction by Kalman filter," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3538–3551, 2019.

[10] T. Yang, X. Wang, B. Yao et al., "Small moving vehicle detection in a satellite video of an urban area," *Sensors*, vol. 16, no. 9, p. 1528, 2016.

[11] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4511–4523, 2014.

[12] W. Ao, Y. Fu, and X. Hou, "Needles in a haystack: tracking city-scale moving vehicles from continuously moving satellite," *IEEE Transactions on Image Processing*, vol. 29, pp. 1944–1957, 2020.

[13] F. Kocadag and A. Demirkol, "Real time tracking of TV satellites on moving vehicles using Kalman filter," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 1–5, Noida, India, February 2015.

[14] W. L. Wang, Q. Q. Li, and L. L. Tang, "Algorithm of vehicle detection in low altitude aerial video," *Journal of Wuhan University of Technology*, vol. 32, pp. 155–158, 2010.

[15] Y. Luo, Y. Liang, and Y. Wang, "Traffic flow parameter estimation from satellite video data based on optical flow," *Computer Engineering and Applications*, vol. 54, pp. 204–207, 2018.

[16] L. Gueguen and R. Hamid, "Large-scale damage detection using satellite imagery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1321–1328, Boston, MA, USA, June 2015.

[17] L. H. Hu, "Evaluation research on the application of GF-1 satellite for monitoring major engineering land," *Journal of North China Institute of Science and Technology*, vol. 12, pp. 110–115, 2015.

[18] A. Hawkins, J. Carrico, S. Motiwala, and C. Maclachlan, "Flight dynamics operations and collision avoidance for the Skysat imaging constellation," in *9th International Workshop on Satellite Constellations and Formation Flying (IWSCFF)*, University of Colorado, Boulder, CO, USA, July 2017.

[19] A. K. Murthy, M. Shearn, and B. D. Smiley, "Sky Sat-1: very high-resolution imagery from a small satellite," in *Proceedings of the Sensors, Systems, and Next-Generation Satellites XVIII*, Amsterdam, Netherlands, September 2014.

[20] M. Steckling, U. Renner, and H. P. Röser, "DLR-TUBSAT, qualification of high precision attitude control in orbit," *Acta Astronautica*, vol. 39, no. 9-12, pp. 951–960, 1996.

[21] N. Chen, *Jilin-1: China's first commercial remote sensing satellites aim to fill the void*, Chinese Academy of Sciences, 2016.

[22] X. Zhang, J. Xiang, and Y. Zhang, "Space object detection in video satellite images using motion information," *International Journal of Aerospace Engineering*, vol. 2017, 9 pages, 2017.

[23] M. Yoo, Y. Na, H. Song et al., "Motion estimation and hand gesture recognition-based human–UAV interaction approach in real time," *Sensors*, vol. 22, no. 7, p. 2513, 2022.

[24] A. Yilmaz, O. Javed, and M. Shah, "Object tracking," *ACM Computing Surveys*, vol. 38, no. 4, pp. 13–45, 2006.

[25] X. Mei and H. Ling, "Robust visual tracking using L1 minimization," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 1436–1443, Kyoto, Japan, 2009.

[26] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust L1 tracker using accelerated proximal gradient approach," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 1830–1837, Providence, RI, USA, July 2012.

[27] X. Mei, H. Ling, Y. Wu, E. P. Blasch, and L. Bai, "Minimum error bounded efficient l1 tracker with occlusion detection," in *2011 IEEE conference on computer vision and pattern recognition*, pp. 1257–1264, Colorado Springs, CO, USA, June 2011.

[28] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.

[29] Y. Bai and M. Tang, "Robust tracking via weakly supervised ranking SVM," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 1854–1861, Providence, RI, USA, June 2012.

[30] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *10th European Conference on Computer Vision (ECCV 2008)*, pp. 234–247, Marseille, France, October 2008.

[31] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M. H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *13th European Conference on Computer Vision (ECCV 2014)*, pp. 2661–2675, Zurich, Switzerland, September 2014.

[32] Q. He, R. Yang, J. Lau, and M. H. Yang, "Visual tracking via locality sensitive histograms," in *2013 IEEE conference on computer vision and pattern recognition*, pp. 2427–2434, Portland, OR, USA, June 2013.

[33] X. Jia, H. Lu, and M. H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 1822–1829, Providence, RI, USA, June 2012.

[34] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," *Pattern Recognition Letters*, vol. 49, pp. 250–258, 2014.

[35] Q. Wang, J. Fang, and Y. Yuan, “Multi-cue based tracking,” *Neurocomputing*, vol. 131, pp. 227–236, 2014.

[36] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, “Robust visual tracking via multi-task sparse learning,” in *2012 IEEE conference on computer vision and pattern recognition*, pp. 2042–2049, Providence, RI, USA, June 2012.

[37] Y. Tian, T. Guan, and C. Wang, “Real-time occlusion handling in augmented reality based on an object tracking approach,” *Sensors*, vol. 10, no. 4, pp. 2885–2900, 2010.

[38] H. Shi, Z. Lin, W. Tang, B. Liao, J. Wang, and L. Zheng, “A robust hand tracking approach based on modified tracking-learning-detection algorithm,” in *8th International Conference on Multimedia and Ubiquitous Engineering (MUE 2014)*, pp. 9–15, Zhangjiajie, China, May 2014.

[39] C. Jia, Z. Wang, X. Wu et al., “Tracking-learning-detection (TLD) method with local binary pattern improved,” in *2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1625–1630, Zhuhai, China, December 2015.

[40] S. Avidan, “Support vector tracking,” in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition (CVPR 2001)*, pp. 1064–1072, Kauai, HI, USA, December 2001.

[41] H. Grabner, M. Grabner, and H. Bischof, “Real-time tracking via online boosting,” in *2006 17th British Machine Vision Conference (BMVC 2006)*, pp. 47–56, Edinburgh, United Kingdom, September 2006.

[42] B. Babenko, M. H. Yang, and S. Belongie, “Robust object tracking with online multiple instance learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, 2011.

[43] M. Danelljan, A. Robinson, and F. Shahbaz Khan, “Beyond correlation filters: learning continuous convolution operators for visual tracking,” in *21st ACM Conference on Computer and Communications Security (CCS 2014)*, pp. 472–488, Scottsdale, AZ, United states, November 2014.

[44] R. S. Chavan and M. Patil, “Object tracking based on tracking-learning-detection,” *IEEE Transactions on Software Engineering*, vol. 7, 2011.

[45] S. Hare, A. Saffari, and P. Torr, “Struck: structured output tracking with kernels,” in *2011 IEEE International Conference on Computer Vision (ICCV 2011)*, pp. 263–270, Barcelona, Spain, November 2011.

[46] W. Zhong, H. Lu, and H. Yang, “Robust object tracking via sparse collaborative appearance model,” *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2356–2368, 2014.

[47] J. Zhang, Y. Liu, H. Liu, and J. Wang, “Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection,” *Sensors*, vol. 21, no. 4, p. 1129, 2021.

[48] J. Yang, W. Tang, and Z. Ding, “Long-term target tracking of UAVs based on kernelized correlation filter,” *Mathematics*, vol. 9, no. 23, p. 3006, 2021.

[49] M. Danelljan, G. Häger, and F. S. Khan, “Discriminative scale space tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1561–1575, 2017.

[50] H. K. Galoogahi, A. Fagg, and S. Lucey, “Learning background-aware correlation filters for visual tracking,” in *16th IEEE International Conference on Computer Vision (ICCV 2017)*, pp. 1144–1152, Venice, Italy, October 2017.

[51] B. Du, Y. Sun, S. Cai, C. Wu, and Q. Du, “Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 168–172, 2018.

[52] H. Li and Y. Man, “Moving ship detection based on visual saliency for video satellite,” in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2016)*, pp. 1248–1250, Beijing, China, July 2016.

[53] L. Lei and D. Guo, “Multitarget detection and tracking method in remote sensing satellite video,” *Computational Intelligence and Neuroscience*, vol. 2021, 7 pages, 2021.

[54] Y. Liu, Y. Liao, C. Lin, Y. Jia, Z. Li, and X. Yang, “Object tracking in satellite videos based on correlation filter with multi-feature fusion and motion trajectory compensation,” *Remote Sensing*, vol. 14, no. 3, p. 777, 2022.

[55] L. Meng and J. P. Kerekes, “Object tracking using high resolution satellite imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 1, pp. 146–152, 2012.

[56] F. S. Khan, J. Van De Weijer, and M. Vanrell, “Modulating shape features by color attention for object recognition,” *International Journal of Computer Vision*, vol. 98, no. 1, pp. 49–64, 2012.

[57] M. Danelljan, G. Bhat, and F. S. Khan, “ECO: efficient convolution operators for tracking,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, pp. 6931–6939, Honolulu, HI, USA, July 2017.

[58] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[59] R. Rifkin, G. Yeo, and T. Poggio, “Regularized least-squares classification,” *Nato Science Series Sub Series III Computer and Systems Sciences*, vol. 190, pp. 131–154, 2003.

[60] M. Kristan, J. Matas, A. Leonardis et al., “The seventh visual object tracking VOT2019 challenge results,” in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 2206–2241, Seoul, Korea (South), October 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3543640, 11 pages
https://doi.org/10.1155/2022/3543640

Research Article

Design of Resource-Aware Load Allocation for Heterogeneous Fog Computing Environments

Syed Rizwan Hassan,[1] **Ishtiaq Ahmad**,[1] **Ateeq Ur Rehman**,[2,3] **Seada Hussen**,[4] **and Habib Hamam**[3,5,6,7]

[1]*Department of Electrical Engineering, The University of Lahore, Lahore 54000, Pakistan*
[2]*Department of Electrical Engineering, Government College University, Lahore 54000, Pakistan*
[3]*Faculty of Engineering, Uni de Moncton, Moncton, NB E1A 3E9, Canada*
[4]*School of Electrical and Computer Engineering, Haramaya Institute of Technology, 138 Dire Dawa, Ethiopia*
[5]*International Institute of Technology and Management, Libreville BP1989, Gabon*
[6]*Spectrum of Knowledge Production & Skills Development, Sfax 3027, Tunisia*
[7]*Department of Electrical and Electronic Engineering Science, School of Electrical Engineering, University of Johannesburg, Johannesburg 2006, South Africa*

Correspondence should be addressed to Syed Rizwan Hassan; syedrizwanhassan@gmail.com,
Ateeq Ur Rehman; ateeq.rehman@gcu.edu.pk, and Seada Hussen; seada.hussen@aastu.edu.et

Received 2 February 2022; Revised 24 April 2022; Accepted 28 April 2022; Published 7 June 2022

Academic Editor: Maode Ma

The architecture employed by most of the researchers for the deployment of latency-sensitive Internet of Things (IoT) applications is fog computing. Fog computing architecture offers less delay as compared to the cloud computing paradigm by providing resource constraint fog devices close to the edge of the network. Fog nodes process the incoming data by utilizing available resources which reduces the volume of data to be sent to the cloud server. Fog devices having dissimilar processing capabilities are present in a system. The connection of suitable sensor nodes to the parent fog node plays an essential role in achieving the optimum performance of the system. In this paper, we have designed an algorithm that dynamically assigns appropriate sensor devices to fog nodes to achieve a reduction in network utilization and latency. The proposed algorithm estimates the volume of information detected by an edge device from the rate of sensing frequency of the sensor attached to the edge device. The proposed policy while connecting the network nodes takes into account the heterogeneity and processing capability of the devices. Several evaluations are performed on multiple scales for the evaluation of the proposed algorithm. The outcomes of the evaluations confirm the effectiveness of the proposed algorithm in achieving a reduction in network consumption and end-to-end delay.

1. Introduction

Devices under the everyday use of human beings are connected to the Internet owing to the Internet of Things (IoT) technology. Management of the big data generated by these devices needs high processing and storage units. The growing demand of IoT applications results in a huge increase in interconnection and deployment of IoT devices. The number of IoT devices to be present in the system by 2025 is expected to be 1 trillion [1], whose financial impression will be nearly equivalent to 11% of the world economy.

A large number of applications of diverse kinds of nature are implemented using cloud and fog computing architectures. The applications implemented on the cloud paradigm consist of sensor nodes that contain sensors that are used to detect the environment. The effectiveness of the results deduced from the sensed data depends on the frequency and quality of the information sensed by the sensor nodes. In conventional cloud-based architecture, the sensors are directly linked to the cloud server. The data sensed by the sensor nodes are transmitted directly to the cloud for further processing [2]. Cloud computing provides a centralized approach for the implementation of IoT applications. In cloud architecture, a resourceful cloud server is located in a centralized way where all the data generated by the end devices is to be diffused for processing. Several IoT applications are deployed using cloud computing architecture. However, cloud computing architecture is unable to meet the stringent requirements of delay-sensitive applications. Cloud paradigm being cloud-centric architecture offers high network load and latency while deploying applications on a large scale [3].

On the other hand, fog computing architecture provides resources in a distributed manner. Several benefits are there of deploying applications on fog paradigm including low latency and reduced network consumption and execution cost. Figure 1 depicts the general fog computing architecture which is a three-layer model consisting of sensor layer, fog layer, and cloud layer. The sensor layer consists of edge devices responsible for detecting the change in the environment [4]. The fog layer consists of fog devices having limited power, storage, and computational capabilities [5]. The major contribution of fog architecture is the provision of facilities near the edge devices to reduce network load and delay [6]. Deployment of IoT applications on a large scale is one of the key advantages of fog computing architecture. Fog architecture is extensively used in designing and implementing IoT applications. Several researchers have deployed the fog computing paradigm for the deployment of IoT applications. In [7], the authors have presented an energy-efficient multitier fog computing approach that introduces additional layers in the conventional fog architecture for the provision of IoT-based smart services.

The data sensed by the sensor nodes is communicated to the fog devices for processing by using the fog resources. The sensing frequency of deployed sensors determines the volume of sensed information by the edge nodes. Sufficient processing resources are required to handle the massive amount of information generated by edge devices having sensors with high sensing rates. To accomplish the processing of such a large amount of data often consumes a large part of the resources available at the fog nodes. An essential requirement to implement efficient fog-based applications is the optimized allocation of end devices to fog nodes in such a way that maintains a balance between the sensed load and available processing resources at the fog devices. In this context, an approach is proposed for fog computing-based applications to assign appropriate edge devices to parent fog nodes in such a way as to achieve a reduction in delay and network utilization.

This paper presents the design of a resource-aware load allocation algorithm for the fog-cloud computing paradigm. The proposed approach manages a balance between resources available at the fog paradigm and the volume of sensed load approaching from the sensor layer to achieve effective utilization of network resources. The proposed approach manages the processing load on fog infrastructure by assigning appropriate edge nodes to the fog devices. To maintain optimum interconnection between the fog and edge nodes, the proposed algorithm exploits the sensing frequency information. A comparison is performed between the proposed approach and the traditional cloud-based paradigm by deploying IoT applications on multiple scales to evaluate the efficiency of the designed scheme. The performance metrics under observation during all these evaluations are network consumption and latency.

The article is divided into different sections. The segment below offers a comprehensive review of different research. Section 3 presents the system proposed. Section 4 describes the values of parameters used in the simulations. Section 5 summarizes the results and comparisons, and the last section discusses the conclusion and future directions of research.

2. Related Work

In this section, numerous cloud and fog computing-based systems designed related to this research work are presented.

The advent of cloud technology provides processing of a large volume of sensed data coming from geographically distributed sensors at a common point facilitating shared access to sensed information from different localities. A generic resource model keeping in view of computational resources and sensing resources is presented in [8] which facilitates request, allocation, and reservation services in sensor cloud environments.

Cloud of sensors (CoS) is a three-layer architecture that provides cloud resources near the edge for wireless sensor networks. To efficiently achieve the application requirements, several virtual nodes are assigned to applications by decoupling them from the CoS paradigm; this technique is called CoS virtualization. To resolve resource provision issues in CoS networks, the authors in [9] proposed a partially distributed algorithm, namely, Zeus, with two key functionalities. Initially, it classifies the common tasks from multiple applications and executes them once for all the applications thus minimizing the consumption of resources. Secondly, the virtual nodes are provided at the edge of the network providing reduced latency and scalability.

Three-layer Cloud of Things (CoT) is an architectural concept defining the connectivity of the IoT devices to the cloud server through edge devices. In heterogeneous and complex CoT environments, the main problem is the allocation of resources in an effective way to meet the requirements of applications. The main goal of resource allocation tasks in such environments is to maximize the number of applications to be run on the architecture using available resources. The authors in [10] presented a strategy to manage heterogeneous fog nodes scattered throughout the network to achieve efficient utilization of network resources in a way to achieve minimum delay for diverse nature of applications.

Fog architecture is providing prominent support in the implementation of delay-sensitive applications, whereas cloud computing paradigm being more resourceful was verified to be a better option in the implementation of IoT applications from a processing perspective. The authors in [11] presented a multilevel deadline-based resource

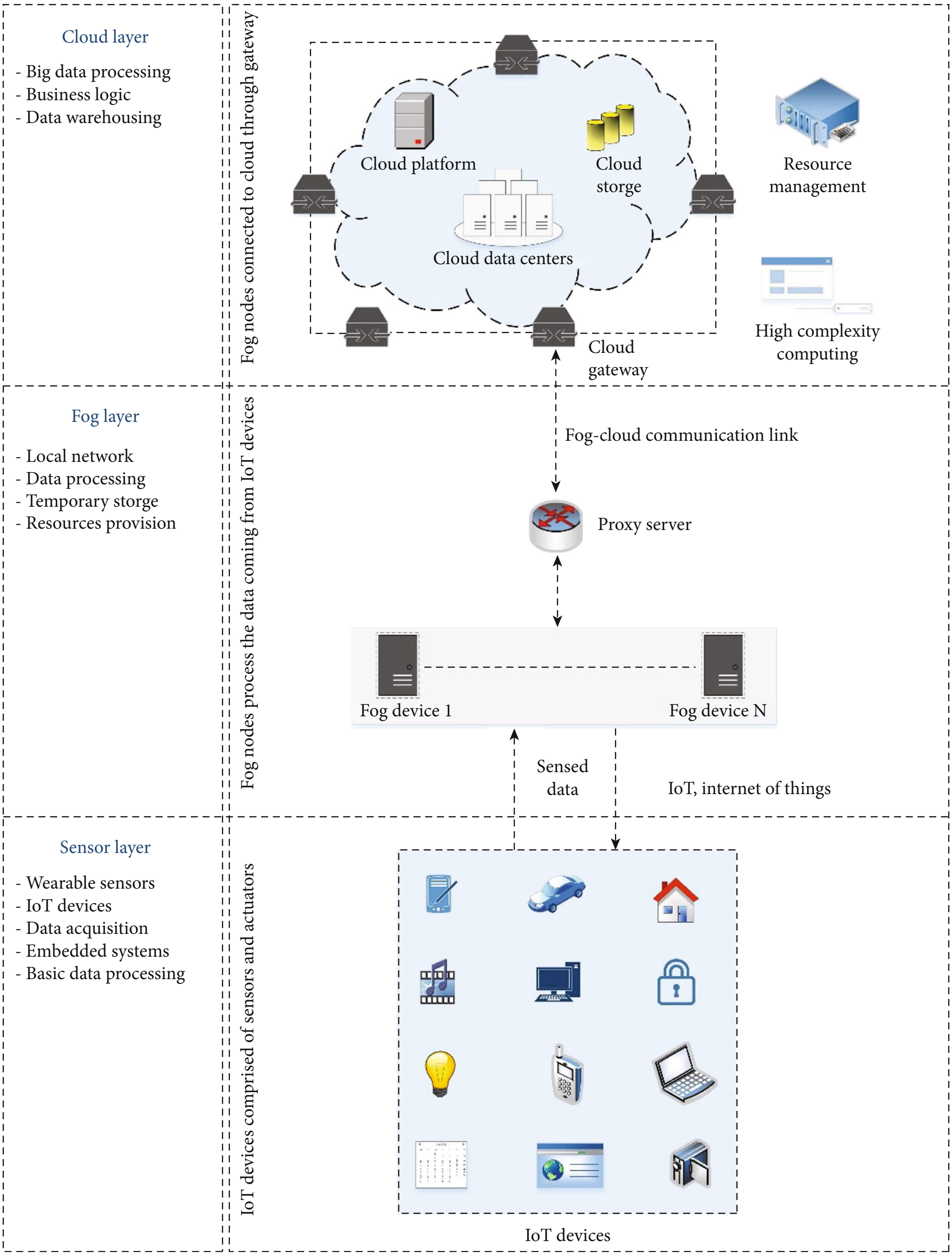


Figure 1: Fog-cloud computing paradigm.

allocation policy to meet dynamic user requirements for fog-cloud networks and simulated the proposed algorithm extending the CloudSim toolkit. A reduction of 12% in processing time and 15% in execution cost is achieved by using the proposed approach. In [12], the authors proposed a resource allocation model for vehicular cloud computing (VCC) networks. They model a multiobjective optimization problem with maximization of acceptance rate and minimization of the cloud provider cost as constraints. The authors improved the nondominated sorting genetic algorithm II (NSGA-II) by amending the initial population according to the matching factor, dynamic crossover probability, and mutation probability to promote excellent individuals and increase population diversity. The results of the evaluations performed expose that the proposed methods achieve improved performance compared to the former models.

Smart cities based on several cyber-physical systems (CPS) have a different level of intelligence [13, 14]. Extensive usage and dependency of applications on CPS can cause a disturbance, damage, and loss on failure of CPS. Cloud on the other hand can provide more sustainable solutions for smart cities. Instantaneous reception of multiple application services on sensor nodes can cause service collisions causing coupling resource management problems causing a denial of services. The authors in [15] proposed an extension of Hungarian algorithm-based fog computing architecture in which the fog layer role is to handle resource provision matters to reduce latency. The result of different experiments performed confirm that the extended algorithm provides a reduction in coupling issues and achieves efficient utilization of available fog resources.

Due to the introduction of the fog computing concept, there is a drastic increase in the deployment of latency-sensitive and large-scale applications on this paradigm. E-healthcare industry is also moving towards providing intelligent medical services to the users near to improve quality of life. Recent work in the provision of medical services using cloud and fog computing paradigm includes remote pain monitoring using cloud paradigm for newly born babies to decrease death ratio in infants [16]. In [17], the authors designed a fall detection application for aged persons using the mobile cloud paradigm and discussed the beneficial aspects of implementation healthcare applications on the said architecture. Healthcare applications have stringent QoS requirements which require low latency and stable connectivity between biopotential sensors and cloud servers. To meet latency-sensitive requirements of such applications, fog computing offers a promising solution by providing resources adjacent to the network border. In [18], the authors have deployed Medical Cyber-Physical Systems (MCPSs) on fog computing paradigm and examined different parameters affecting the cost-effective implementation of fog commuting-based MCPS. They also proposed a heuristic algorithm for optimal and cost-effective implementation of fog-based MCPS.

The authors in [19] presented an energy-efficient cloud-based application for the continuous monitoring of patients in a persistent vegetative state by detecting their facial muscles. For remote monitoring access, the proposed design contains a mobile platform. The fog computing paradigm is employed for the design of a remote pain monitoring application in [20]. The presented design detects and processes the biopotential signals of patients admitted in hospitals by using fog resources and provides access to information related to patients through a web platform. The authors compared the proposed design with the cloud-based design and performed several simulations to confirm the effectiveness of the proposed design. The authors in [21] presented a fog computing-based approach for the design of an efficient car parking system that provides a reduction in latency and network consumption as compared to a cloud-based approach. A task assignment approach that allocates modules according to processing resources available at the network devices present in the system is presented in [22]. A module allocation approach for heterogeneous fog-cloud computing environments is presented in [23]. The presented algorithm efficiently allocates the application modules to the fog nodes while considering connection latency, computational power, and volume of sensed information. A comparison of the proposed approach with the traditional computing architectures is also presented by the authors. A cloud-fog based approach for the provision of on-demand health monitoring services is presented in [24]. In the presented approach, the cloud server acts as a backbone for the provision of resources and reservations for the provision of services near the edge which is resolved using the resource constraint fog nodes. Table 1 presents a qualitative comparison of the proposed approach with the existing schemes.

Table 1: Qualitative comparison of the proposed system with the existing systems.

Reference	Paradigm	Delay	Processing cost	Network load
[19]	Cloud	High	High	High
[20]	Fog	Moderate	Low	Low
[21]	Fog	Moderate	Low	Low
[22]	Fog	Moderate	Moderate	Moderate
[23]	Fog-cloud	Moderate	Low	Low
[24]	Cloud-fog	High	Medium	Low
*Proposed	Fog-cloud	Minimum	Low	Low

Table 2: Comparison of the proposed algorithm with the existing resource assignment schemes.

Reference	Observed parameters during resource assignment		
	Sensing rate	Capacity	Heterogeneous devices
[22]	✗	✓	✗
[25]	✗	✓	✗
[26]	✗	✓	✗
[27]	✗	✓	✗
*Proposed	✓	✓	✓

In [25], the authors proposed a strategy that balances the load generated by the cameras in a face recognition IoT application for video surveillance. Network usage, computational complexity, and accuracy are the performance metrics chosen for the evaluation of the proposed strategy. In [26], a QoS-aware load assignment policy is proposed that assigns latency-sensitive services on devices situated at the verge of the system to decrease cost and delay in the fog-cloud network. Primarily, the idea of a fog colony is presented in [27]. This concept is repeated in various related works. The fog colony is based on several fog cells and behaves as a micro data center. A fog cell is an application module placed on a node to offer services to linked edge nodes. In [22], the authors proposed a strategy that efficiently uses network resources and places modules according to the available capacity of network nodes. This algorithm searches for eligible network nodes to meet the module placement requirements and places modules on eligible fog nodes unless the fog layer is exhausted. Table 2 shows the comparison between the proposed algorithm and previous load assignment strategies in terms of observed parameters. The sensing frequency parameter for the approximation of

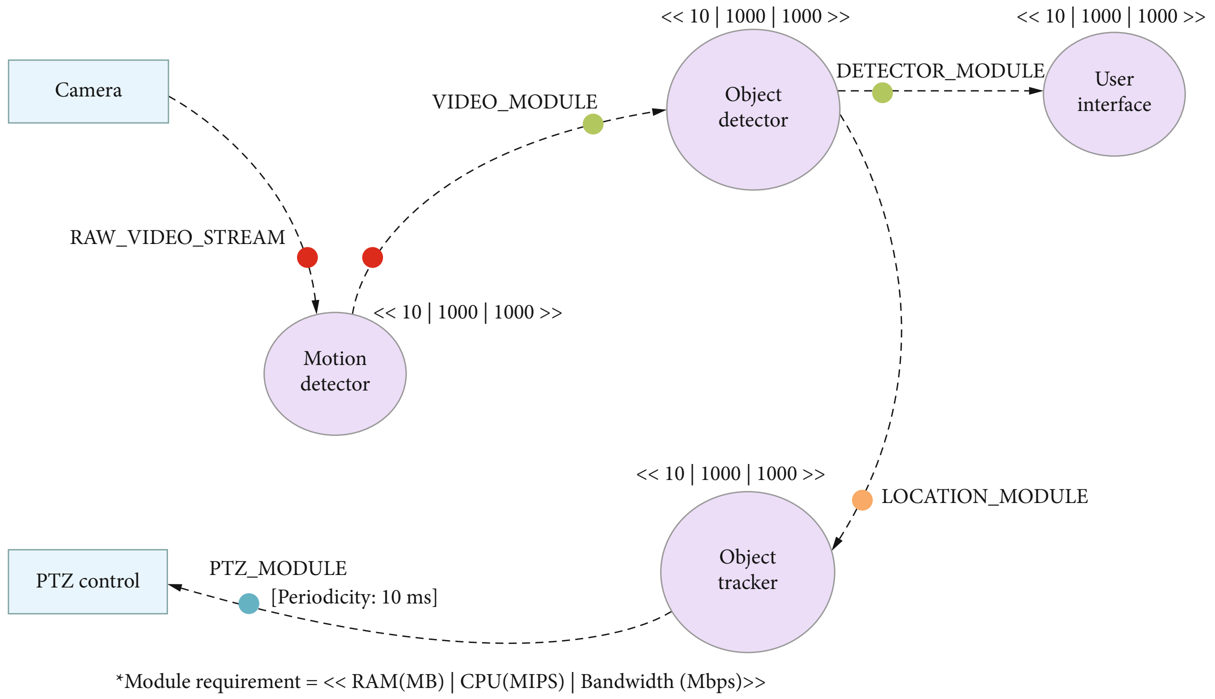


Figure 2: Directed acyclic graph of intelligent surveillance application.

information size is not incorporated in most of the existing models. Our proposed model also considers the heterogeneity of edge and fog devices.

3. System Model and Problem Formulation

Fog-cloud is an emerging concept of connecting resource deficient edge devices to resourceful cloud servers through resource-limited fog devices for implementing IoT applications. The fog-cloud architecture provides resourceful cloud servers in a centralized manner and delivers limited processing capabilities through fog devices adjacent to the sensor nodes. The fog-cloud architecture due to its distribution of resources in a decentralized manner is effective in implementing applications on large scale. This paradigm offers mobility, reduced network load, and minimum latency for implementing IoT applications. Figure 1 shows a general diagram of the fog-cloud computing paradigm in which resource-limited fog nodes are providing resources near to the edge devices. The cloud server is behaving as a centralized unit for the collection of all the information arriving from the edge nodes after preliminary processing through fog nodes. The first tier of this network is based on IoT devices which consist of sensors and actuators. The second tier consists of resource-constrained fog devices to provide basic computational functionalities and connections towards the cloud. The third tier is based on cloud servers having huge storage and computational resources available in a centralized manner. Each layer is responsible for the execution of a specific module of the application.

There is a hierarchical decrease in the resources available on each layer. The cloud server is available on top of the architecture with plenty of resources to fulfil the computational and storage demands of the applications deployed. Resource constraint fog devices are part of the second layer to link IoT devices of the first layer with the cloud on-demand. Every node in the network is responsible for the execution of some application tasks. In dynamic fog-cloud networks, the computational burden is on the fog layer as they have to provide basic processing capabilities with their limited resources to reduce latency by providing minimized load towards the cloud server.

Dynamic fog-cloud environments consist of devices with different available processing capabilities. The parameters that limit the processing functionality of any fog device are CPU and RAM. The function defining the fog node is constrained by these two parameters. If the i_{th} fog node in the paradigm is represented as f_i, then the processing capability of that node is expressed as

$$P_c(f_i) = <\mathrm{CPU}_i, \mathrm{RAM}_i> . \quad (1)$$

The computational resources available in the fog paradigm is the sum of individual processing capabilities available at each fog device existing in the paradigm and is expressed as

$$N = \sum_{i=1}^{M} P_c\{f_i\}, \quad (2)$$

where M is the total number of fog nodes present in the network. The information sensed by the sensor nodes is transmitted to the connected parent fog device. The resources available at fog devices are consumed for the execution of tasks related to data sensed by the edge nodes. The edge

```
Input: Set of Fog devices $F$ and Edge devices $K$
Output: Assignment of appropriate edge nodes to fog devices
Fog devices $f_i \in Layer$  2, Edge devices $I_i \in Layer$  3
for each $I_i$ do
    if($R_i < R_{\mathrm{limit}}$)                    ▷$R_{\mathrm{limit}}$ is $predefined\, sen \sin g\, rate$ limit
        add $I_i$ to $K_L = \{\}$
    end
    else
        add $I_i$ to $K_H = \{\}$
    end
end
for each $f_i$do
        for $I_i \in K_H$ do
                if($P_c(f_i) < S(I_i)$)
                        add $I_i$ to $E_i = \{\}$
                        $P_c(f_i) = P_c(f_i) - S(I_i)$ ;
                end
                else
                        for $I_i \in K_L$ do
                            if($P_c(f_i) < S(I_i)$)
                                add $I_i$ to $E_i = \{\}$
                                $P_c(f_i) = P_c(f_i) - S(I_i)$ ;
                            end
                        end for
                end
        end for
end for
```

Algorithm 1: Resource-aware load allocation algorithm for a fog-cloud paradigm.

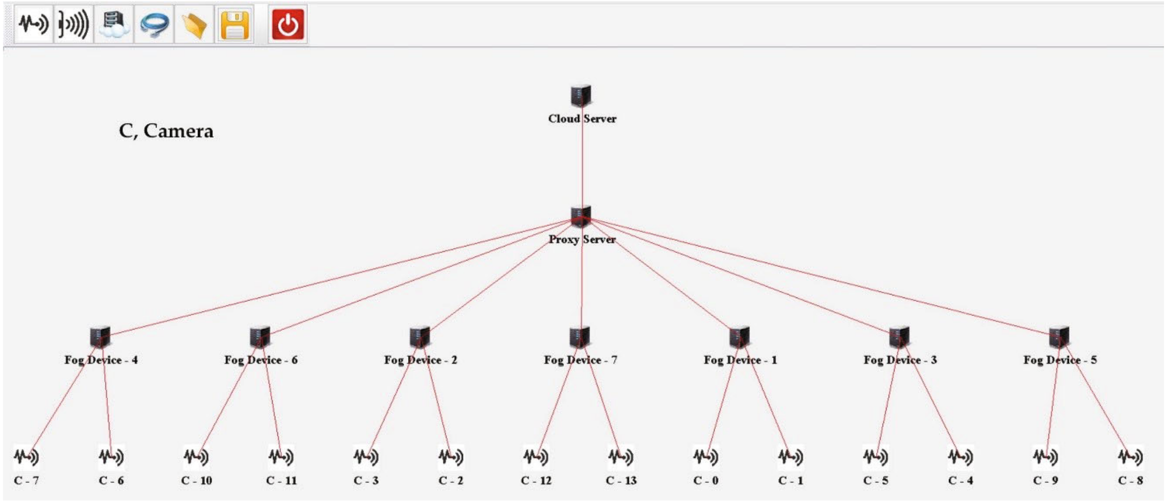


Figure 3: One of the scenarios created for the evaluation of the proposed paradigm.

devices are connected to the parent fog nodes to which they have to deliver the sensed information for instant processing. If there are a K number of total edge devices present in a system ($K = \{I_1, I_2, I_3, I_4 \cdots I_k\}$), then the edge devices connected to a fog device f_i are denoted by E_i. The sensing rate of the i_{th} edge device I_{th} is symbolized as R_{th}. The volume of sensed information by the sensor device depends upon the sensing rate of the sensor attached to the end devices. In the fog-cloud paradigm, the received data demanding high processing resources than the available at fog nodes is communicated to the parent cloud for execution. To achieve optimum performance of the fog-cloud paradigm, the volume of sensed load at fog nodes is an important parameter to be considered. This parameter was not exploited in literature work published before related to resource utilization and allocation in fog-cloud paradigms. For thei_{th}edge device,I_i, the total volume of sensed data is directly proportional to the sensing rate of the sensor attached and is expressed as$S(I_i)$. Consequently, the volume of sensed data that a fog device has to process depends on

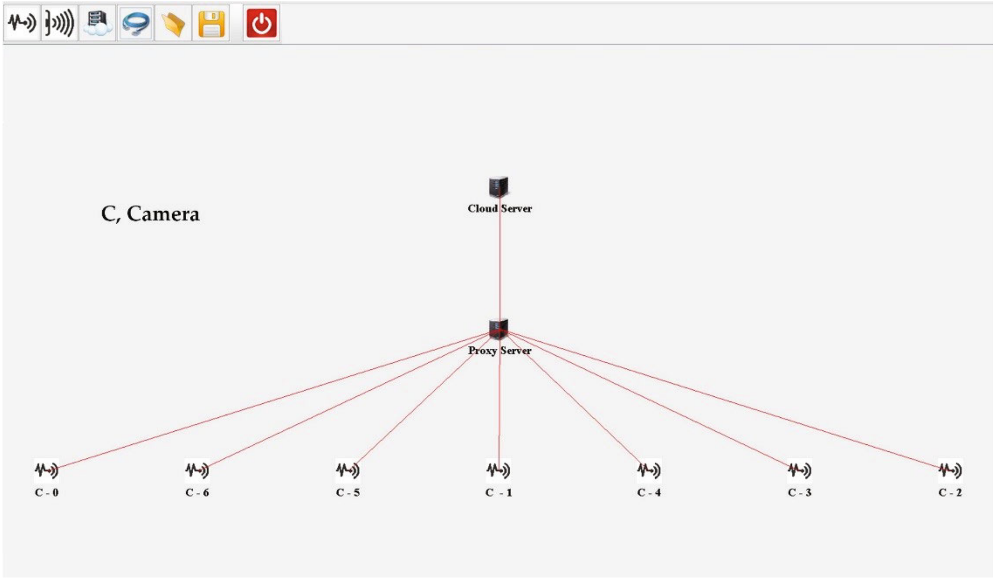


Figure 4: One of the scenarios created for the evaluation of the cloud-based approach.

Table 3: Network configurations related to network nodes used in simulations.

Specification	Cloud server	Gateway node	Fog device	End device
Rate per execution	0.01	0.0	0.0	0.0
Random access memory (MB)	40000	4000	2000 - 4000	1000
Downlink capacity (MB)	9.7656	9.7656	9.7656	9.7656
Processing power (MIPS)	448100	2800	2000 - 4000	500
Uplink capacity (MB)	0.0	9.7656	9.7656	9.7656

the sensed information volume at each edge device connected to that fog device which is given by

$$L(f_i) = \sum_{\forall I_i \in E_i} S(I_i). \tag{3}$$

For optimum performance of fog-cloud computing paradigm, the connection of appropriate edge devices according to available resources at parent fog device plays a pivotal role.

In the traditional deployment of applications on the fog-cloud computing paradigm, the edge devices are connected to fog devices irrespective of the sensing rate of sensors at the edge nodes which results in high network consumption and latency. In such situations, fog nodes having excess processing resources to handle the high volume of information might have a connection to edge nodes producing a low volume of data resulting in wastage of fog resources. Consequently, the resource constraint fog nodes having a connection to edge nodes with high volumetric production of sensed information can overburden the fog nodes to tackle such information. In this research, a policy is proposed that deploys edge nodes according to the availability of resources at the fog layer to reduce network burden and delay for efficient utilization of the fog-cloud network resources.

Table 4: Types and characteristics of tuples used in simulations.

Tuple	Tuple length	CPU length (MIPS)
PTZ_MODULE	28	100
VIDEO_MODULE	2000	2000
LOCATION_MODULE	100	1000
DETECTOR_MODULE	2000	500
CAMERA	20000	1000

This research redefines and simulates the intelligent surveillance application using distributed camera networks [28] to demonstrate the effectiveness of the proposed scheme. For the better understanding and representation of application components in a distributed computing environment, Distributed Data Flow Model (DDF) [29] was used in this research for the deployment of applications on the fog-cloud paradigm. Figure 2 presents the directed acyclic graph (DAG) of the intelligent surveillance application based on a distributed network of cameras deployed for the evaluation of the proposed algorithm. In the DAG model, the five modules of the deployed application are represented as a vertex that processes the information approaching from the preceding module. The arcs linking various modules describe the flow of data between different modules.

The DAG model of the application consists of five modules as shown in the above diagram. The motion detection

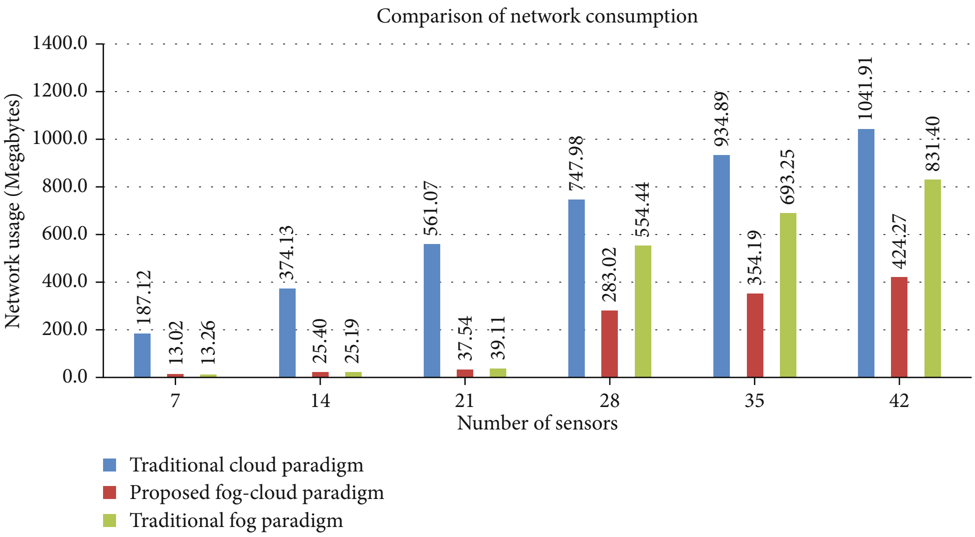


Figure 5: Comparison of network consumption.

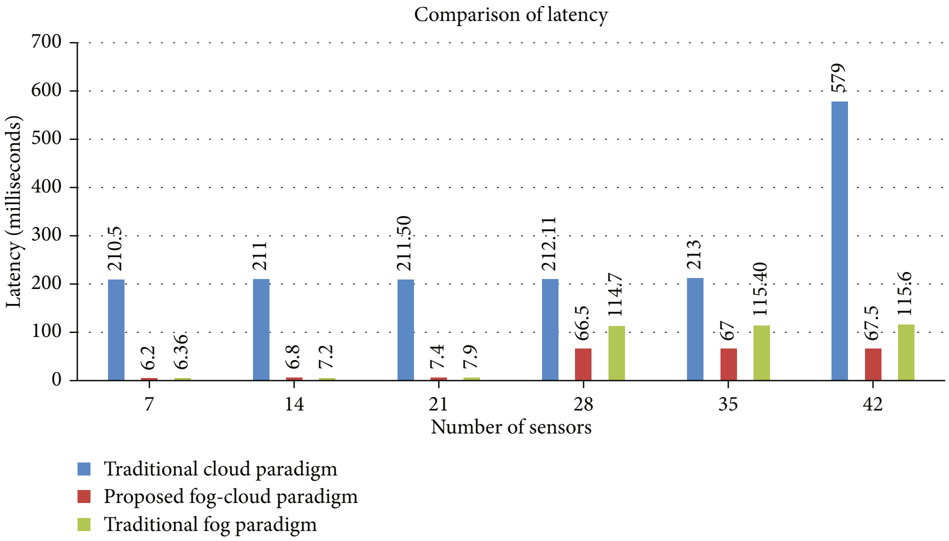


Figure 6: Comparison of latency.

module is placed in cameras for capturing video streams for detecting the motion of an object. This module forwards the information to the object detector module on the detection of motion of an object. Afterwards, the object detection module tracks the object and calculates the coordinates of the object. The object tracker module receives the coordinates previously calculated by the object detector modules and calculates the PTZ configurations for the camera for effective monitoring of the detected object. The PTZ control module after receiving the PTZ configurations adjusts the cameras accordingly. Finally, filtered video streams collected from the object detector module are conveyed to the user's device via the user interface module for better visualization of the tracked object. In the computing atmosphere, the basic unit used for intramodule communication is a tuple that is characterized by its specific length containing information to be processed and the number of resources required for its processing. Tuple mapping is described in DAG using coloured circles. For example, reception of a tuple of type RAW_VIDEO_STREAM on the module "Motion Detector" will result in a release of tuple type VIDEO_MODULE.

4. Proposed Solution

In this paper, a resource-aware load allocation algorithm is proposed that effectively manages the connection between edge nodes and parent devices by taking into account the available processing resources at the fog layer and volume of sensed information at end devices. The proposed algorithm allocates appropriate edge devices to each fog device present in the network after searching throughout the edge layer. To balance the processing load on fog nodes according to their computational resources, the algorithm effectively places edge nodes under fog nodes according to the sensing rate of sensors present at the edge nodes. The proposed approach classifies

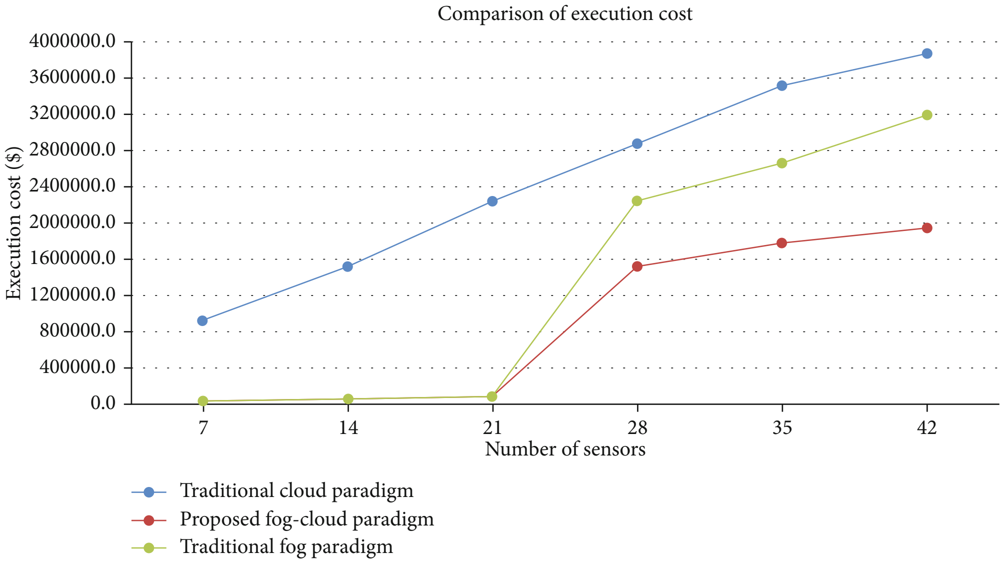


Figure 7: Comparison of execution cost.

and registers the edge devices as edge nodes with high sensing rates and low sensing rates. Afterwards, from the categorized edge nodes, a combination of edge nodes is allocated to the fog devices according to resources existing at the fog nodes. The proposed Algorithm 1 is given below.

Fog nodes and edge devices are given as input to the proposed algorithm. Initially, the algorithm categorizes the edge devices on the basis of sensing rates of the sensors attached to these edge nodes. If the sensing rate of the edge device is less than the predefined rate, then it is placed in the set K_L, otherwise the edge device is placed in the set K_H. Afterwards, the algorithm searches throughout the set$K_H$andK_L for the selection of appropriate edge devices. The algorithm assigns suitable edge devices to fog nodes for optimal performance. If the resources required for the processing of the sensed volume by an edge device are less than the available at the fog device, then that edge device is assigned to that fog device as child node.

5. Results and Discussion

For validating the effectiveness of the proposed scheme, an intelligent surveillance application is implemented on different scales. In each experimental scenario, the number of cameras monitoring the area under surveillance is increased. There are a total number of seven areas that are under surveillance during all the simulations. In all the simulated scenarios, the cameras are connected to fog nodes that are associated with the cloud server. One fog device per area under surveillance is assigned which provides resources close to the boundary of the network to observe and detect activity in that area. The fog nodes provide resources near to cameras for the processing of video streams recorded by cameras. The number of cameras per surveilled area is varied in each physical topology created in the simulation. Initially, two cameras per fog node are connected which are increased in each new topology created. The sensors created in the simulations are according to the strategy of [30]. In each scenario, the number of cameras per fog node is increased to analyze the delay and network consumption. The view of one of each scenario created in iFogSim for the proposed paradigm and traditional cloud computing paradigm is shown in Figures 3 and 4.

The tabulated network configurations, tuple types, simulation parameters, and sensing frequencies used in the simulations are shown in Tables 3 and 4. For evaluation, the intelligent surveillance application is implemented on the proposed load aware resource allocation fog-cloud paradigm and the traditional cloud and fog computing paradigms. The information detection frequencies of cameras deployed in the simulations take values between 5 ms and 20 ms.

A comparative analysis is performed between the proposed scheme and traditional cloud and fog paradigms by creating various simulation scenarios on multiple scales. Cost of processing at cloud, network consumption, and end-to-end delay are the parameters under observation during all these evaluations. Figure 5 presents a comparison between the consumption of networks during implementing the application on different paradigms. The proposed algorithm successfully decreases the load on the network as compared to fog and cloud-based implementations.

In the proposed approach, the parent fog nodes are assigned to the child devices according to the volume of data sensed by them. The volume of the sensed load is associated with the sensing rate of the sensing device, so edge devices having higher sensing rates are assigned to the fog nodes having higher data processing capacity to reduce the load on the network. Therefore, the load is aligned with the network resources available which in turn decreases the overall burden on the system. On the contrary, in the cloud architecture, all the sensed load is handed directly to the cloud server for processing resulting in high network utilization. Due to the inability of provision of fog resources according to sensed load, the traditional fog paradigm provides reduced network consumption as compared to the cloud paradigm but is not much network efficient as compared to the proposed model.

A comparison between the proposed paradigm with the traditional cloud and fog paradigms in terms of offered latency is presented in Figure 6. In cloud-based implementation, all the sensed data from the system is to be processed by the cloud server; thus, rise in latency is proportional to the number of sensors linked to the cloud, whereas in the fog computing paradigm, the processing of data is also provided at the intermediate level by the fog nodes resulting in a decrease in the information to be processed by the cloud server and decreasing the round-trip time. The core feature of the designed strategy is to reduce the latency and burden on the network by allocating the suitable fog resources to the edge nodes according to the rate at which the information is sensed by the edge nodes. The proposed model estimates the volume of information arriving from the edge devices by analyzing the sensing frequency of the sensors installed at the edge devices. Afterwards, the algorithm links the suitable fog devices to the sensor nodes according to the resources of the fog devices. The proposed policy reduces the amount of data to be processed at the cloud server by providing suitable fog resources according to the demand from the edge devices which in turn decreases the processing cost at the cloud as depicted in Figure 7.

The proposed algorithm estimates the volume of information generated by the edge nodes by incorporating the sensing rate information of the sensors attached to these nodes. Afterwards, the suitable edge nodes are linked to the parent fog devices. This provision of fog computing resources according to the volume of sensed information is not offered in traditional fog computing designs. This optimum balance between the detected volume of information to be processed at edge nodes and the processing capacity available at the parent fog device is the key feature of the proposed algorithm. Due to this salient feature, most of the information generated at the sensor nodes is processed at the linked fog device resulting in a reduction of information to be processed at the cloud server. Thus, a significant reduction in the cost of execution is observed as compared to traditional fog designs.

6. Conclusions

The algorithm proposed in this research effectively balances the available resources offered by the fog paradigm and the volume of information generated at the edge of the network. To achieve this, the proposed scheme manages the connection between the fog nodes and edge devices. The proposed strategy estimates the volume of information detected at the edge nodes and assigns appropriate parent fog devices from the available nodes at the fog paradigm. This efficient management of sensed load and processing resources of the network by the proposed algorithm effectively reduces the latency and network consumption of the system. The intelligent surveillance through distributed camera network application was implemented on different scales to compare the proposed algorithm with the traditional cloud and fog architectures. iFogSim toolkit is used to perform these simulations. The results of the comparison show that the proposed algorithm prominently reduces the processing cost at cloud, delay, and network consumption. The proposed strategy is capable to execute any type of application. My future work consists of deploying more applications on the proposed design and modification of the proposed algorithm for the examination of multiple parameters. Moreover, future research includes the analysis and design of expected glitches triggered due to node failure in the system.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Syed Rizwan Hassan and Ishtiaq Ahmad contributed to actualization, validation, methodology, formal analysis, investigation, software, and initial draft. Ateeq Ur Rehman, Seada Hussen and Habib Hamam contributed to actualization, validation, methodology, formal analysis, investigation, and initial draft. All authors read and approved the final version.

References

[1] A. S. Petrenko, S. A. Petrenko, K. A. Makoveichuk, and P. V. Chetyrbok, "The IIoT/IoT device control model based on narrow-band IoT (NB-IoT)," in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pp. 950–953, Moscow and St. Petersburg, Russia, 2018.

[2] H. Shahid, M. A. Shah, A. Almogren et al., "Machine learning-based mist computing enabled Internet of Battlefield Things," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 4, pp. 1–26, 2021.

[3] M. Villari, M. Fazio, S. Dustdar, O. Rana, and R. Ranjan, "Osmotic computing: a new paradigm for edge/cloud integration," *IEEE Cloud Computing*, vol. 3, no. 6, pp. 76–83, 2016.

[4] K. A. Awan, I. U. Din, A. Almogren, H. A. Khattak, and J. J. Rodrigues, "Edge trust-a lightweight data-centric trust management approach for green internet of edge things," *Wireless Personal Communications*, 2021.

[5] V. Moysiadis, P. Sarigiannidis, and I. Moscholios, "Towards distributed data management in fog computing," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 7597686, 14 pages, 2018.

[6] F. Firouzi, B. Farahani, and A. Marinšek, "The convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT)," *Information Systems*, vol. 107, article 101840, 2021.

[7] M. Ammad, M. A. Shah, S. U. Islam et al., "A novel fog-based multi-level energy-efficient framework for IoT-enabled smart environments," *IEEE Access*, vol. 8, pp. 150010–150026, 2020.

[8] S. Bose, D. Sarkar, and N. Mukherjee, "A framework for heterogeneous resource allocation in sensor-cloud environment," *Wireless Personal Communications*, vol. 108, no. 1, pp. 19–36, 2019.

[9] I. L. Santos, L. Pirmez, F. C. Delicato et al., "Zeus: a resource allocation algorithm for the cloud of sensors," *Future Generation Computer Systems*, vol. 92, pp. 564–581, 2019.

[10] T. C. Xavier, I. L. Santos, F. C. Delicato et al., "Collaborative resource allocation for Cloud of Things systems," *Journal of Network and Computer Applications*, vol. 159, article 102592, 2020.

[11] R. K. Naha, S. Garg, A. Chan, and S. K. Battula, "Deadline-based dynamic resource allocation and provisioning algorithms in Fog- Cloud environment," *Future Generation Computer Systems*, vol. 104, pp. 131–141, 2020.

[12] W. Wei, R. Yang, H. Gu, W. Zhao, C. Chen, and S. Wan, "Multi-objective optimization for resource allocation in vehicular cloud computing networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.

[13] K. Haseeb, I. U. Din, A. Almogren, I. Ahmed, and M. Guizani, "Intelligent and secure edge-enabled computing model for sustainable cities using green internet of things," *Sustainable Cities and Society*, vol. 68, article 102779, 2021.

[14] I. U. Din, A. Bano, K. A. Awan, A. Almogren, A. Altameem, and M. Guizani, "LightTrust: lightweight trust management for edge devices in industrial Internet of Things," *IEEE Internet of Things Journal*, 2021.

[15] T. Wang, Y. Liang, W. Jia, M. Arif, A. Liu, and M. Xie, "Coupling resource management based on fog computing in smart city systems," *Journal of Network and Computer Applications*, vol. 135, pp. 11–19, 2019.

[16] S. Tejaswini, N. Sriraam, and G. Pradeep, "Cloud-based framework for pain scale assessment in NICU-a primitive study with infant cries," in *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, pp. 1–4, Bangalore, India, 2018.

[17] F. Muheidat, L. Tawalbeh, and H. Tyrer, "Context-aware, accurate, and real time fall detection system for elderly people," in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, pp. 329–333, Laguna Hills, CA, USA, 2018.

[18] H. K. Apat, K. Bhaisare, B. Sahoo, and P. Maiti, "Energy efficient resource management in fog computing supported medical cyber-physical system," in *2020 International conference on computer science, Engineering and Applications (ICCSEA)*, pp. 1–6, Gunupur, India, 2020.

[19] B. Gj, "Internet of Things (IoT) and cloud computing based persistent vegetative state patient monitoring system: a remote assessment and management," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pp. 301–305, Belgaum, India, 2018.

[20] S. Shukla, M. F. Hassan, M. K. Khan, L. T. Jung, and A. Awang, "An analytical model to minimize the latency in healthcare internet-of-things in fog computing environment," *PLoS One*, vol. 14, no. 11, article e0224934, 2019.

[21] K. S. Awaisi, A. Abbas, M. Zareei et al., "Towards a fog enabled efficient car parking architecture," *IEEE Access*, vol. 7, pp. 159100–159111, 2019.

[22] M. Taneja and A. Davy, "Resource aware placement of IoT application modules in Fog-Cloud computing paradigm," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pp. 1222–1228, Lisbon, Portugal, 2017.

[23] S. R. Hassan, I. Ahmad, J. Nebhen, A. U. Rehman, M. Shafiq, and J.-G. Choi, "Design of latency-aware IoT modules in heterogeneous fog-cloud computing networks," *CMC-COMPUTERS MATERIALS & CONTINUA*, vol. 70, no. 3, pp. 6057–6072, 2022.

[24] C. S. Nandyala and H.-K. Kim, "From cloud to fog and IoT-based real-time U-healthcare monitoring for smart homes and hospitals," *International Journal of Smart Home*, vol. 10, no. 2, pp. 187–196, 2016.

[25] S. S. N. Perala, I. Galanis, and I. Anagnostopoulos, "Fog computing and efficient resource management in the era of Internet-of-Video Things (IoVT)," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, Florence, Italy, 2018.

[26] S. Azizi, F. Khosroabadi, and M. Shojafar, "A priority-based service placement policy for fog-cloud computing systems," *Computational Methods for Differential Equations*, vol. 7, pp. 521–534, 2019.

[27] O. Skarlat, M. Nardelli, S. Schulte, M. Borkowski, and P. Leitner, "Optimized IoT service placement in the fog," *Service Oriented Computing and Applications*, vol. 11, no. 4, pp. 427–443, 2017.

[28] B. Song, C. Ding, A. T. Kamal, J. A. Farrell, and A. K. Roy-Chowdhury, "Distributed camera networks," *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 20–31, 2011.

[29] M. Ashouri, P. Davidsson, and R. Spalazzese, "Quality attributes in edge computing for the Internet of Things: a systematic mapping study," *Internet of Things*, vol. 13, article 100346, 2021.

[30] H. Gupta, A. Vahid Dastjerdi, S. K. Ghosh, and R. Buyya, "iFogSim: a toolkit for modeling and simulation of resource management techniques in the Internet of Things, edge and fog computing environments," *Software: Practice and Experience*, vol. 47, pp. 1275–1296, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1495934, 7 pages
https://doi.org/10.1155/2022/1495934

Research Article

Advanced FMECA Method Based on Intuitionistic 2-Tuple Linguistic Variables and the Triangular Fuzzy Analytic Hierarchy Process

Guangze Pan,[1] **Dan Li**,[1,2] **Qian Li**,[1] **Yaqiu Li**,[1,3] **and Yuanhang Wang**[1]

[1]*Center for Reliability and Environmental Engineering, China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 511370, China*
[2]*Center for Reliability and Environmental Engineering, Guangdong Provincial Key Laboratory of Electronic Information Products Reliability Technology, Guangzhou 511370, China*
[3]*Key Laboratory of Active Medical Devices Quality & Reliability Management and Assessment, Guangzhou 511370, China*

Correspondence should be addressed to Yaqiu Li; ermao13@163.com

Received 22 April 2022; Accepted 23 May 2022; Published 6 June 2022

Academic Editor: Maode Ma

Failure mode effects and criticality analysis (FMECA) is a commonly adopted approach to defining, assessing, and reducing possible failures in designs, systems, processes, products, and services. Traditional FMECA ranks the failure modes of products based on a risk priority number (RPN), which is obtained by multiplying the risk elements. Conventional FMECA has the shortcomings of badly handling unknown information and unreasonably assessing RPNs. To deal with these issues, an advanced FMECA method based on intuitionistic 2-tuple linguistic variables (I2LVs) and the triangular fuzzy analytic hierarchy process (TFAHP) is proposed. In this method, the fuzzy evaluation of risk elements given by different FMECA members is represented by I2LVs, which can efficiently handle unknown information. The TFAHP method is adopted to assess the weights of risky elements and rank the risk priorities of different failure modes. Finally, an application case of an insulated-gate bipolar transistor is used to verify the effectiveness and robustness of the proposed method.

1. Introduction

Failure mode effects and criticality analysis (FMECA) is typically adopted to identify, evaluate, and reduce existing or underlying errors and failures in system designs or processes [1]. Due to its simplicity and high efficiency, FMECA has been widely applied in some industries, such as the nuclear, aerospace, transportation, and manufacturing industries [2–9]. In traditional FMECA [10–13], every failure mode is assessed using three risk elements: detection (D), severity (S), and occurrence (O). The risk factors for the failure modes are integers between 1 and 10. By multiplying the D , S, and O values, a risk priority number (RPN) can be acquired. Although traditional FMECA has been proven to be a useful way to assess possible product failures in various areas, some shortcomings and limitations remain. For example, the hazard analysis is highly subjective and ignores the uncertainties of the actual intermediate state and the fuzzy nature of the language information. Moreover, risk factors are not weighed, meaning that they are taken to be equally valuable. These shortcomings lead to errors between traditional FMECA evaluations and actual results, which significantly limits their effectiveness.

To overcome these shortcomings and limitations, a reasonable evaluation can be made by using fuzzy numbers and linguistic variables instead of exact values. George et al. [14] and Mangeli et al. [15] converted language descriptions into triangular and trapezoidal fuzzy numbers or linguistic variables in FMECA. Xiao [16] and Liu et al. [17] used D numbers to more flexibly and intuitively represent the attribute information in an FMECA multiple criteria decision. Wang et al. [18] identified an exceptional fuzzy number and proposed a corresponding fuzzy RPN to identify the risk priority of failures. However, these

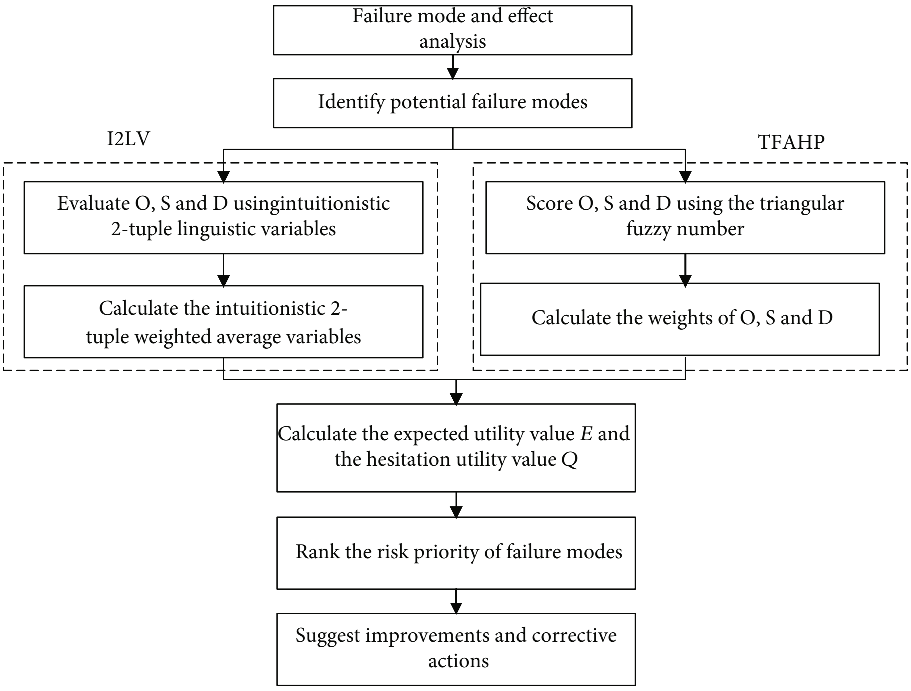

Failure mode and effect analysis
Identify potential failure modes
I2LV
TFAHP
Evaluate O, S and D usingintuitionistic 2-tuple linguistic variables
Calculate the intuitionistic 2-tuple weighted average variables
Score O, S and D using the triangular fuzzy number
Calculate the weights of O, S and D
Calculate the expected utility value E and the hesitation utility value Q
Rank the risk priority of failure modes
Suggest improvements and corrective actions

Figure 1: Implementation process of the advanced FMECA method.

Table 1: Judgment matrix.

Risk factor	O	S	D
O	(0.5,0.5,0.5)	(0.35,0.43,0.51)	(0.54,0.63,0.72)
S	(0.49,0.57,0.65)	(0.5,0.5,0.5)	(0.63,0.72,0.82)
D	(0.28,0.37,0.45)	(0.18,0.28,0.37)	(0.5,0.5,0.5)

evaluations also have shortcomings; e.g., the evaluation implies that the linguistic description of the subject is affiliated with the set and cannot use linguistic descriptions that are not affiliated with the set, such as the hesitation that a decision-maker cannot judge. Therefore, intuitionistic 2-tuple linguistic variables (I2LVs) [19, 20] consisting of language terms and explicit numbers can be used to describe the risk factors. I2LVs use qualitative linguistic terms to express criterion membership and nonmembership. This can summarize the ambiguity of linguistic information better than intuitionistic fuzzy numbers and linguistic variables can, thereby reflecting the actual situation more reliably and truthfully.

In addition, given the limitations of traditional FMECA (which does not weigh the risk elements), many researchers have tried to use the multiple-criteria decision-making (MCDM) methods instead of typical RPN methods to prioritize failure modes. Liu et al. [21] combined vague numbers and the VIKOR method to propose a new FMEA method that is particularly applicable to MCDM problems with confusing and incommensurable (comprising various units) standards. Alencar et al. [22] proposed an MCDM model to rank the possible causes of failure by considering more attributes and using the same methodological support as in MCDM. Ju et al. [23] and Geum et al. [24] used grey relation projection (GRP) and grey relation analysis (GRA) methods to decide the weight and risk priority of failure modes. The triangular fuzzy analytic hierarchy process (TFAHP) [25] is a widely used comprehensive evaluation method. Compared with traditional evaluation methods, TFAHP introduces triangular fuzzy numbers in an expert scoring process, making the evaluation results more reasonable, accurate, and operable.

To clarify the fuzzy information of FMECA and improve the accuracy of the analysis, a new FMECA method based on I2LVs and TFAHP is proposed. The I2LVs are adopted to describe the vague evaluation of the risk elements by FMECA members, while TFAHP is used to assess the weights of the risk elements and comprehensively rank the risk priorities of the failures. This method is suitable for the FMECA of various products that have multiple failure modes.

2. Preliminaries

2.1. Intuitionistic 2-Tuple Linguistic Variables. The 2-tuple linguistic variables (2LVs) refer to 2-tuple group evaluation data (l_i, ε_i), where l_i is the language term in the language term set $L = \{l_0, l_1, \cdots, l_{2\tau}\}$, and $\varepsilon_i \in (-0.5, 0.5]$ is the symbol transfer value, which indicates the error between the integrated language term and the closest original one.

Definition 1. Let $L = \{l_0, l_1, \cdots, l_{2r}\}$ be the language term set and $\widehat{L}$ be the extended language term set of L; then, the I2LVs are as follows:

$$H = \left\{ \left\langle l_{\alpha(x)}, l_{\beta(x)} \right\rangle \middle| x \in X, l_{\alpha(x)}, l_{\beta(x)} \in \widehat{L} \right\}, \tag{1}$$

Table 2: Results of the I2LV-TFAHP method.

Failure mode	E	Q	Rank
FM_1	0.681	0.126	1
FM_2	0.476	0.087	5
FM_3	0.456	0.101	6
FM_4	0.595	0.130	3
FM_5	0.363	0.109	7
FM_6	0.638	0.157	2
FM_7	0.584	0.185	4

Table 3: Priority ranks evaluated by the five FMECA methods.

Failure mode	I2LV-TFAHP	TFAHP	I2LV	I2LV-TOPSIS	Traditional method
FM_1	1	1	1	1	2
FM_2	5	6	7	6	4
FM_3	6	7	5	5	4
FM_4	3	4	3	2	5
FM_5	7	5	6	7	7
FM_6	2	3	4	3	1
FM_7	4	2	2	4	3

where the language terms $l_{\alpha(x)}$ and $l_{\beta(x)}$ indicate the membership and nonmembership degrees of element x, respectively, and satisfy $0 \le \alpha(x) + \beta(x) \le 2\tau$.

Definition 2. Let U and V be continuous and strictly monotonic utility functions of $\widehat{L}$. If, for any I2LV, U and V satisfy the following functions:

$$\begin{cases} U\left(l_{\alpha(x)}\right) = V\left(\operatorname{neg}\left(l_{\alpha(x)}\right)\right), \\ V\left(l_{\beta(x)}\right) = U\left(\operatorname{neg}\left(l_{\beta(x)}\right)\right), \\ U(l_0) = V(l_{2\tau}) = 0, \\ U(l_{2t}) = V(l_0) = 1, \end{cases} \tag{2}$$

then U and V are called the utility functions of membership degree $l_{\alpha(x)}$ and nonmembership degree $l_{\beta(x)}$, respectively.

Definition 3. Let $h_i = \langle l_{\alpha(h_i)}, l_{\beta(h_i)} \rangle (i = 1, 2, \cdots, m)$ be an I2LV and $w = (w_1, w_2, \cdots, w_n)^T$ be its weight vector, satisfying $\sum_{i=1}^n w_i = 1$ and $w_i \ge 0$. Then, the intuitionistic 2-tuple weighted average variables (I2WAVs) are

$$\begin{aligned} h &= I2WAV(h_1, h_2, \cdots, h_m) \\ &= \left\langle U^{-1}\left[\sum_{i=1}^m w_i U\left(l_{\alpha(h)}\right)\right], V^{-1}\left[\sum_{i=1}^m w_i V\left(l_{\beta(k)}\right)\right]\right\rangle. \end{aligned} \tag{3}$$

Definition 4. Let $h = \langle l_{\alpha(h)}, l_{\beta(h)} \rangle$ be an I2LV; then, the expected utility function of h is

$$E(h) = \frac{U\left(l_{\alpha(h)}\right) + V\left(l_{\beta(h)}\right)}{2}. \tag{4}$$

Definition 5. Let $h = \langle l_{\alpha(h)}, l_{\beta(h)} \rangle$ be an I2LV; then, the hesitation utility function of h is

$$E(h) = \frac{U\left(l_{\alpha(h)}\right) + V\left(l_{\beta(h)}\right)}{2}. \tag{5}$$

Definition 6. Let h_1 and h_2 be two I2LVs; then

(i) if $E(h_1) > E(h_2)$, then $h_1 > h_2$

(ii) if $E(h_1) = E(h_2)$ and $Q(h_1) < Q(h_2)$, then $h_1 > h_2$

(iii) if $E(h_1) = E(h_2)$ and $Q(h_1) = Q(h_2)$, then $h_1 = h_2$

2.2. TFAHP Method. The core of the analytic hierarchy process (AHP) is to construct a judgment matrix by using an integer between 1 and 9, with its inverse number used as a scale. This evaluation often does not consider the fuzzy nature of the subjective judgment. If the ratio of the weights of the two factors is not easy to determine, it is only known that the range of change is $l - u$, and the maximum possible value is m; then, the triangular fuzzy number (l, m, u) can be used in the evaluation.

In comparison with the traditional AHP method, TFAHP uses triangular fuzzy numbers in the expert scoring process, which makes the scoring relatively reasonable and accurate [26]. Moreover, during the weight calculation, the problem of judging and adjusting the consistency of the matrix in the traditional AHP is skillfully solved by adopting a possibility matrix [27].

First of all, a set of triangular fuzzy judgment matrices is established as follows:

$$\begin{cases} A^{(k)} = \left(a_{ij}^{(k)}\right)_{n\times n} \\ a_{ij}^{(k)} = \left(l_{ij}^{(k)}, m_{ij}^{(k)}, u_{ij}^{(k)}\right) \\ \text{s.t.,} 0 \leqslant l_{ij}^{(k)} \leqslant m_{ij}^{(k)} \leqslant u_{ij}^{(k)} \leqslant 1 \end{cases} \tag{6}$$

where n is the number of factors, k is the expert serial number, K is the total number of experts (where $k = 1, 2, \cdots, K$), $a_{ij}^{(k)}$ is the ratio of the importance of factor i to factor$j$, $l_{ij}^{(k)}$ and $u_{ij}^{(k)}$ are the upper and lower bounds of the triangular fuzzy number, respectively (where $l_{ij}^{(k)} + u_{ji}^{(k)} = 1$, $u_{ij}^{(k)} + l_{ji}^{(k)} = 1$, $l_{ii}^{(k)} = 0.5$, $u_{ii}^{(k)} = 0.5$), and m_{ij} is the median of the triangular fuzzy number $m_{ij}^{(k)} + m_{ji}^{(k)} = 1$.

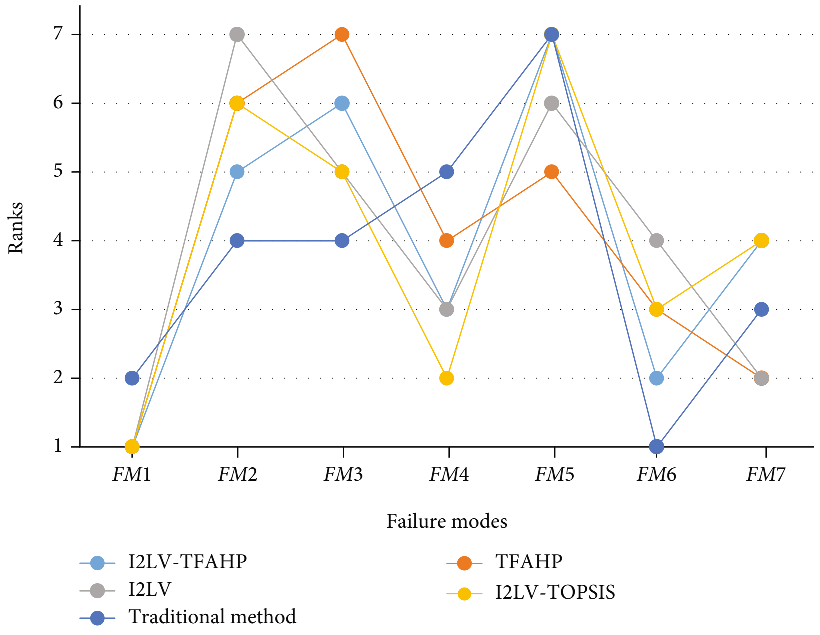


Figure 2: Comparison of the results of the five FMEA methods.

Table 4: Expert weights in different cases.

Cases	TE_1	TE_2	TE_3	TE_4
Case 0	0.2	0.35	0.15	0.3
Case 1	0.1	0.4	0.3	0.2
Case 2	0.25	0.25	0.25	0.25
Case 3	0.3	0.2	0.35	0.15
Case 4	0.4	0.1	0.4	0.1

The set of triangular fuzzy judgment matrices is merged according to the following function:

$$a_{ij} = \frac{a_{ij}^{(1)} + a_{ij}^{(2)} + \cdots + a_{ij}^{(K)}}{K}. \tag{7}$$

Secondly, the single-level triangular fuzzy weights of the judgment matrix are calculated.

$$c_i = \frac{\sum_{j=1}^{n} a_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}} = \left(\frac{\sum_{j=1}^{n} l_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{ij}}, \frac{\sum_{j=1}^{n} m_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}}, \frac{\sum_{j=1}^{n} \mu_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n} l_{ij}}\right). \tag{8}$$

The single-level triangular fuzzy weights c_i are compared with each other. Let $c_{i1} = (c_{l1}, c_{m1}, c_{\mu 1})$ and $c_{i2} = (c_{l2}, c_{m2}, c_{\mu 2})$; then, the likelihood matrix $P = (p_{ij})_{n\times n}$ is solved as follows:

$$\begin{aligned} p(c_{i1} \geq c_{i2}) = 0.5 \max &\left\{1 - \max\left\{\frac{c_{m2} - c_{l1}}{c_{m1} - c_{l1} + c_{m2} - c_{l2}}, 0\right\}, 0\right\} \\ &+ 0.5 \max\left\{1 - \max\left\{\frac{c_{\mu 2} - c_{m1}}{c_{\mu 1} - c_{m1} + c_{\mu 2} - c_{m2}}, 0\right\}, 0\right\}. \end{aligned} \tag{9}$$

Then, the likelihood matrix is transformed into a fuzzy matrix with consistent features.

$$r_{ij} = \frac{r_i - r_j}{2(n-1)} + 0.5, \tag{10}$$

$$r_i = \sum_{k=1}^{n} P_{ik}. \tag{11}$$

Finally, the final weights are calculated using the following formula:

$$w_i = \frac{\sum_{j=1}^{n} r_{ij} + (n/2) - 1}{n(n-1)}. \tag{12}$$

3. Methods

Supposing that there are p FMECA team experts $TE = \{TE_1, TE_2, \cdots, TE_p\}$ evaluating m failure modes $FM = \{FM_1, FM_2, \cdots, FM_m\}$ with respect to n risk factors $RF = \{RF_1, RF_2, \cdots, RF_n\}$, the expert weight vector is $w = (w^{(1)}, w^{(2)}, \cdots, w^{(p)})^T$. The expert TE_k gives an evaluation of risk factor RF_j of failure mode FM_i as a I2LV $b_{ij}^{(k)} = \langle l_{\alpha_{ij}^{(k)}}, l_{\beta_{ij}^{(k)}} \rangle$, with the I2LV matrix being $B^{(k)} = (b_{ij}^{(k)})_{m\times n}$ $(k = 1, 2, \cdots, p)$. The set of triangular fuzzy judgment matrices of risk factors is $A^{(k)} = (a_{ij}^{(k)})_{n\times n}$.

The proposed FMECA method includes the steps shown in Figure 1.

Step 1. According to I2WAV and the expert weight vector w, the I2LV matrix $B^{(k)}$ is integrated into the group I2LV

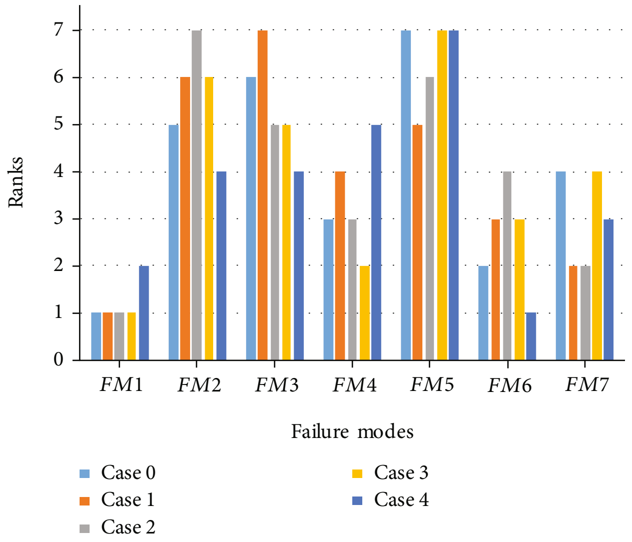


Figure 3: Sensitivity analysis for the I2LV-TFAHP method.

Table 5: Failure mode priority ranks under various cases.

Failure mode	Case 0	Case 1	Case 2	Case 3	Case 4
FM_1	1	1	1	1	2
FM_2	5	6	7	6	4
FM_3	6	7	5	5	4
FM_4	3	4	3	2	5
FM_5	7	5	6	7	7
FM_6	2	3	4	3	1
FM_7	4	2	2	4	3

matrix $B = (b_{ij})_{m\times n} = (\langle l_{\alpha_{ij}}, l_{\beta_{ij}}\rangle)_{m\times n}$.

$$b_{ij} = \mathrm{I2WAV}\left(b_{ij}^{(1)}, b_{ij}^{(2)}, \cdots, b_{ij}^{(p)}\right) = \left\langle U^{-1}\left[\sum_{k=1}^{p} w^{(k)} U\left(l_{\alpha_i^{(k)}}\right)\right], V^{-1}\left[\sum_{k=1}^{p} w^{(k)} V\left(l_{\beta_{ij}^{(k)}}\right)\right]\right\rangle. \tag{13}$$

Step 2. The TFAHP method is adopted to assess the weight vector of risk elements $\varpi = (\varpi_1, \varpi_2, \cdots, \varpi_n)^T$.

Step 3. According to the group I2WAV and the weight vector of risk elements, the group I2WAV for every risk element of failure modes A_i is integrated to obtain the comprehensive I2WAV $b_i = \langle l_a, l_b\rangle$.

$$b_i = I2WAV(b_{i1}, b_{i2}, \cdots, b_{in}) = \left\langle U^{-1}\left[\sum_{j=1}^{n} \varpi_j U\left(l_{\alpha_{ij}}\right)\right], V^{-1}\left[\sum_{j=1}^{n} \varpi_j V\left(l_{\beta_{ij}}\right)\right]\right\rangle. \tag{14}$$

Step 4. According to Definitions 4 and 5, the expected utility value $E(b_i)$ and the hesitation utility value $Q(b_i)$ of the comprehensive I2WAV b_i are calculated.

Step 5. According to Definition 6, the expected utility value and the hesitation utility value are sorted to acquire the failure mode's risk priorities.

4. Case Study

The section uses the new FMEA method according to I2LV and TFAHP (I2LV-TFAHP method) to assess the failure modes of a crimp-type insulated-gate bipolar transistor (IGBT).

Step 1. The FMECA team consists of four experts, $\{TE_1, TE_2, TE_3, TE_4\}$. The risk elements are detection (D), severity (S), and occurrence (O). The weight of the experts in each risk factor is $w = (0.2, 0.35, 0.15, 0.3)^T$, and the linguistic term set is $L = \{l_0, l_1, l_2, l_3, l_4, l_5, l_6\} = \{$extremely low, very low, low, moderate, high, very high, extremely high$\}$. Through detailed analysis of the IGBT, the FMECA team identifies seven potential failure modes: fretting wear (FM_1), short circuit (FM_2), open circuit (FM_3), microcorrosion (FM_4), boundary warpage (FM_5), gate oxide-layer destruction (FM_6), and spring failure (FM_7).

Step 2. The experts use I2LV to evaluate the failure modes. The I2LV matrices $B^{(k)} = (b_{ij}^{(k)})_{m\times n} (k = 1, 2, 3, 4 ; i = 1, 2, 3 ; j = 1, 2, 3, 4, 5, 6, 7)$ are provided.

Step 3. According to I2WAV and the experts' weight vector w, the group I2LV matrix is obtained.

Step 4. Experts use triangular fuzzy numbers to score and average the risk factors to obtain a judgment matrix (Table 1). The risk element weights calculated by the TFAHP method are $\varpi = (0.319, 0.485, 0.196)^T$.

Step 5. According to Equation (14), the comprehensive I2WAV is calculated.

$$\begin{cases} b_1 = \langle l_{0.68}, l_{0.48}\rangle, b_2 = \langle l_{2.59}, l_{2.89}\rangle, b_3 = \langle l_{2.43}, l_{2.96}\rangle, \\ b_4 = \langle l_{3.18}, l_{2.04}\rangle, b_5 = \langle l_{1.85}, l_{3.49}\rangle, b_6 = \langle l_{3.36}, l_{1.70}\rangle, \\ b_7 = \langle l_{2.95}, l_{1.94}\rangle. \end{cases} \tag{15}$$

Step 6. The expected utility value and hesitation utility value of the comprehensive I2WAV $b_i (i = 1, 2, 3, 4, 5, 6, 7)$ are calculated according to Equations (4) and (5).

Step 7. The failure mode risk priority ranks are determined based on $E(b_i)$ and $Q(b_i)$, as shown in the last column of Table 2.

Table 3 shows that FM_1 has the highest expected utility degree in fretting wear failure and, therefore, should be offered a top risk priority. The order of the risk priorities of the seven failure modes is $FM_1 > FM_6 > FM_4 > FM_7 > FM_2 > FM_3 > FM_5$.

By comparing TFAHP, I2LV, I2LV-TOPSIS, and the traditional method, the rationality of the I2LV-TFAHP method is verified. The evaluation outcomes are displayed in Table 3 and Figure 2.

Table 4 and Figure 2 demonstrate the following facts.

TFAHP, I2LV, and I2LV-TOPSIS all indicate that the failure mode with the highest risk priority is FM_1 (fretting wear), which is consistent with the conclusion of I2LV-TFAHP.

I2LV-TFAHP ranks FM_6 (gate oxide-layer destruction) in second place, while TFAHP and I2LV rank FM_7 (spring failure) in second place and the I2LV-TOPSIS method ranks FM_4 (microcorrosion) in second place. The TFAHP, I2LV, and I2LV-TOPSIS methods rank FM_6 in fifth, fourth, and third places, respectively. Using engineering knowledge during the operation of IGBT, the failure frequency of the gate oxide-layer destruction is higher than that of the spring failure as well as microcorrosion, as shown in the risk element *O*. Thus, in the I2LV-TFAHP method, it is reasonable to rank the FM_6 risk priority in second place.

The I2LV-TFAHP ranks FM_5 (boundary warpage) in seventh place, while the TFAHP and I2LV methods rank FM_3 (open circuit) and FM_2 (short circuit) in seventh place. The open circuit and short circuit of the IGBT make it unable to work, and the effect of boundary warpage on the IGBT function is small, as reflected in the risk factors. Therefore, it is also reasonable for the I2LV-TFAHP method to rank FM_5 in seventh place.

The first four failure modes with the highest risk, as evaluated by typical methods, are not the same as those of the four other methods, particularly when all of the new methods select FM_1 as the highest-risk failure mode. Moreover, the traditional method ranks FM_2 and FM_3 in equal fourth place, meaning that it cannot distinguish between their risks.

The above information indicates that the proposed I2LV-TFAHP method is more reasonable and accurate than other methods.

Sensitivity analysis is performed by changing the expert weights, as shown in Table 4, where case 0 demonstrates the previously mentioned application, and cases 1 to 4 indicate cases with different weights.

The ranking outcomes of the failure mode risk priorities in various cases are displayed in Figure 3 and Table 5. They show that a change in expert weight has a very slight influence on the risk priority rank, which means that the suggested method is sufficiently robust in ranking the risk priorities of the failure modes identified in the FMECA.

5. Conclusions

In this study, an advanced FMECA method based on I2LV and TFAHP is proposed to deal with the shortcomings and limitations of the conventional FMECA method. It also suggests a different way to prioritize the risks of different failure modes. I2LV can effectively deal with the fuzzy nature of linguistic variables, while TFAHP is adopted to evaluate the risk element weights and the comprehensive ranking of the failure mode risk priorities. An application example is provided to demonstrate the failure mode risk priorities in the FMECA provided by the IGBT. Compared with TFAHP, I2LV, I2LV-TOPSIS, and the traditional method, the proposed I2LV-TFAHP method is more accurate and reasonable and has greater robustness when analysing risk priority ranks.

Data Availability

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

Conflicts of Interest

The authors declare that there is no conflict of interest in publishing this article.

Acknowledgments

This research was funded by the Key-Area Research and Development Program of Guangdong Province (2019B111109001), the 2019 Industrial Technology Basic Public Service Platform Project (2019-00896-3-1), the Guangzhou Science and Technology Project (201904010016) and the Guangdong Basic and Applied Basic Research Foundation (2021A1515110679).

References

[1] G. F. Li, Y. Li, C. H. Chen, J. L. He, T. W. Hou, and J. H. Chen, "Advanced FMEA method based on interval 2-tuple linguistic variables and TOPSIS," *Quality Engineering*, vol. 4, pp. 1–10, 2020.

[2] S. Carpitella, A. Certa, J. Izquierdo, and C. M. la Fata, "A combined multi-criteria approach to support FMECA analyses: a real-world case," *Reliability Engineering & System Safety*, vol. 169, no. 1, pp. 394–402, 2018.

[3] A. Certa, F. Hopps, R. Inghilleri, and C. M. la Fata, "A Dempster-Shafer theory-based approach to the Failure Mode, Effects and Criticality Analysis (FMECA) under epistemic uncertainty: application to the propulsion system of a fishing vessel," *Reliability Engineering & System Safety*, vol. 159, no. 3, pp. 69–79, 2017.

[4] H. A. Khorshidi, I. Gunawan, and M. Y. Ibrahim, "Data-driven system reliability and failure behavior modeling using fmeca," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1253–1260, 2016.

[5] E. K. Youn and K. H. Choi, "A study on reliability analysis of electric railway catenary system using fmeca," *Transactions of the Korean Institute of Electrical Engineers*, vol. 64, no. 11, pp. 1618–1625, 2015.

[6] C. Spreafico, D. Russo, and C. Rizzi, "A state-of-the-art review of FMEA/FMECA including patents," *Computer Science Review*, vol. 25, no. 8, pp. 19–28, 2017.

[7] A. E. Brom, O. V. Belova, and A. Sissinio, "Lifecycle costs for energy equipment: FMECA & lifecycle costing models as "decision making" tools for cost reduction during the whole equipment life," *Procedia Engineering*, vol. 152, no. 12, pp. 173–176, 2016.

[8] A. E. Brom, I. N. Omelchenko, and O. V. Belova, "Lifecycle costs for energy equipment FMECA for gas turbine," *Procedia Engineering*, vol. 152, no. 12, pp. 177–181, 2016.

[9] J. Singh, S. Singh, and A. Singh, "Distribution transformer failure modes, effects and criticality analysis (FMECA)," *Engineering Failure Analysis*, vol. 99, no. 2, pp. 180–191, 2019.

[10] R. Ahmad, S. Kamaruddin, I. A. Azid, and I. P. Almanar, "Failure analysis of machinery component by considering external factors and multiple failure modes - a case study in the processing industry," *Engineering Failure Analysis*, vol. 25, no. 10, pp. 182–192, 2012.

[11] J. Li and H. Xu, "Reliability analysis of aircraft equipment based on fmeca method," *Physics Procedia*, vol. 25, pp. 1816–1822, 2012.

[12] M. Catelani, L. Ciani, L. Cristaldi, M. Faifer, and M. Lazzaroni, "Electrical performances optimization of photovoltaic modules with FMECA approach," *Measurement*, vol. 46, no. 10, pp. 3898–3909, 2013.

[13] Z. Q. Cai, S. D. Sun, S. B. Si, and N. Wang, "Modeling of failure prediction Bayesian network based on fmeca," *System Engineering Theory and Practice*, vol. 33, pp. 187–193, 2013.

[14] J. J. George, V. R. Renjith, P. George, and A. S. George, "Application of fuzzy failure mode effect and criticality analysis on unloading facility of LNG terminal," *Journal of Loss Prevention in the Process Industries*, vol. 61, pp. 104–113, 2019.

[15] M. Mangeli, A. Shahraki, and F. H. Saljooghi, "Improvement of risk assessment in the FMEA using nonlinear model, revised fuzzy TOPSIS, and support vector machine," *International Journal of Industrial Ergonomics*, vol. 69, pp. 209–216, 2019.

[16] F. Xiao, "A novel multi-criteria decision making method for assessing health-care waste treatment technologies based on D numbers," *Engineering Applications of Artificial Intelligence*, vol. 71, pp. 216–225, 2018.

[17] H. C. Liu, J. X. You, X. J. Fan, and Q. L. Lin, "Failure mode and effects analysis using D numbers and grey relational projection method," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4670–4679, 2014.

[18] Y. M. Wang, K. S. Chin, G. K. K. Poon, and J. B. Yang, "Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1195–1207, 2009.

[19] P. Liu and S. M. Chen, "Multiattribute group decision making based on intuitionistic 2-tuple linguistic information," *Information Sciences*, vol. 430, pp. 599–619, 2018.

[20] K. Du and H. Yuan, "Interval-valued intuitionistic 2-tuple linguistic Bonferroni mean operators and their applications in multi-attribute group decision making," *International Journal of Fuzzy Systems*, vol. 21, no. 8, pp. 2373–2391, 2019.

[21] H. C. Liu, J. X. You, X. Y. You, and M. M. Shan, "A novel approach for failure mode and effects analysis using combination weighting and fuzzy VIKOR method," *Applied Soft Computing*, vol. 28, pp. 579–588, 2015.

[22] M. H. Alencar, A. Filho, and A. Almeida, "An MCDM model for potential failure causes ranking from FMECA," in *11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference*, vol. 6, pp. 4536–4544, Helsinki, Finland, 2012.

[23] Y. B. Ju, D. Ju, A. Wang, and M. Ju, "GRP method for multiple attribute group decision making under trapezoidal interval type-2 fuzzy environment," *Journal of Intelligent and Fuzzy System*, vol. 33, no. 6, pp. 3469–3482, 2017.

[24] Y. Geum, Y. Cho, and Y. Park, "A systematic approach for diagnosing service failure: service-specific FMEA and grey relational analysis approach," *Mathematical and Computer Modelling*, vol. 54, no. 11-12, pp. 3126–3142, 2011.

[25] M. G. Dong, S. Y. Li, and H. M. Zhang, "Approaches to group decision making with incomplete information based on power geometric operators and triangular fuzzy AHP," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7846–7857, 2015.

[26] G. Nirmala and G. Uthra, "AHP based on triangular intuitionistic fuzzy number and its application to supplier selection problem," *Materials Today: Proceedings*, vol. 16, pp. 987–993, 2019.

[27] L. Coffey and D. Claudio, "In defense of group fuzzy AHP: a comparison of group fuzzy AHP and group AHP with confidence intervals," *Expert Systems with Applications*, vol. 178, p. 114970, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8856789, 9 pages
https://doi.org/10.1155/2022/8856789

Research Article

An Efficient CNN for Radiogenomic Classification of Low-Grade Gliomas on MRI in a Small Dataset

Jun Liu,[1] **Feng Deng**,[2] **Geng Yuan**,[3] **Changdi Yang**,[3] **Houbing Song**,[4] **and Liang Luo**[5]

[1] *Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA*
[2] *Beijing Key Laboratory of High Dynamic Navigation Technology, Beijing Information Science & Technology University, Beijing, China*
[3] *Department of Electrical & Computer Engineering, College of Engineering, Northeastern University, Boston, MA, USA*
[4] *Security and Optimization for Networked Globe Laboratory (SONG Lab), Embry-Riddle Aeronautical University, Daytona Beach, FL, USA*
[5] *School of Computer Science and Engineering, University of Electronic Science and Technology, Chengdu, China*

Correspondence should be addressed to Liang Luo; luoliang@uestc.edu.cn

Received 1 April 2022; Accepted 28 April 2022; Published 6 June 2022

Academic Editor: Maode Ma

Gliomas, often known as low-grade gliomas, are malignant brain tumors. Codeletion of chromosomal arms 1p/19q has been connected with a good response to treatment in low-grade gliomas (LGG) in several studies. For treatment planning, the ability to anticipate 1p19q status is crucial. This research's purpose is to develop a noninvasive approach based on MR images using our efficient CNNs. While public networks like VGGNet, GoogleNet, and other well-known public networks can use transfer learning to identify brain cancer on MRI, the model contains a large number of components that are unrelated to brain tumors. We build a model from the bottom-up, rather than relying on transfer learning. Our network structure flexibly uses a deep convolution stack mixed with dropout and dense operation, which reduces overfitting and enhances performance. We increase the number of samples by augmenting the dataset. The Gaussian noise is introduced during the model training. To address the issue of data imbalance, we use stratified k-fold cross-validation during training to find the best model. Our proposed model is compared with models fine-tuned through transfer learning, such as MobileNetV2, InceptionResNetV2, and VGG16. Our model achieves better results than these models on the same small dataset. In the test set, when deciding whether or not an image should be 1p/19q codeleted, the proposed architecture achieved an F1-score of 96.50%, precision of 96.50%, recall of 96.49%, and accuracy of 96.50%. By comparing with the transfer model, we found that transfer learning does not outperform CNN on a small dataset.

1. Introduction

Low-grade gliomas (LLG) [1] are brain tumors that arise from astrocytes and oligodendrocytes, which are two separate types of brain cells [1]. Low-grade gliomas can cause a variety of symptoms depending on where they are in the brain. The tumor in the area of the brain that governs language may prevent the patient from speaking or understanding. A brain tumor diagnosis can be devastating for patients. The majority of tumors are discovered as a result of a symptom that prompts doctors to perform a brain MRI or CT scan.

MRI is the most effective method for detecting brain malignancies. The scans provide a massive amount of image data. The radiologist examines these images. Tumors of the brain are difficult to diagnose and treat. The sizes and locations of brain tumors vary dramatically. As a result, fully comprehending the nature of the tumor is quite challenging. For MRI analysis, a qualified neurosurgeon is required. The absence of skilled doctors and a lack of information regarding tumors can make generating reports from MRIs extremely difficult and time-consuming. A manual inspection may be susceptible to errors due to the complexities involved in brain tumors and their characteristics. Machine learning-based automated classification systems have consistently outperformed manual classification.

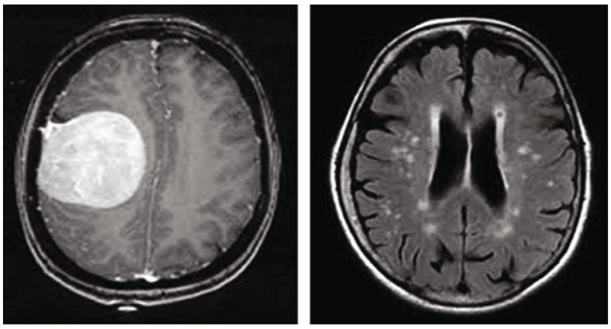
The study of the relationship between cancer imaging features and gene expression is known as radiogenomics. Biomarkers that determine the genetics of a disease without the use of an intrusive biopsy can be created using radiogenomics. A biomarker is a biological indicator of some state or condition. The presence or lack of biomarkers is important in avoiding intrusive biopsies because certain treatments for brain tumors are more successful in the presence or absence of a biomarker. The detection of biomarkers can ensure that patients receive the most effective treatment for their specific situation [2].

Low-grade gliomas (LLG) [2–4] are tumors that are considered formed from glial cells, have infiltrative development, and lack malignant histopathological characteristics. One of the biomarkers that appear to be essential in low-grade gliomas is 1p/19q chromosomal codeletion. When 1p/19q codeletion is discovered in low-grade gliomas, studies demonstrate that they respond better to chemotherapy and radiotherapy. The novelty and promising results of combining deep learning with radiogenomics are what make this study noteworthy. The detection of 1p/19q codeletion using deep learning works better with T2 images than with T1 postcontrast images [2].

In 2017, deep learning was firstly used by Akkus et al. [2] to predict 1p19q from LGG MRI; tumor segmentation, image registration, and CNN-based 1p/19q status classification are the three primary steps of their method. When data augmentation is not performed, their multiscale CNNs overfit the original training data. Lombardi et al. [4] used popular public networks, including AlexNet, VGG19, and GoogleNet, for 1p19q categorization through transfer learning [5–7]. According to their description, even with limited datasets, the results offered by transfer learning are robust. Abiwinanda et al. [8] used five different CNN designs, with the second design with two convolutional layers, one maxpool layer, and one ReLU layer, then come 64 hidden neurons, achieving the highest accuracy.

Why are there just thousands of training examples? Maithra Raghu et al. [9] wondered. They looked upon transfer learning in small data settings. They discovered that there was a significant performance difference between transfer learning and training from scratch for a big model (ResNet), but not for a smaller model. For a little amount of data, the large model built by ImageNet can have too many parameters. They discovered that transfer learning provides limited performance increases for the evaluated medical imaging tasks after a rigorous performance evaluation and examination of hidden representations of neural networks. Transfer learning had little effect on the performance of medical imaging tasks, and the model trained from the ground up was near as well as the ImageNet transfer model.

The following are our main contributions:



Yes No

Figure 1: Brain MRI.

(i) Using the 3×3 convolution and LeakyReLU, we create a dedicated convolutional neural network for detecting brain tumors on MRI images

(ii) During training, we use a customizable combination of dropout and Gaussian noise to reduce overfitting and increase performance

(iii) Stratified k-fold is used to correct problems in training induced by data imbalance

(iv) Our proposed model is compared to MobileNetV2, InceptionResNetV2, and VGG16 that have been fine-tuned through transfer learning

2. Materials and Methods

We use the provided dataset to train our planned network. Meanwhile, on the same dataset, we compare the performance of MobileNetV2, InceptionResNetV2, VGG16, etc., which were fine-tuned using transfer learning approaches.

2.1. Experimental Data. The Kaggle small brain tumor dataset [10] provided the brain MRI dataset that was utilized to evaluate the planned study. The dataset contains 253 brain MRI images in two folders: yes and no. There are 155 tumorous brain MRI images in folder yes, and there are 98 nontumorous brain MRI images in folder no.

Figure 1(a) is a brain with a tumor, and Figure 1(b) is a brain tumor.

2.2. Network Architectures. In Figure 1, there are 14 layers in the model. Convolutional kernels with smaller convolutions—3×3—were found to produce positive outcomes, as these smaller convolutions may capture some of the finer characteristics of the edges. This network's convolutional layers all employ 3×3 kernels. It is starting with 16 kernels per layer; the architecture progresses to 32 kernels per layer, 64 kernels per layer, and finally 128 kernels per layer.

This network depicted in Figure 2 is made up of the convolution layer, pooling layer, dropout layer, LeakyReLU layer, dense layer [11], flatten layer, and softmax layer, with the input picture.

Table 1 specifies the network activities utilized by each layer, as well as the size of the convolution kernel and the size of the input.

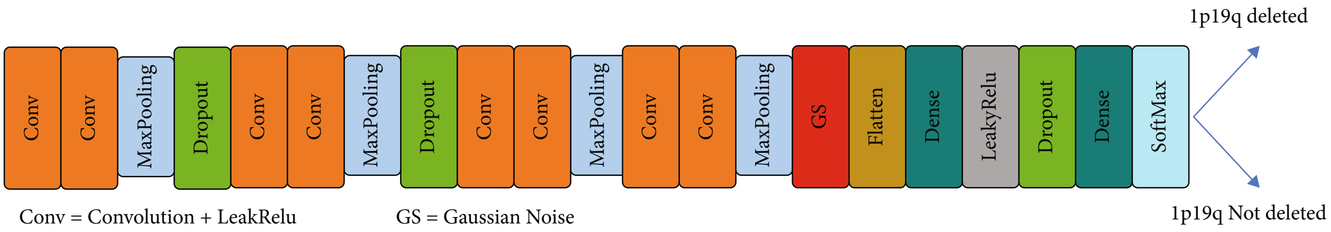


Figure 2: Network architecture.

Table 1: Network layer.

Type	Filter shape	Input size
Input layer	N/A	256 × 256 × 1
Conv2D	16 × 3 × 3	254 × 254 × 16
LeakyReLU	N/A	254 × 254 × 16
Conv2D	16 × 3 × 3	252 × 252 × 16
LeakyReLU	N/A	252 × 252 × 16
Maxpooling	2 × 2	126 × 126 × 16
Dropout	N/A	126 × 126 × 16
Conv2D	32 × 3 × 3	124 × 124 × 32
LeakyReLU	N/A	124 × 124 × 32
Conv2D	32 × 3 × 3	122 × 122 × 32
LeakyReLU	N/A	122 × 122 × 32
Maxpooling	2 × 2	61 × 61 × 32
Dropout	N/A	61 × 61 × 32
Conv2D	64 × 3 × 3	59 × 59 × 64
LeakyReLU	N/A	59 × 59 × 64
Conv2D	64 × 3 × 3	57 × 57 × 64
LeakyReLU	N/A	57 × 57 × 64
Maxpooling	2 × 2	28 × 28 × 64
Conv2D	128 × 3 × 3	26 × 26 × 128
LeakyReLU	N/A	26 × 26 × 128
Conv2D	128 × 3 × 3	24 × 24 × 128
LeakyReLU	N/A	24 × 24 × 128
Maxpooling	5 × 5	4 × 4 × 128
Gaussian noise	N/A	4 × 4 × 128
Flatten	N/A	2048
Dense	N/A	1024
LeakyReLU	N/A	1024
Dropout	N/A	1024
Dense	N/A	2
Softmax	N/A	2

(i) All images of brain tumors that are fed into the network are scaled to 256 × 256

(ii) This network uses 3 × 3 kernels for all convolutional layers. The architecture starts with 16 kernels per layer, then 32, 64, and finally 128 kernels per layer

In Figure 3, we incorporate more hidden layers and therefore more nonlinear functions, enhancing the decision function capabilities and introducing fewer parameters, inspired by a VGGNet stack of three 3 × 3 convolutional layers, instead of a single 7 × 7 layer.

However, our network differs from the VGGNet structure, which is made up of a stack of 3 × 3 convolutions and ReLUs. Our network is made up of 3 × 3 convolutions and LeakyReLUs.

(i) Because negative values are kept and saturation concerns are avoided when employing tanh, LeakyReLU was chosen as the activation function

(ii) This model employs 2 × 2 maxpooling at first, then 7 × 7 maxpooling later. The number of neurons in a layer is reduced when it is dense

(iii) Dense [11] (fully connected) is used twice just before the softmax layer to reduce the number of neurons to two, reflecting the binary prediction of either “codeletion” or “no codeletion”

(iv) Gaussian noise was purposely supplied to the training data to minimize overfitting, which can think of it as a form of random data augmentation. For corrosion processes with genuine inputs, Gaussian noise (GS) is a natural choice. During training, the error is reduced, and at the same time, the interference items generated by noise are penalized to achieve the purpose of reducing the square of the weight. The noise distribution’s standard deviation was set to 0.5

The following formula can be used to compute the probability density of a Gaussian distribution:

$$F(x\,;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{(x-\mu)^{2}}{2\sigma^{2}}\right)dx. \qquad (1)$$

Assume we introduce Gaussian noise to the inputs in Figure 4. Before moving on to the next layer, the squared weight amplifies the noise variation. The squared error increases as a result of this. When the input is noisy, minimizing the squared error tends to minimize the square of the weights. Let us assume y^{noisy} is output by GS at one time:

$$y^{\text{noisy}}=\Sigma_{i}w_{i}x_{i}+\Sigma_{i}w_{i}\varepsilon_{i}, \qquad (2)$$

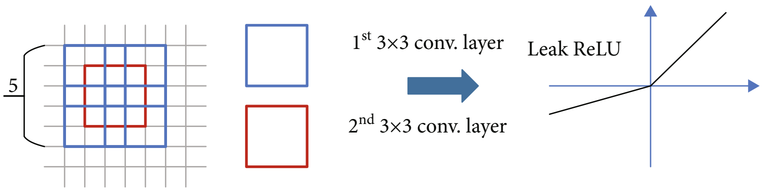


Figure 3: 3 × 3 convolutions and LeakyReLU.

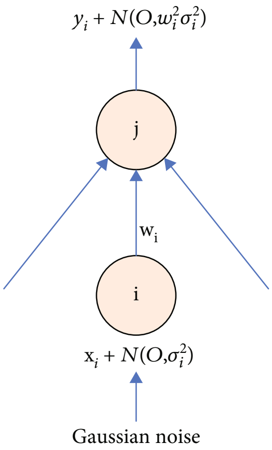


Figure 4: Introduce Gaussian noise to the inputs.

where ε_i is sampled from $N(0,\sigma_i^2)$:

$$\begin{aligned} E\left[\left(y^{\text{noisy}} - t\right)^2\right] &= E\left[\left(y + \Sigma_i w_i \varepsilon_i - t\right)^2\right] \\ &= E\left[\left((y - t) + \Sigma_i w_i \varepsilon_i\right)^2\right] \\ &= (y - t)^2 + E[2(y - t)\Sigma_i w_i \varepsilon_i] + E\left[\left(\Sigma_i w_i \varepsilon_i\right)^2\right] \\ &= (y - t)^2 + E\left[\Sigma_i w_i^2 \varepsilon_i^2\right] \\ &= (y - t)^2 + \Sigma_i w_i^2 \sigma_i^2. \end{aligned} \tag{3}$$

Because ε_i is independent of ε_i and ε_i is independent of $(y - t)$, σ_i^2 is equivalent to L2 penalty [12]. The error is minimized during the training process, and the noise-induced interference items are penalized in order to reduce the square of the weight and achieve a comparable result to L2 regularization.

(i) The goal of flatten is to one-dimensionalize the multidimensional input, which is accomplished by transitioning from convolutional to fully connected layers

(ii) During the forward and backward propagation phases, dropout avoids neurons at random. The number of neurons that are not updated is determined by the dropout value. The dropout rate was set to 0.3

(iii) The probability of each of the binary outcomes—“codeletion” and “no codeletion”—is included in the output layer using a softmax classifier

2.3. Hyperparameters. Several hyperparameter values were explored in this study.

2.3.1. Learning Rate. The learning rate is the amount of time we spend moving in a particular direction to find the global minima. Starting with a greater learning rate usually works fine because the initial weight values are rather random. We typically grow closer and closer to either the global or local minima as we proceed through the training phase. Because we do not want to overshoot the minima, annealing, the learning rate is a typical method. To put it in another way, as the training phase progresses, we begin to take smaller and smaller moves in a specific direction. When there is no change in the loss value, we will continue to reduce the learning rate by the square root of 0.1 until it reaches a reduction of $0.5e - 6$.

2.3.2. Early Stopping. Overfitting models to training data can be prevented or limited by early stopping techniques. Moreover, when the findings are static, early halting procedures prevent needless computations. If there has not been a change of at least 0.001 in 10 (epochs), the model provided here will terminate training. We usually start with smaller weights when we begin the network. The partial network weights may grow in size as the training time increases. We can limit the network’s capabilities to a specific range by stopping training at the appropriate time. The steps are as follows:

(i) The validation set is used to collect test results after every 5 epochs. Stop training if the test error on the validation set increases as the epoch increases

(ii) After stopping, use the weights as the network’s final parameters

2.3.3. Batch Size. The batch size specifies how many photos are handled during forward propagation to produce a loss value for backpropagation. Batch size is typically set to a power of two and is restricted by the available memory. Furthermore, while a bigger batch size allows for faster training, weights update less frequently and may not deliver the greatest outcomes. The batch size was set to 16.

2.3.4. Number of Epochs. When training your model, the number of epochs denotes the number of times the complete training dataset is iterated over. Validation determines how well the model generalizes to new data at the end of each epoch. The number of epochs was set to 32.

2.4. Stratified k-Fold Cross-Validation. k-fold cross-validation is an excellent technique to examine the bias-variance tradeoff and guarantee that the model has low bias and variance. The following is a summary of the testing procedure:

(a) The data from the original is split into k sections at random by unrepeat sampling

(b) Each time, as the test set, one of these is chosen. The other $k-1$, on the other hand, is employed as a model training set

(c) Repeat the second step k times more to give a probability for each subgroup and the rest to be the training set

(d) After each training set, obtain a model

(e) Use this model to test on the relevant test set, as well as to calculate and save the model's evaluation metrics

(f) Using the current k-fold cross-validation procedure, calculate the average of all the test results from the k sets as a measure of the model's accuracy and performance

Choosing a good "k" number ensures that the testing procedure provides the most accurate assessment of the performance of our model. When increasing the number of splits k, the variance increases while the bias decreases. Lowering k, on the other hand, increases the bias while decreasing the variance. The tradeoff between bias and variance is a difficult problem.

The predictors X and response Y can both be written as variables:

$$Y = f(X) + \epsilon \sim \mathcal{N}(0, \sigma_{\epsilon}). \tag{4}$$

The quadratic error's expected value can then be represented as

$$\mathrm{SE}(x) = E\left[(Y - f(x))^2\right]. \tag{5}$$

After some arithmetic, we get

$$\mathrm{SE}(x) = (E[f(x)] - f(x))^2 + E\left[(f(x) - E[f(x)])^2\right] + \sigma_e^2. \tag{6}$$

In the formula above, $(E[f(x)] - f(x))^2$ is bias^2, $E[(f(x) - E[f(x)])^2$ is variance, and σ_e^2 is irreducible error.

The square of a random variable X should have the following expected value:

$$\begin{aligned} E\left[X^2\right] &= \mathrm{Var}[X] + E[X]^2, \\ E[y] &= E[f + \epsilon] = E[f] = f. \end{aligned} \tag{7}$$

This is how it works:

$$\begin{aligned} E\left[(y-f)^2\right] &= E\left[y^2 + f^2 - 2yf\right] \\ &= E\left[y^2\right] + E\left[f^2\right] - E[2yf] \\ &= \mathrm{Var}[y] + E[y]^2 + \mathrm{Var}[f] + E[f]^2 - 2fE[f] \\ &= \mathrm{Var}[y] + Var[f] + (f - E[f])^2 \\ &= \mathrm{Var}[y] + \mathrm{Var}[f] + E[f - f]^2 \\ &= \sigma^2 + Var[f] + \mathrm{Bias}[f]^2. \end{aligned} \tag{8}$$

The relationship between mistake and the bias-variance tradeoff is depicted in Figure 5. The best model is one that minimizes both bias and variance at the same time, resulting in the lowest error rate. The vertical dashed line represents a model with just the correct level of complexity. This model will have high accuracy ratings on both train and test data, indicating that it is generalizable. We hope that the model's bias and variance are very low, but this is not always possible. We must weigh the benefits and drawbacks and strike a balance. In practice, we use $k = 3$. The most significant benefit of k-fold cross-validation is that all data is used in training and prediction, thereby avoiding overfitting and accurately reflecting the concept of crossover.

Because of the disparity in the number of photos of brain tumors versus healthy brains in the dataset, if we use k-fold on an unbalanced dataset, we may end up with no or very few minority classes in our training data. We utilize stratified k-fold to avoid this issue. Stratified k-fold is a k-fold variant that produces hierarchical folds: each set has nearly the same percentage of samples for each target class as the entire set. If the dataset is divided into four categories and the ratio is $2:3:3:2$, the divided sample ratio is approximately $2:3:3:2$.

2.5. Experiments. For the experiments, we use TensorFlow as the backend Keras Python package on an Ubuntu 18.04 X86_64 server. One NVIDIA 2080ti GPU is used.

2.5.1. Data Preparation. Data preparation steps are included deleting a third class, standardizing the data, and implementing cross-validation [12], to shuffle the training data. Because this is a small dataset, there were insufficient examples to train the neural network. In addition, data augmentation was useful in addressing the data imbalance issue.

The image is preprocessed before being processed into the proposed structure. The original MR image is scaled to 225×2251 pixels in the first step. Image augmentation techniques such as flipping, mirroring, and rotating are used to generate redundant data for the network, which is frequently used to avoid network overfitting and improve system resilience.

ImageDataGenerator is a Keras class that describes the image data preparation and augmentation setup. We can rotate the image at any angle between 0 and 360 degrees using the ImageDataGenerator class. For flipping along the vertical or horizontal axis, the ImageDataGenerator class has options for horizontal flip and vertical flip. The key advantage of utilizing the Keras ImageDataGenerator class

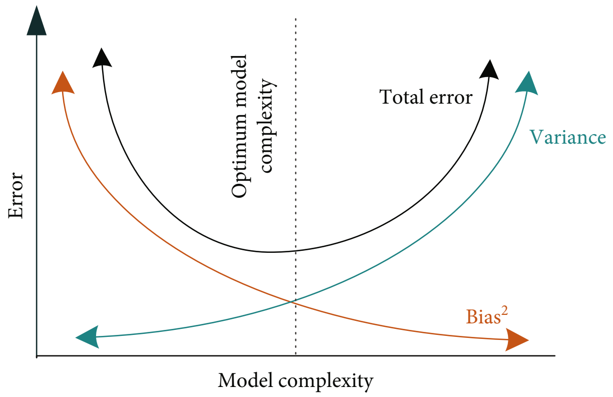


Figure 5: Bias and variance relationship.

is that it is intended for real-time data enhancement. The model generates augmented images on the fly while it is still being trained.

2.5.2. Proposed Workflow. We build, evaluate, and train our model to improve performance and use stratified k-fold cross-validation in model training as depicted in Figure 6.

(i) We divided the data into training and testing datasets at random and built the model using the training set and estimated its accuracy using the test set

(ii) Then, we acquire the best quality model by fine-tuning the model via 3-fold cross-validation to enhance the estimate's accuracy

(iii) On the test set, evaluate the model's expected accuracy

(iv) Output evaluation statistics include precision, recall, F1-score, and confusion matrix

2.5.3. Evaluation Method

(1) Confusion Matrix. A confusion matrix is a technique for determining whether or not a classification method is effective. If your dataset has more than two classes or an uneven number of observations in each, classification accuracy alone can be deceiving.

In Table 2, we can clearly see the number of correct identifications and the number of incorrect identifications for each category.

(2) Precision. The model properly predicted the percentage of patients with 1p/19q codeletion based on the total number of patients with 1p/19q codeletion referred to as precision. It has the following formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (9)$$

(3) Recall. The fraction of 1p/19q codeleted patients recognized by the model is divided by the total number of 1p/19q codeleted and 1p/19q nondeleted patients to compute recall. It has the following formula:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (10)$$

(4) F1-Score. The purpose of the F1-score was to combine precision and recall measurements into a single value. It is an important metric for class imbalance problems; due to an imbalance in the number of brain and nonbrain tumors in this brain MRI dataset, the F1-score was created to operate effectively with data that is unbalanced. It has the following formula:

$$\text{F1-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \quad (11)$$

3. Results

The harmonic mean of precision and recall is calculated using the F1-score. In relation to all other classes, the scores for each class indicate how accurate the classifier was in classifying the data points in that class. The number of samples of the real answers that fall into that group is the support.

We train the model using stratified 3-fold cross-validation. F1-score, precision, and recall are all factors that must be considered. Table 3 demonstrates that the model achieves good values.

In the test set, we employed 171 photos, 86 of which are 1p19q deleted and 85 of which are 1p19q not deleted; the suggested architecture received an F1-score of 0.9650 in Table 4.

Figure 7 shows the confusion matrix for the classification of 1p19q status on the test set. We can be certain that all 125 1p19q deleted pictures were detected accurately.

We compare pretrained MobileNetV2 [13], Inception-ResNetV2 [14], VGG16 [15], etc., which fine-tuned using the transfer learning approach and other approaches.

Table 5 demonstrates that for classification on small datasets, transfer learning is not superior to ordinary CNN. This is due to the insufficient number of training samples in small datasets to learn complex sets of deep feature sets. With reasonable design, CNNs without transfer learning can attain and surpass transfer learning. Our method yields the best outcomes. Simultaneously, we examine the indicators listed in the above table and discover that the deep learning approach outperforms the machine learning SVM method by a wide margin.

4. Discussion

We provide a reliable and noninvasive approach for predicting 1p/19q chromosomal arm deletion in this work. Having a sufficient amount of datasets is a significant difficulty when applying deep learning approaches to medical imaging. Despite the fact that the initial data amount was limited, our data volume expanded as a result of data augmentation approaches. With larger patient populations and more

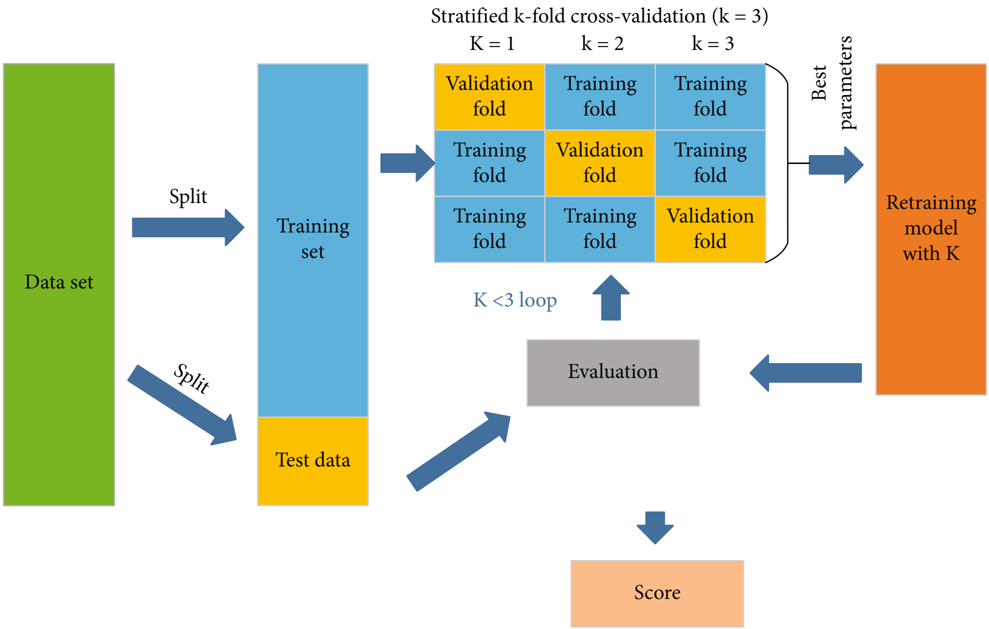


Figure 6: Proposed workflow.

Table 2: Confusion matrix.

	1p19q deleted	1p19q not deleted
1p19q deleted	TP	FP
1p19q not deleted	FN	TN

TP = true positive; FN = false negative; FP = false positive; TN = true negative.

Table 3: Classification performance.

Fold	Train/test	Precision	Recall	F1-score	Support
Fold-1	Train	0.9850	0.9562	0.9704	274
	Test	0.9517	0.9841	0.9598	477
Fold-2	Train	0.9881	0.9679	0.9241	274
	Test	0.9851	0.9635	0.9742	477
Fold-3	Train	0.9886	0.9526	0.9703	274
	Test	0.9517	0.9841	0.9598	477

Table 4: Statistics for test set.

	Precision	Recall	F1-score	Support
1p19q deleted	0.9761	0.9535	0.960	86
1p19q not deleted	0.9540	0.9764	0.970	85
Avg/total	0.9650	0.9649	0.9650	171

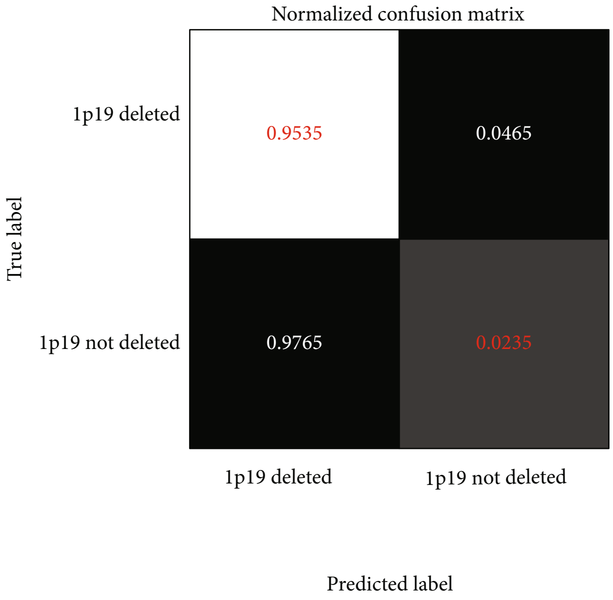


Figure 7: Confusion matrix.

varied data, it is possible that additional performance gains will be gained.

As large convolution kernels are inefficient in terms of cost. We are reducing the number of irrelevant features conceivable by restricting the number of parameters. This drives the deep learning algorithm to learn traits that are common to a variety of scenarios, allowing it to generalize more effectively. Smaller odd-sized kernel filters would be preferable. However, 1 × 1 is removed from the list of possible ideal filter sizes since the features recovered would be fine-grained and local, with no information from nearby pixels. Furthermore, it does not extract any useful features. Through experiments, we found that although VGG16 also uses a 3 × 3 convolution kernel, it is prone to overfitting due to the complexity of the network, and the dataset is small. As a result, VGG16 categorization precision and recall of 1p/19q chromosomal arm deletion are not very good.

Table 5: Performance comparison.

Model	Precision	Recall	F1-score	Accuracy
Ours	0.9650	0.9649	0.9650	0.9650
MobileNetV2 [14]	1.0	0.8709	0.9200	0.9200
InceptionV3 [15]	0.923076	1	0.9600	0.9428
InceptionResNetV2 [14]	0.90625	0.9354	0.9153	0.90
AlexNet [16]	0.93620	0.95650	0.9462	0.9483
VGG16 [16]	0.89660	0.9286	0.9123	0.9138
Shwetha and Madhavi [17]	0.89	0.92	0.90	0.88
DenseNet-169 [10]	—	—	—	0.9412
DenseNet-SVM [10]	—	—	—	0.9412
CNNs [18]	—	—	—	0.9500
SVM [19]	0.70	0.71	0.70	0.71

Because of the deep architecture of current networks like GoogleNet and ResNet, feature maps from these networks frequently have a very large receptive field. However, studies [20] reveal that the network gathers information from a considerably narrower portion of the receptive field, which is referred to as the valid receptive field in this research. In this experiment, we found that the recall rate was not high by using InceptionResNetV2 and VGG16. As a result, a large receptive field does not increase the performance of medical images on small datasets considerably.

We discovered that MobileNetV2 is significantly higher than InceptionResNetV2 and VGG16 in the fields of precision. It employs depth-wise separable convolutions and divides an ordinary 3 × 3 convolution into two convolutions, which is the same as the 3 × 3 convolution we employ. It makes use of ReLU6. ReLU6 is a standard ReLU with a maximum output limit of 6, allowing for high numerical resolution even when the mobile device's float16/int8 accuracy is low. However, ReLU6 is not as accurate to the server as the LeakyReLU we used.

The model's capacity to learn mapping rules from the input space can be increased by adding Gaussian noise during training, as can the model's generalization ability and fault tolerance. Because the training samples change frequently, adding noise to the network can lead it to lose track of them, resulting in smaller network weights and a more robust network while lowering the generalization error. Since new samples are selected from the domain adjacent to known samples, the structure of the input space is smoothed. This smoothing may make the learning mapping function easier for the network, leading to better and faster learning. After adding Gaussian noise to our model training, we can see significant improvement in performance.

Since medical imaging data is scarce, transfer learning approaches are used to fine-tune medical imaging models using popular public models (e.g., VGGNet and GoogleNet) generated from large public ImageNet datasets. However, these models create a large number of characteristics that are unrelated to medical imaging, jeopardizing the accuracy of medical diagnosis [21]. Our model does not involve transfer learning, and the parameters it generates are specific to the medical imaging dataset that was used. As a result, the reliability of brain tumor diagnosis has substantially improved. Simultaneously, we discover that our method beats transfer learning on small datasets but that transfer learning performs better on large datasets.

5. Conclusion

The results of our CNN approach for 1p/19q codeletion status classification noninvasively are promising. We create a brain tumor detection model that does not rely on transfer learning. Our network structure employs a deep convolution stack strategy when training with Gaussian noise, reducing overfitting and improving performance. Compared to transfer learning models, our model gives more accurate findings. With basic, lightweight models equivalent to ImageNet topology, we discovered that transfer learning offered no performance benefit in small datasets. By properly designing the network and optimizing the hyperparameters during training, CNNs without transfer learning can reach and surpass transfer learning.

Data Availability

This study makes use of datasets that are freely available to the public. This dataset can be found at the following link: https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection.

Conflicts of Interest

There are no conflicts of interest declared by the authors.

Acknowledgments

The National Natural Science Foundation of China (62176048), The National Natural Science Foundation of China (41871348), Beijing Information Science & Technology University (2020KYNH224), and Beijing Key Laboratory of High Dynamic Navigation Technology (HDN2019002) all contributed to this work.

References

[1] J. Amin, M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, "Brain tumor detection and classification using machine learning: a comprehensive survey," *Complex & Intelligent Systems*, 2021.

[2] Z. Akkus, I. Ali, J. Sedlář et al., "Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 469–476, 2017.

[3] D. Bhattacharya, N. Sinha, and J. Saini, "Determining chromosomal arms 1p/19q co-deletion status in low graded glioma by cross correlation-periodogram pattern analysis," *Scientific Reports*, vol. 11, no. 1, p. 23866, 2021.

[4] G. Lombardi, V. Barresi, A. Castellano et al., "Clinical management of diffuse low-grade gliomas," *Cancers*, vol. 12, no. 10, article 3008, 2020.

[5] R. Chelghoum, A. Ikhlef, A. Hameurlaine, and S. Jacquir, "Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images," in *IFIP Advances in Information and Communication Technology*, Springer, 2020.

[6] M. F. Alanazi, M. U. Ali, S. J. Hussain et al., "Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model," *Sensors*, vol. 22, no. 1, p. 372, 2022.

[7] Z. N. K. Swati, Q. Zhao, M. Kabir et al., "Brain tumor classification for MR images using transfer learning and fine- tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, 2019.

[8] N. Abiwinanda, M. Hanif, S. T. Hesaputra, A. Handayani, and T. R. Mengko, *Brain Tumor Classification Using Convolutional Neural Network World Congress on Medical Physics and Biomedical Engineering 2018*, Springer, Singapore, 2019.

[9] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: understanding transfer learning for medical imaging," *Advances in neural information processing systems*, vol. 32, 2019.

[10] J. Kang, Z. Ullah, and J. Gwak, "MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, article 2222, 2021.

[11] *Keras API Reference / Layers API / Core Layers / Dense Layer*-March 2022, https://keras.io/api/layers/core_layers/dense/.

[12] A. Panigrahi and M. R. Patra, "Chapter 6 - network intrusion detection model based on fuzzy-rough classifiers," in *Handbook of Neural Computation*, P. Samui, S. Sekhar, and V. E. Balas, Eds., pp. 109–125, Academic Press, 2017.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "Mobilenetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, Salt Lake City, UT, USA, 2018.

[14] F. Taghiyev, *Brain Tumor Detection Using TensorFlow 2.x*, 2022, March 2022, https://www.kaggle.com/code/faridtaghiyev/brain-tumor-detection-using-tensorflow-2-x/notebook.

[15] M. F. I. Soumik, *Brain Tumor Detection InceptionV3*, 2022, March 2022, https://www.kaggle.com/code/mdfarhanisraksoumik/brain-tumor-detection-inceptionv3-auc-99-84.

[16] S. Kuraparthi, M. K. Reddy, C. N. Sujatha et al., "Brain tumor classification of MRI images using deep convolutional neural network," *Traitement du Signal*, vol. 38, no. 4, pp. 1171–1179, 2021.

[17] V. Shwetha and C. H. R. Madhavi, "Classification of brain tumors using hybridized convolutional neural network in brain MRI images," *International Journal of Circuits, Systems and Signal Processing*, vol. 16, pp. 561–570, 2022.

[18] A. Sinha, *Brain Tumour Detection with CNN 96% Accuracy*, 2022, March 2022, https://www.kaggle.com/code/ethernext/brain-tumour-detection-with-cnn-96-accuracy/notebook.

[19] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters ——improve semantic segmentation by global convolutional network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4353–4361, Honolulu, Hawaii, USA, 2017.

[20] Brendon Im, 2022, March 2022, https://www.kaggle.com/code/brendonim/brain-mri-tumor-detection-using-svm.

[21] J. Liu, F. Deng, G. Yuan, X. Lin, H. Song, and Y. Wang, "An explainable convolutional neural networks for automatic segmentation of the left ventricle in cardiac MRI," in *Proceedings of the CECNet 2021*, Beijing, China, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5986295, 11 pages
https://doi.org/10.1155/2022/5986295

Research Article

Research on Efficiency in Credit Risk Prediction Using Logistic-SBM Model

Dongmei Li and Liping Li

School of Management, Shanghai University, 333 Nanchen Road, Baoshan District, Shanghai 200444, China

Correspondence should be addressed to Liping Li; liliping@shu.edu.cn

Received 1 March 2022; Accepted 12 April 2022; Published 3 June 2022

Academic Editor: Maode Ma

Network lending, an innovative financial lending product, is separated from traditional financial media and implemented on the Internet platform. We study the credit risk prediction of online loan based on risk efficiency analysis. Moreover, we put forward the concept of borrower risk efficiency and apply it to risk prediction. The main task of this study is to establish risk efficiency characteristics on the basis of referring to various risk characteristics and carry out risk prediction after passing the screening of a series of features. The framework is realized by combining logistic regression and slack-based measure (SBM), and feature selection and verification are carried out through machine learning and statistics. Firstly, the efficiency risk characteristics are extracted and the risk efficiency is calculated by MaxDEA. Secondly, the features are screened and verified by Python. Then, the efficiency value obtained by SBM method is used as a new index for the training and testing of logistic model together with the initial related indexes. Moreover, in order to prove the effectiveness of the proposed credit risk prediction control scheme based on risk efficiency, the research compares the prediction before and after adding the risk efficiency feature. The simulation results demonstrated that the logistic-SBM model is more suitable for credit risk prediction than the commonly used logistic method, which realized the efficient prediction of credit risk based on the logistic-SBM model. Finally, some suggestions are put forward to China's regulatory authorities and the platform itself to control the credit risk of Internet lending industry.

1. Introduction

In "Interim Measures for the Management of Business Activities of Online Lending Information Intermediaries" promulgated in 2016, online lending is defined as direct lending between individuals including natural persons, legal persons, and other organizations through the Internet platform. Internet finance peer-to-peer (P2P) network finance is a branch of Internet finance, which is the product of the combination of Internet and finance. The academic definition of Internet finance has something in common with Internet finance, which is a new financial business model for traditional financial institutions and Internet enterprises to achieve financing. Davis and Gelpern and Slattery believe that P2P online lending has injected fresh vitality into the traditional lending market to meet the needs of investors and consumers [1, 2]. Financial technology based on P2P is one of the new breakthroughs in financial service institutions [3]. The main business models of Internet finance include Internet payment, online lending, equity crowd funding, Internet fund sales, Internet insurance, Internet trust, and internet consumer finance. Lenders have a greater impact on borrowers than do borrowers on lenders [4]. As technologies of big data and block chain advance, the financial credit risk in the context of the Internet has become a popular research subject [5]. P2P online lending originated in foreign countries. The earliest P2P online lending platform in the world is Zopa in the UK, which was established in London in March 2005. The new financial industry represented by peer-to-peer lending has gradually become a new source of volatility due to the increasing complexity of the Chinese financial market [6]. In 2007, China established its first P2P network lending enterprise. P2P lending platforms have different backgrounds and transparency [7]. Platform background is related to operational risk [8]. The embryonic period of the development of P2P online lending financial enterprises is from 2007 to 2012. From 2013 to 2015, the development of P2P online lending financial enterprises has entered a period of vigorous expansion. From 2017 to now, it is a period of consolidation and standardization of P2P online lending financial enterprises. There are more than 10000 P2P online lending financial enterprises, in which more than 5000 were operated at the same time. The annual transaction scale is about 3 trillion yuan, and the bad debt loss rate is very high. Through continuous rectification, the People's Bank of China issued the "fintech development plan (2019-2021)" in September 2019, proposing to "further enhance the technology application ability of the financial industry and realize the deep integration and coordinated development of Finance and technology." By the beginning of 2020, there are already a lot fewer P2P online lending institutions in operation.

In China, the scope of definition of online lending includes both individual-to-individual lending, individual-to-business lending, and corporate organization-to-business organization lending. Since the birth of the first P2P in China in 2007, online lending has developed rapidly. To a certain extent, it is not only the result of the continuous advancement of modern information technology but also the inevitable product of the diversification of lending needs. However, the problems exposed have become more prominent during the development. Investors should pay attention to information asymmetry and credit risk impact [9]. Therefore, the problems of online lending industry in China have not only the common problems of other countries' online lending but also the specific problems of our country. Internet financial risk is not only directly related to the operation and development of the Internet financial system itself but has also a very important impact on the country's macroeconomic operation because of its rapid development speed and growing scale of development [10].

2. Literature Review

Since 2013, innovative Internet financial services such as Yirendai, Crowdfunder, and Renrendai have been born in China, promoting the reform of financial service models and accelerating financial marketization. Although there are a large number of online lending investors, they basically lack professional lending knowledge [11]. Moreover, the amount of online lending is small. When the lender lacks the effective information of the borrower, the bidding will often follow suit blindly and other irrational behaviors, which will inevitably increase the credit risk of online lending [12, 13]. However, the risk of the industry has also become obvious. The theory of information asymmetry was first put forward by Akerlof (1970) [14] by observing the phenomenon of used car market. In online lending, information asymmetry can also lead to the possibility of borrowers' default [15, 16]. The imbalance of these factors will lead to the platform's resources, and opportunities cannot play a role, resulting in the collapse of the platform [17]. The survival of the platform depends on the age, scale, and life cycle of the enterprise [18]. The management ability of platform operators plays a key role in the success or failure of small and micro platforms [19, 20]. For instance, in February 2017, 55 problematic platforms were involved in illegal fundraising, difficulty with cash withdrawal, fraud, absconding with money, and loss of connection and other risky breaches. Recent years have seen the rapid development of Internet finance in China, and various peer-to-peer (P2P) lending platforms have been released [8]. There is diversity of default behaviors of borrowers with different credit grades in online P2P loan market [21]. Reputation plays an important role in the long-term development of P2P lending platform [22]. These negative news have greatly affected investors' investment confidence and have had a very bad impact on the social reputation of the entire industry. Therefore, it is particularly important to scientifically evaluate Internet financial risks. The issue of risk and regulation of P2P lending platform in China is taken seriously. The P2P industry has promulgated the regulation that online loan platform must be online for fund deposit business, which makes bank deposit gradually normalized [23]. The difference between P2P online loan and traditional financial institutions lies in the transaction system of P2P online loan, which adopts the interest rate auction system when the transaction is concluded. Herzenstein and Barasinska [24] studied the interest rate of the American prosper online lending platform in 2011 and 2014, respectively. They found that the borrowers would set the maximum interest rate they were willing to pay for borrowing the funds, and then, the investors would decide whether to borrow according to the loan information provided by the prosper online lending platform. This innovative financial lending model provides investors with a new way of financial management. Liu et al. mainly find that investors' herd behavior exists significantly [25]. P2P mode can make the idle funds of investors not only increase in value but also meet the borrower's demand for funds to increase a loan channel. In this lending mode, the lending process no longer depends on offline financial institutions but relies on the network lending platform to match the needs of both sides and to realize the transaction. The reasons for choosing logistic-SBM model are as follows: DEA can be used to explore the new intersecting fields including management science, mathematics, mathematical economics, and operations research. DEA uses multiple inputs and outputs to measure the relative efficiency of each DMU. In the process of risk management for borrowers of Internet financial loan products, the DEA method can take each borrower as each DMU to obtain its efficiency value, rather than just studying the traditional indicators of the borrower. At present, there are few researches on the real customer credit data in China. Therefore, this study selected the logistic regression method for big data analysis through the comparison of different mathematical model methods. In this study, according to the characteristics of the source data, data envelopment analysis was used to process the source data and then, the data was trained in the logistic regression model to improve the accuracy of the model prediction. This method not only provides an innovative method to study the credit risk analysis of

Internet Financial borrowing customers but also expands the research space in this field, which has both theoretical and practical significance. Based on the present situation of the P2P lending platform development in our country, its development in the process of credit risk, transaction risk, legal risk, and so on is analyzed. In addition, corresponding regulatory measures were put forward to strengthen the development of P2P lending platform in China, which is greatly important.

3. Methods

The notion of probability is very closely related to the notion of symmetry [26]. Credit risk prediction is essential to predict the probability of default of borrowers. The specific research methods are as follows.

First is data preparation. This study divides the credit data of Internet financial technology companies into a sample set and a test set.

Then, SBM-DEA model was established. According to the above five indicators, the efficiency value of each customer was obtained by using DEA model through MaxDEA software. DEA_score was added to the next dataset.

The third step is feature processing. The feature processing methods include feature binning, correlation coefficient, IV, and random forest model.

The fourth step was to test the logistic-SBM model. The prediction results of the model are observed directly through the mixed matrix diagram. The AUC value of the model was calculated and tested. The model was tested by the K-S test.

In the last step, we compared the values of corresponding evaluation measures of two models.

The logistic-SBM model was established through MaxDEA and Python software.

3.1. Data Source Preparation. We used the real credit data of an Internet financial technology company as the analysis object. The company is mainly engaged in small loans, online finance, and other Internet financial products. The platform has a variety of data sources, high data quality, and rich data information. The loan customer risk management model to be studied in this paper selected the loan records of the platform. The sample population data was sampled and divided into a sample set and test set.

For the data selection, the loan with the end of repayment and the loan with default were selected for modelling. The target variable was selected according to the user's "repayment status" characteristics. If the loan has been repaid in which the default has not occurred, the value is 0. If there is overdue loan in which default occurs, the value is 1. Finally, 14028 transaction data that have been paid off were selected as the sample set, among which 10237 cases have been successfully paid off, accounting for 72.98% of the total number of samples. Besides, 3791 cases have overdue loans, accounting for 27.02% of the total number of samples.

3.2. SBM-DEA Method. This study used the SBM-DEA method (short for SBM method) to preprocess the data, because it can distinguish each customer to measure their respective efficiency value, rather than dividing them into different categories. This method can improve the prediction accuracy of the model and make the prediction of the initial logistic model more effective. The nonoriented SBM model is used in this study. The nonoriented SBM model is as follows:

$$\begin{aligned} \min \quad & \rho = \frac{1-(1/m)\sum_{i=1}^{m}\left(s_i^-/x_{ik}\right)}{1+(1/q)\sum_{r=1}^{q}\left(s_r^+/y_{rk}\right)} \\ & \text{s.t.} \quad X\lambda + s^- = x_k \\ & Y\lambda - s^+ = \mathrm{y}_k \\ & \lambda, s^-, s^+ \geq 0 \end{aligned} \tag{1}$$

The SBM model uses $p*$ to represent the efficiency value of the evaluating DMU. It measures the inefficiency from both input and output, which is called the nonoriented model. In the unsupervised SBM model, there is no zero in the input and output data. In the SBM model, the inefficiency of input and output is reflected as follows:

$$\begin{aligned} & \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{ij}}, \\ & \frac{1}{q}\sum_{r=1}^{q}\frac{s_r^+}{y_{rk}}. \end{aligned} \tag{2}$$

If the efficiency value ($p*$) of the SBM model is equal to 1, it means that the DMU evaluated is strongly efficient, while the efficiency of radial model is weakly efficient. The projection value (target value) of the evaluated DMU is

$$\begin{aligned} \widehat{x_k} &= x_k - s^-, \\ \widehat{y_k} &= y_k + s^+. \end{aligned} \tag{3}$$

The reasons for SBM indicator selection are as follows: the input indicators include borrower's liability information, credit risk score, and income information. These three indicators can mainly summarize the borrower's asset flow and external risk evaluation information. The output indicators are the borrower's loan amount and period, which are the most important indicators to describe the borrower's loan situation. Input and output indicators of the SBM method are shown in Table 1.

According to the correlation of indicators obtained in the initial stage of logistic regression and the experience summary in daily business, three input indicators and two output indicator were finally selected. Therefore, the following five indicators were selected as the input and output indicators of the SBM method.

According to the above five indicators, the efficiency value of each customer was obtained by using the SBM method through MaxDEA software. DEA_score was added to the next dataset. DEA_score distribution diagrams are shown in Figure 1.

Table 1: Input and output indicators of the SBM method.

Indicators		Indicator description
Input indicator	Income per month	Monthly income amount of the borrowing customer
	M_final_score	Credit risk score of external credit institutions to the buyer
	External_debt	Amount of external liabilities of the borrower
Output indicator	Loan amount	Loan amount of the borrower
	Product period	Number of loan periods of the borrower

3.3. Logistic-SBM Modelling Process. Due to the wide and complex dimensions of the data used in this study and the large amount of data involved, the logistic DEA model consisted of a series of steps. The logistic DEA model selected the input and output values of the DEA model according to the initial index of the logistic model method. Then, Max-DEA software is used to calculate the efficiency value of each customer as a decision unit (DMU). As a new index, the efficiency value obtained by DEA would be used in the training and testing of the logistic model together with the initial relevant index. Finally, the model was used to test the default probability of loan customers, which verifies the effectiveness and accuracy of the model. It was helpful to analyze the contribution of DEA index to the accuracy of the logistic regression model.

3.3.1. Feature Binning. Through the observation of the collected datasets, it was found that many data types are inconsistent, in which many of them were character type. Because these character indicators may play a great role in the model, we used weight of evidence (WOE) to transform many character indicators into measurable numerical indicators. According to the chi-square value of each pair of adjacent intervals, the two intervals with the smallest value are combined. The formula used in this step is as follows:

$$x = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(A_{ij}-E_{ij}\right)^2}{E_{ij}},$$
$$E_{\mathrm{ij}} = \frac{N_i \times C_i}{N}. \tag{4}$$

A_{ij} is the ith interval and the number of jth instances, E_{ij} is the desired frequency of A_{ij}, N is the total number of samples, N_i the number of samples in the ith group, and C_i is the proportion of the jth sample in the whole.

Feature information table is shown in Table 2. The continuous characteristic variable was discrete. Discrete feature states were often merged to reduce the number of states. It is convenient to transform all variables to similar scales. At the same time, some missing features will be brought into the model as an independent box. The reduction of extreme values and meaningless fluctuations in characteristics have an impact on the score and increase the stability and robustness of the model.

3.3.2. Correlation Coefficient. The correlation coefficient was obtained by calculating the correlation of each feature. The correlation coefficient formula is as follows:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}. \tag{5}$$

Among them, $\mathrm{Cov}(X, Y)$ is the covariance of X and Y; $\mathrm{Var}[X]$ and $\mathrm{Var}[Y]$ are the variance of X and Y, respectively.

If the absolute value of characteristic correlation coefficient was greater than 0.7, it was considered as a strong correlation feature. If there was strong correlation between features, some features can be deleted and one of them can be retained, as shown in Table 3. Delete the total debt ratio indicator.

3.3.3. IV. IV (Information Value) measures the amount of information about a variable. From the formula, it is equivalent to a weighted sum of the WOE values of the independent variables, in which the size of the value determines the influence of the independent variables on the target variables. The feature Information Value (IV) index can measure the concentration of the feature containing predictor variables. Weight of evidence (WOE) is a supervised coding method. The calculation formula is

$$\mathrm{WOE}_i = \log\left(\frac{G_1/G_{\mathrm{total}}}{B_1/B_{\mathrm{total}}}\right). \tag{6}$$

The IV is mainly used to code the input variables and evaluate the predictive ability. The value of characteristic variable IV indicates the predictive ability of the variable. The feature information degree of the remaining features was calculated, including the IV of the other features. After grouping, the formula for calculating the IV of each group is shown in Table 4.

According to the reference threshold of IV, the features with IV less than or equal to 0.02 are defined as nonpredictive features. Therefore, all features of this class were deleted. According to the characteristic IV shown in Table 5, “Marriage” and “Birth_month” features were deleted.

3.3.4. Random Forest Model. Random forest model is an integrated algorithm, which generates many trees and gets the result by voting or calculating the average. For grouped variables, cart Gini value is used as the evaluation standard. The steps of random forest model feature importance

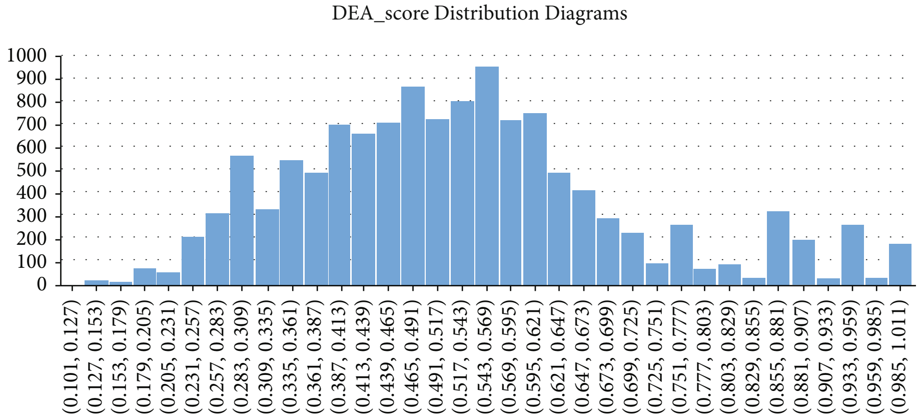


Figure 1: DEA_score distribution diagrams.

Table 2: Feature information table.

No.	Features	Feature interpretation	Class number
1	DEA_score	Efficiency score of borrowing	5
2	Education	Borrower's highest education	5
3	Marriage	Marital status of the borrower	4
4	Home type	Type of residence of the borrower	4
5	Company	Type of work unit of the borrower	5
6	Pay method	How the borrower pays wages	3
7	Job type	Job type of the borrower	4
8	Product name	Types of borrowers' lending products	4
9	Sales department	Which business department is responsible for the borrower's lending behavior	3
10	Bank	Ownership of bank card signed by the borrower	5
11	Family aware	Is the borrower aware of his borrowing behavior	3
12	Pro_id	The registered residence of a borrower	5
13	Birth month	Month of birth of the borrower	3
14	Birthday	Date of birth of the borrower	4
15	Inapv_edr	External debt ratio of borrowers	5
16	Inapv_idr	Internal debt ratio of the borrower	5
17	Inapv_tdr	Total debt ratio of the borrower	5
18	Age	Age of borrower	6
19	Entry date	Working days of the borrower	5

selection were as follows. The formula for calculating the Gini index is

$$\mathrm{GI}_m = \sum_{k=1}^{|k|} \sum_{k \neq k'} p_{mk} p'_{mk} = 1 - \sum_{k=1}^{|k|} p_{mk}^2. \tag{7}$$

The meaning of each indicator in the formula is as follows: k means that there are k categories.

P_{mk} means the proportion of the category k in the node m.

The importance of the feature x-j at the node m is the Gini exponential change before and after the node m branch and is calculated as follows:

$$\mathrm{VIM}_{jm}^{(\mathrm{Gini})} = \mathrm{GI}_M - \mathrm{GI}_l - \mathrm{GI}_r. \tag{8}$$

Among them, GI_l and GI_r, respectively, represent the Gini index of the two new nodes after branching.

When the node where the feature$x_j$appears in the decision tree$i$is in the set$M$, the calculation formula of the importance of$x_j$in the$i$th tree is

$$\mathrm{VIM}_{ij}^{(\mathrm{Gini})} = \sum_{m \in M} \mathrm{VIM}_{jm}^{(\mathrm{Gini})}. \tag{9}$$

Assuming that there are n trees in the RF, then the importance of x_j in the nth tree is

$$\mathrm{VIM}_{j}^{(\mathrm{Gini})} = \sum_{i=1}^{n} \mathrm{VIM}_{ij}^{(\mathrm{Gini})}. \tag{10}$$

Table 3: Index correlation.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	DEA_score	1.00	0.03	0.01	0.04	0.12	-0.16	-0.04	-0.29	0.00	0.08	-0.25	0.04	0.01	0.00	0.25	-0.12	0.14	0.02	0.07
2	Education	0.03	1.00	0.06	0.11	0.18	-0.16	-0.01	-0.04	-0.04	0.05	0.03	0.06	-0.01	0.01	0.10	-0.02	0.10	-0.10	0.10
3	Marriage	0.01	0.06	1.00	-0.04	0.03	-0.02	-0.03	0.02	-0.06	0.03	0.04	0.02	0.01	0.00	0.00	0.01	0.01	-0.37	-0.14
4	Home type	0.04	0.11	-0.04	1.00	0.20	-0.14	-0.05	-0.16	-0.05	0.01	0.02	0.21	-0.01	0.01	0.05	-0.04	0.04	0.08	0.17
5	Company	0.12	0.18	0.03	0.20	1.00	-0.41	-0.16	-0.30	-0.18	0.12	0.02	0.24	-0.02	0.02	0.23	-0.07	0.19	0.04	0.35
6	Pay_method	-0.16	-0.16	-0.02	-0.14	-0.41	1.00	0.11	0.29	0.25	-0.15	0.03	-0.19	-0.01	0.00	-0.32	0.09	-0.23	0.01	-0.26
7	Job type	-0.04	-0.01	-0.03	-0.05	-0.16	0.11	1.00	0.09	0.06	0.00	0.00	-0.03	0.00	-0.01	0.00	-0.04	-0.04	0.02	-0.05
8	Product name	-0.29	-0.04	0.02	-0.16	-0.30	0.29	0.09	1.00	0.05	-0.10	0.27	-0.13	-0.02	-0.02	-0.34	0.25	-0.17	-0.13	-0.18
9	Sales department	0.00	-0.04	-0.06	-0.05	-0.18	0.25	0.06	0.05	1.00	-0.06	-0.31	-0.33	0.00	0.01	-0.09	0.01	-0.11	0.03	-0.08
10	Bank	0.08	0.05	0.03	0.01	0.12	-0.15	0.00	-0.10	-0.06	1.00	-0.02	0.09	-0.01	0.00	0.18	-0.07	0.10	0.00	0.09
11	Family aware	-0.25	0.03	0.04	0.02	0.02	0.03	0.00	0.27	-0.31	-0.02	1.00	0.10	0.00	-0.03	-0.26	0.19	-0.05	-0.06	-0.04
12	Pro_id	0.04	0.06	0.02	0.21	0.24	-0.19	-0.03	-0.13	-0.33	0.09	0.10	1.00	0.00	0.01	0.09	-0.03	0.08	0.05	0.17
13	Birth month	0.01	-0.01	0.01	-0.01	-0.02	-0.01	0.00	-0.02	0.00	-0.01	0.00	0.00	1.00	0.02	0.01	0.00	0.00	-0.01	-0.01
14	Birthday	0.00	0.01	0.00	0.01	0.02	0.00	-0.01	-0.02	0.01	0.00	-0.03	0.01	0.02	1.00	0.01	0.00	0.00	0.02	0.00
15	Inapv_edr	0.25	0.10	0.00	0.05	0.23	-0.32	0.00	-0.34	-0.09	0.18	-0.26	0.09	0.01	0.01	1.00	-0.20	0.72	0.06	0.19
16	Inapv_idr	-0.12	-0.02	0.01	-0.04	-0.07	0.09	-0.04	0.25	0.01	-0.07	0.19	-0.03	0.00	0.00	-0.20	1.00	0.35	-0.09	-0.11
17	Inapv_tdr	0.14	0.10	0.01	0.04	0.19	-0.23	-0.04	-0.17	-0.11	0.10	-0.05	0.08	0.00	0.00	0.72	0.35	1.00	0.04	0.14
18	Age	0.02	-0.10	-0.37	0.08	0.04	0.01	0.02	-0.13	0.03	0.00	-0.06	0.05	-0.01	0.02	0.06	-0.09	0.04	1.00	0.33
19	Entry date	0.07	0.10	-0.14	0.17	0.35	-0.26	-0.05	-0.18	-0.08	0.09	-0.04	0.17	-0.01	0.00	0.19	-0.11	0.14	0.33	1.00

Table 4: Group IV calculation formula.

Group	WOE	IV
Group 1	$\log\left(\frac{G_1/G_{\text{total}}}{B_1/B_{\text{total}}}\right)$	$\left(\frac{G_1}{G_{\text{total}}}-\frac{B_1}{B_{\text{total}}}\right)\log\left(\frac{G_1/G_{\text{total}}}{B_1/B_{\text{total}}}\right)$
Group 2	$\log\left(\frac{G_2/G_{\text{total}}}{B_2/B_{\text{total}}}\right)$	$\left(\frac{G_2}{G_{\text{total}}}-\frac{B_2}{B_{\text{total}}}\right)\log\left(\frac{G_2/G_{\text{total}}}{B_2/B_{\text{total}}}\right)$
……	……	……
Group n	$\log\left(\frac{G_n/G_{\text{total}}}{B_n/B_{\text{total}}}\right)$	$\left(\frac{G_n}{G_{\text{total}}}-\frac{B_n}{B_{\text{total}}}\right)\log\left(\frac{G_n/G_{\text{total}}}{B_n/B_{\text{total}}}\right)$
Total		$\sum\left(\frac{G_i}{G_{\text{total}}}-\frac{B_i}{B_{\text{total}}}\right)\log\left(\frac{G_1/G_{\text{total}}}{B_1/B_{\text{total}}}\right)$

Table 5: IV of each feature.

Feature No.	Features	IV
1	DEA_score	0.104
2	Education	0.025
3	Marriage	0.011
4	Home type	0.117
5	Company	0.083
6	Pay method	0.079
7	Job type	0.067
8	Product name	0.796
9	Sales department	0.204
10	Bank	0.054
11	Family aware	0.319
12	Pro_id	0.067
13	Birth_month	0.004
14	Birthday	0.020
15	Inapv_edr	0.168
16	Inapv_idr	0.168
17	Age	0.149
18	Entry date	0.046

Finally, the importance scores obtained through normalization are processed. The formula is as follows:

$$\text{VIM}_j=\frac{\text{VIM}_j}{\sum_{i=1}^{c}\text{VIM}_i}. \tag{11}$$

The variable importance score is represented by VIM, and the Gini index is represented by GI. Assuming that there aremfeaturesx_1,x_2, andx_m, the Gini index score of each featurex_iis now calculated. Features are ranked from high to low according to their importance, and the top n features are selected.

Order of feature importance is shown in Table 6. Firstly, the feature variables in the random forest were sorted in descending order according to VI (variable importance). Then, the indexes with unimportant proportion were removed from the current feature variables to obtain a new feature set.

Table 6: Order of feature importance.

No.	Features	Importance	Cum_importance
1	Product name	0.204	0.204
2	Family aware	0.086	0.29
3	Age	0.066	0.356
4	Inapv_idr	0.065	0.420
5	Entry date	0.060	0.480
6	Birthday	0.060	0.540
7	Inapv_edr	0.058	0.599
8	DEA_score	0.056	0.655
9	Home type	0.054	0.709
10	Pro_id	0.047	0.756
11	Sales department	0.046	0.803
12	Job type	0.046	0.849
13	Education	0.044	0.893
14	Company	0.043	0.936
15	Bank	0.041	0.977
16	Pay_method	0.023	1.000

The result of feature importance was obtained by the random forest algorithm. The results would be retained three decimal places and sorted according to the importance from high to low. At the same time, the cumulative importance was calculated. According to the feature importance ranking, it was obvious that the feature of "Pay_method" showed the low importance.

3.3.5. Logistic-SBM Model Variables. In the application of P2P network credit loan, the logistic model was adopted due to its high discrimination ability in the field of default loan customer identification. The logistic formula is

$$\begin{aligned} E(p)&=f\left(\beta_0+\sum\beta_i x_i\right),\\ f(x)&=\frac{\exp(x)}{1+\exp(x)}. \end{aligned} \tag{12}$$

The overdue status of a group of applicants in the performance period is $\{y_1, y_2, \cdots, y_n\}$ and $y_i \in \{0,1\}$. The likelihood function and log likelihood function are

$$\text{L}(p)=\prod P(Y=y_i)=\prod p^{y_i}(1-p)^{1-y_i},$$

$$\begin{aligned} l(p)=\log(\text{L}(p))&=\log\left(\prod P(Y=y_i)\right)=\sum(y_i\log(p)\\ &\quad+(1-y_i)\log(1-p))=\sum\left(y_i\left(\beta_0+\sum\beta_i x_{ij}\right)\right.\\ &\quad\left.-\log\left(1+\exp\left(\beta_0+\sum\beta_i x_{ij}\right)\right)\right). \end{aligned} \tag{13}$$

The parameter estimation formula is as follows:

$$\hat{p} = \mathrm{argmax}\, l(p),$$
$$\hat{p} = \frac{\sum y_i}{n},$$

$$\begin{aligned} l(p) &= \sum (y_i \log (p) + (1 - y_i) \log (1 - p)) \\ &= \sum \left(y_i \left(\beta_0 + \sum \beta_i x_{ij} \right) - \log \left(1 + \exp \left(\beta_0 + \sum \beta_i x_{ij} \right) \right) \right). \end{aligned} \tag{14}$$

The parameter estimation formula is as follows:

$$\frac{\partial \mathrm{l}}{\partial \beta_q} = \sum \left(y_i - \frac{1}{\exp \left(-\beta_0 - \sum \beta_i x_{ij} \right)} \right) x_{iq}. \tag{15}$$

Estimate the β_q by the gradient descent method; the formula is as follows:

$$\begin{aligned} \beta_q^{r+1} &= \beta_q^r - h\delta, \\ \delta &= \frac{\partial l}{\partial \beta_q} \bigg| \beta_q = \beta_q^r. \end{aligned} \tag{16}$$

It is very important to select variables from the dataset. Considering the correlation coefficient, validity, and importance of index data, 15 variables were selected in the final logistic-SBM model for empirical study. Logistic-SBM model variables are shown in Table 7.

4. Result Analysis and Inspection

Model verification is used to measure the predictive ability of the developed model, including internal and external tests. The internal test is the comparison between the prediction situation of the test set in the sample and the actual situation. The external test is the comparison between the prediction situation and the actual situation of the dataset except the model after passing the model. The primary goal of the developed model is to distinguish whether the borrower is in default. Besides, the accuracy of model prediction, confusion matrix analysis, and the Kolmogorov-Smirnov test can all be used as criteria for judging the quality of this model.

4.1. Confusion Matrix Analysis. Accuracy is an important concept and indicator in model evaluation. The performance of the resulting classier can then be evaluated in terms of the recall (or sensitivity) and precision of the classier on an evaluation dataset. Recall and precision are defined in terms of the number of true positives (TP), misses (FN), and false alarms (FP) of the classier (cf. Table 8).

In Table 7, the first line expresses prediction results from the prediction model; the first column expresses the actual results in the original data. True positive (TP) expresses the amount that the positive samples are correctly classified as positive; false negative (FN) expresses the amount that the positive samples are misclassified as negative; false positive (FP) expresses the amount that the negative samples are misclassified as positive; true negative (TN) expresses the amount that the negative samples are correctly classified as negative. As the common evaluation measures, the accuracy-specific expressions are shown as follows:

Table 7: Logistic-SBM model variables.

Feature No.	Features
1	DEA_score
2	Education
3	Home type
4	Company
5	Job type
6	Product name
7	Sales department
8	Bank
9	Family aware
10	Pro_id
11	Birthday
12	Inapv_edr
13	Inapv_idr
14	Age
15	Entry date

$$A(\text{accuracy}) = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}. \tag{17}$$

The borrower results predicted by the model were compared with the marked good and bad borrowers. From this result, the model has a strong predictive ability. 77.49% of borrowers were accurately predicted, and only 22.51% of borrowers were incorrectly predicted. Among them, the first quadrant is the number of borrowers that the model predicts to be nondefaulting and actually not defaulting. In the second and third quadrants, the number of errors is predicted. The fourth quadrant indicates that the model predicts the number of defaults and actual defaults. The accuracy of the model was that the ratio of the number of accurate predictions to the total number was 77.49%, in which the accuracy rate was high.

4.2. AUC-ROC Curve Observation. The AUC-ROC curve is a performance measurement for classification problems under various threshold settings. ROC (receiver operating characteristic curve) is a probability curve, and AUC (area under the curve) represents the degree or measure of separability which represents how many models can distinguish categories. The higher the AUC, the better the model predicts 0 as 0 and 1 as 1. The ROC curve of the logistic-DEA model is shown in Figure 2.

4.3. K-S Test. The KS indicator measures the largest gap between the cumulative distribution of responding

Table 8: Confusion matrix for binary classification.

		Prediction positive 1	Prediction negative 0	Total N
Actually positive	1	True positives (TP)	False positives (FN)	N-pos
Actually negative	0	False negatives (FP)	True negatives (TN)	N-neg
Total M		M-pos	M-neg	

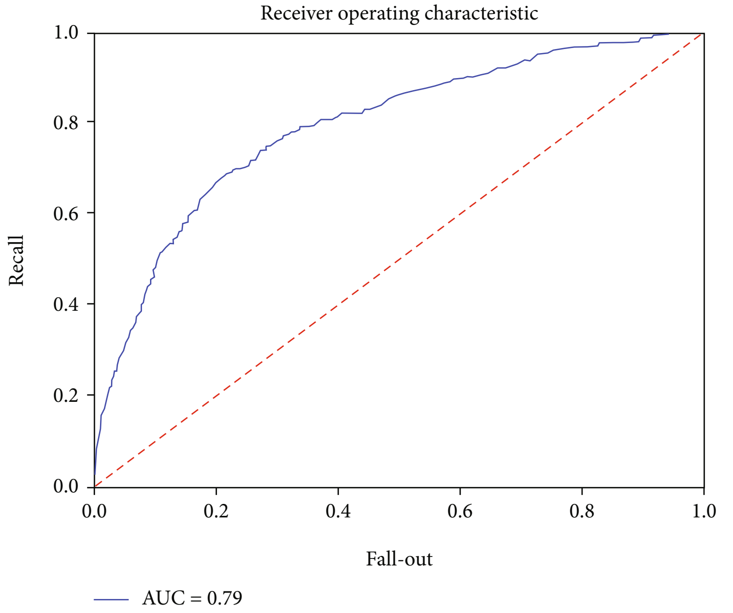


Figure 2: AUC-ROC curve.

customers and nonresponding customers. The calculation formula was as follows:

$$\mathrm{KS} = \mathrm{MAX}(\mathrm{ABS}(\mathrm{CPD}(S_i) - \mathrm{CPG}(S_i))). \tag{18}$$

$\mathrm{CPD}(S_i)$ is the proportion distribution of accumulated good customers, $\mathrm{CPG}(S_i)$ is the proportion distribution of accumulated bad customers.

Firstly, the scores of samples were ranked from large to small and then, the cumulative proportion of good and bad samples in each quantize interval was calculated. The larger the distance between the two, the higher the KS value, indicating that the model area has the ability to distinguish good and bad customers. In the actual business, if the KS value is less than 20%, the accuracy of the model is poor. If the KS value is between 20% and 30%, it means that the model discrimination effect is general. If the KS value is between 30% and 60%, the model is very effective.

The KS value was obtained by the logistic-SBM model, as shown in Figure 3. The KS value of the logistic-SBM model is 33.3%, indicating the good prediction effect and the better effect of distinguishing default customers of the model.

4.4. Comparison of Model Evaluation. Precision, specificity, and recall are important concepts and indicators in model evaluation too. As the common evaluation measures, sensitivity, specificity, G-Measure, and F-Measure are used to make the evaluation. F-Measure is also called F-Score. F-Measure is the weighted harmonic average of precision (P) and recall (R). It is an evaluation standard of the model and is often used to evaluate the quality of the classification model. The F-Measure function synthesizes the results of P and R when the parameter $\alpha = 1$; the weight of P and R is the same. When F-Measure is higher, the model is more effective. Their specific expressions are shown as follows:

$$R(\mathrm{recall}) = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},$$

$$S(\mathrm{specificity}) = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}},$$

$$P(\mathrm{precision}) = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},$$

$$G\text{-Measure} = \sqrt{\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}} = \sqrt{RS},$$

$$\begin{aligned} F\text{-Measure} = F_\alpha &= \frac{\left(1+\alpha^2\right)(\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})\mathrm{TP}+\mathrm{FN}) \times (\mathrm{TP}/(\mathrm{TP}+\mathrm{FP})\mathrm{TP}+\mathrm{FP})}{\alpha^2(\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})\mathrm{TP}+\mathrm{FN}) + (\mathrm{TP}/(\mathrm{TP}+\mathrm{FP})\mathrm{TP}+\mathrm{FP})} \\ &= \frac{\left(1+\alpha^2\right)PR}{\alpha^2 P + R}, \end{aligned}$$

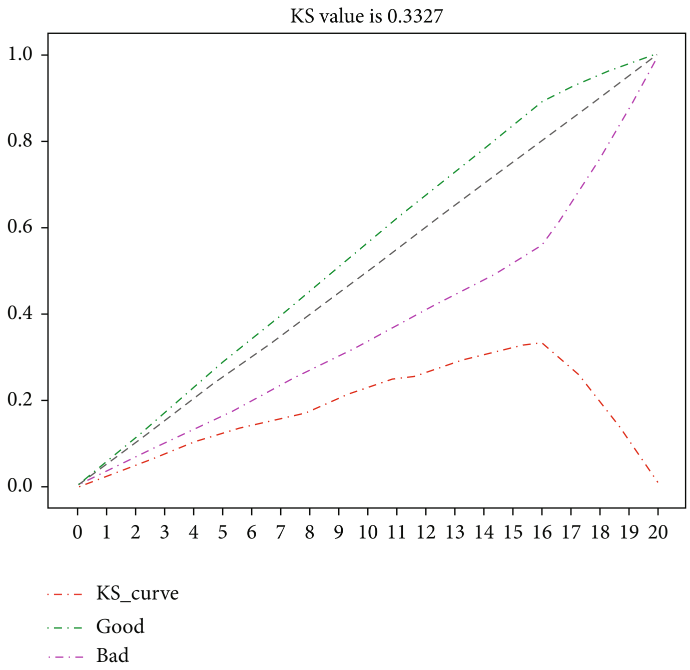


Figure 3: K-S value.

Table 9: Performance comparison of two models.

	Logistic-SBM	Logistic
Accuracy (%)	77.49	76.88
Recall (sensitivity) (%)	92.18	91.7
Precision (%)	79.87	79.54
Specificity (%)	38.73	37.78
G-Measure (%)	59.75	58.86
F-Measure (%)	85.59	85.19

$$F_1 = \frac{(\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})\mathrm{TP}+\mathrm{FN}) \times (\mathrm{TP}/(\mathrm{TP}+\mathrm{FP})\mathrm{TP}+\mathrm{FP})}{(\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})\mathrm{TP}+\mathrm{FN}) + (\mathrm{TP}/(\mathrm{TP}+\mathrm{FP})\mathrm{TP}+\mathrm{FP})} = \frac{PR}{P+R}. \tag{19}$$

We compare the mean values of corresponding evaluation measures of two models. The performance comparison of two models is shown in Table 9.

The relationship of two models can fully explain that the logistic-SBM model presented by this article has the optimal performance relative to the logistic model. The higher the value of related evaluation indicators, the better the effect of the model. Simulation results show that the logistic-SBM is more suitable for credit risk evaluation than the popularly used logistic with consideration of related evaluation indicators. According to the above research results, it can be known that using data envelopment analysis to preprocess the data and increase the efficiency value in the logistic regression model can improve the accuracy of the model.

5. Concluding Remarks

With the rapid development of the Internet, P2P has been applied in various fields [27]. At present, the risk management of borrowers in the P2P network lending platform mainly includes the following: first, the basic information authentication of borrowers. Mine their identity information and credit level from many aspects, and rate the borrowers. Feature variables are extracted from the basic information to determine the characteristics of credit management. The second is the combination of credit line management and credit risk. The loan limit of the borrower corresponds to the corresponding credit risk level.

Credit risk has four main characteristics: asymmetry, accumulation, unsystematic, and endogenous. The good operation of a platform requires strict audit of borrowers. Only through high-quality borrowers to minimize the risk of P2P network credit transactions can the P2P platform maintain stable operation. The grade assigned by the P2P lending site is the most predictive factor of default, but the accuracy of the model is improved by adding other information, especially the borrower's debt level [28]. The results suggest that borrower's social information can be used not only for credit screening but also for default reduction and debt collection [29].

Relevant suggestions have been put forward, which provide reference for the credit management of the P2P network lending industry in China. Regulatory authorities and the platform itself should take some measures to control the credit risk of the P2P Internet lending industry. The specific recommendations were as follows: (1) improve the social credit investigation system, and realize information sharing; (2) improve and implementation of policies; and

(3) undertake social responsibility, and actively develop through innovation.

Data Availability

This study collected partial loan records from an inclusive finance platform in China from 2014 to 2018.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] K. E. Davis and A. Gelpern, "Peer-to-peer financing for development: regulating the intermediaries," *NYUJ Int'l L. & Pol.*, vol. 42, p. 1209, 2009.

[2] P. Slattery, "Square pegs in a round hole: SEC regulation of online peer-to-peer lending and the CFPB alternative," *Yale J. on Reg.*, vol. 30, p. 233, 2013.

[3] B. Budiharto, S. N. Lestari, and G. Hartanto, "The legal protection of lenders in peer to peer lending system," *Law Reform*, vol. 15, no. 2, pp. 275–289, 2019.

[4] Q. Wang, X. Xiong, and Z. Zheng, "Platform characteristics and online peer-to-peer lending: evidence from China," *Finance Research Letters*, vol. 38, article 101511, 2021.

[5] X. Liu, "A visualization analysis on researches of internet finance credit risk in coastal area," *Journal of Coastal Research*, vol. 103, no. sp1, pp. 85–89, 2020.

[6] X. Fang, B. Wang, L. Liu, and Y. Song, "Heterogeneous traders, the leverage effect and volatility of the Chinese P2P market," *Journal of Management Science and Engineering*, vol. 3, no. 1, pp. 39–57, 2018.

[7] W. Zhang, Y. Zhao, P. Wang, and D. Shen, "Investor sentiment and the return rate of P2P lending platform," *Asia-Pacific Financial Markets*, vol. 27, no. 1, pp. 97–113, 2020.

[8] L. Ma, Y. Li, D. Li, H. Li, Y. Wang, and C. Ren, "Risk identification and decision making for P2P companies: an empirical study in the Bohai coast regions," *Journal of Coastal Research*, vol. 106, no. sp1, pp. 191–196, 2020.

[9] Z. Abdul Halim, J. How, P. Verhoeven, and M. K. Hassan, "Asymmetric information and securitization design in Islamic capital markets," *Pacific-Basin Finance Journal*, vol. 62, p. 101189, 2020.

[10] X. Lv, L. Zhou, and X. Guo, "Research on P2P network loan risk evaluation based on generalized DEA model and R-type clustering analysis under the background of big data," *Journal of Financial Risk Management*, vol. 6, no. 2, pp. 163–190, 2017.

[11] Y. Guo, W. Zhou, C. Luo, C. Liu, and H. Xiong, "Instance-based credit risk assessment for investment decisions in P2P lending," *European Journal of Operational Research*, vol. 249, no. 2, pp. 417–426, 2016.

[12] M. Herzenstein, U. M. Dholakia, and R. L. Andrews, "Strategic herding behavior in peer-to-peer loan auctions," *Journal of Interactive Marketing*, vol. 25, no. 1, pp. 27–36, 2011.

[13] J. H. Zeng and S. Yang, "Herding behavior of lenders in P2P lending markets and its rational test: evidence from PaiPaiDai market," *Modern Finance and Economics (Journal of Tianjin University of Finance and Economics)*, p. 7, 2014.

[14] G. A. Akerlof, "The Market for Lemons: Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, vol. 84, no. 3, pp. 488–500, 1970.

[15] S. C. Berger and F. Gleisner, "Emergence of financial intermediaries in electronic markets: the case of online P2P lending," *BuR Business Research Journal*, vol. 2, no. 1, pp. 39–65, 2009.

[16] G. N. Weiss, K. Pelger, and A. Horsch, *Mitigating adverse selection in P2P lending–empirical evidence from prosper. com*Available at SSRN 1650774, 2010.

[17] M. A. Razi, J. M. Tarn, and F. A. Siddiqui, "Exploring the failure and success of DotComs," *Information Management & Computer Security*, vol. 12, no. 3, pp. 228–244, 2004.

[18] G. D. Bruton and Y. Rubanik, "Resources of the firm, Russian high-technology startups, and firm growth," *Journal of Business Venturing*, vol. 17, no. 6, pp. 553–576, 2002.

[19] Y. Honjo, "Business failure of new firms: an empirical analysis using a multiplicative hazards model," *International Journal of Industrial Organization*, vol. 18, no. 4, pp. 557–574, 2000.

[20] R. Sullivan, "Entrepreneurial learning and mentoring," *International Journal of Entrepreneurial Behavior & Research*, vol. 6, no. 3, pp. 160–175, 2000.

[21] M. Lan, *Online P2P lending industry: an international analysis*, University of Nottingham, 2019.

[22] X. Chen, X. Hu, and S. Ben, "How do reputation, structure design and FinTech ecosystem affect the net cash inflow of P2P lending platforms? Evidence from China," *Electronic Commerce Research*, vol. 21, no. 4, pp. 1055–1082, 2021.

[23] M. H. Akhtar, I. S. Chaudhry, M. R. Sheikh, and A. Shahzadi, "Business model, risk and financial stability of banks: a multi-country analysis," *Pakistan Journal of Social Sciences (PJSS)*, vol. 40, no. 1, pp. 401–414, 2020.

[24] N. Barasinska and D. Schäfer, "Is Crowdfunding Different? Evidence on the Relation between Gender and Funding Success from a German Peer-to-Peer Lending Platform," *German Economic Review*, vol. 15, no. 4, pp. 436–452, 2014.

[25] R. Liu, N. Chen, and Y. Li, *"The Herd Behavior on Peer-to-Peer Online Lending Markets: Evidence from China," Discrete Dynamics in Nature and Society, Vol*, vol. 2021, 2021.

[26] J. D. Velimirovic and A. Janjic, "Risk assessment of circuit breakers using influence diagrams with interval probabilities," *Symmetry*, vol. 13, no. 5, p. 737, 2021.

[27] H. Wang, K. Fan, H. Li, and Y. Yang, "A dynamic and verifiable multi-keyword ranked search scheme in the P2P networking environment," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2342–2355, 2020.

[28] C. Serrano-Cinca, B. Gutierrez-Nieto, and L. Lopez-Palacios, "Determinants of default in P2P lending," *PLoS One*, vol. 10, no. 10, article e0139427, 2015.

[29] R. Ge, J. Feng, B. Gu, and P. Zhang, "Predicting and deterring default with social media information in peer-to-peer lending," *Journal of Management Information Systems*, vol. 34, no. 2, pp. 401–424, 2017.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 4349009, 10 pages
https://doi.org/10.1155/2022/4349009

Research Article

Radar Target Recognition and Location Based on CapsNetv2

Jiaxing Hao,[1] Xuetian Wang,[1] Sen Yang,[2] and Hongmin Gao[1]

[1]*School of Information and Electronics, Beijing Institute of Technology, No. 5 zhongguancun South Street, Beijing 100081, China*
[2]*Department of UAV Engineering, Army Engineering University, 97 Heping West Road, Shijiazhuang 050003, China*

Correspondence should be addressed to Jiaxing Hao; 873328461@qq.com

Received 29 March 2022; Accepted 14 May 2022; Published 26 May 2022

Academic Editor: Maode Ma

For precise detection and positioning of weapons and equipment under complex ground backgrounds and weather-changing aerial backgrounds. Compared with the traditional convolutional neural networks, the Capsule Network (CapsNet) is more suitable for identifying weapons and equipment in complex backgrounds because it uses vectors as input for the first time, which can well retain the characteristic information such as the direction and the angle of the target. Therefore, this paper proposes a radar target classification algorithm based on the combination of CapsNetv2 and infrared lidar, which simplifies the convolutional layer of the traditional 9×9 capsule network through a 1×1 reduction layer and a 3×3 convolution kernel, and adopts a double-layer capsule layer. Two prediction frames are obtained to improve the recognition accuracy; at the same time, the output volume retains the direction and the angle, which can more accurately classify the radar targets in various complex backgrounds. Applying the method proposed in this article to the MSTAR dataset shows that the radar target positioning is accurate. The rate increases to 99.5%. Finally, compared with the AlexNet and the YOLOv4 methods designed by Alex Krizhevsky, the proposed radar target recognition method can accurately and quickly identify weapons and equipment from complex backgrounds. The results obtained from the CapsNetv2 are accurately compared with other methods' in complex backgrounds. The proposed method significantly improves the efficiency of military inspections.

1. Introduction

Radar technology plays an important role in the field of modern target detection. It is widely used in military and civil transportation fields due to its all-weather and omnidirectional work characteristics. Target recognition is one of the basic tasks in computer vision. Identifying the target area and obtaining the accurate position of the target lay the foundation for the next information processing of the carrier and improve the perception ability of machine recognition. The current main model is to process the visible light image accordingly. However, the visible light image is susceptible to environmental lighting. Under low light, dark or shadow conditions blocked by surrounding interference, the processing data becomes more complicated. To achieve high-reliability classification and recognition effects, modern pattern recognition theories and methods are usually used for classifier design, such as statistical-based pattern classification methods, feature extraction methods, and neural network-based pattern classification methods.

Statistics-based classification and recognition algorithms use probability models to obtain the feature vector distribution of each category and classify the unknown samples. For example, Shen Yanyan obtained the likelihood function by extracting the ocean wave radar echoes and used Bayesian classifiers for classification [1]. Liu Jingrui and others established a weather radar warning system using probability statistical models to distinguish between strong and weak rainfalls [2]. However, the weather environment is complex and changeable, and there are many interference factors, making the actual task of processing the radar signals much more difficult than processing and identifying the visible light images.

The traditional feature methods mainly match the known features by extracting the feature points of the target. Commonly used feature matching methods include the histogram of oriented gradient (HOG) feature [3], the scale-invariant feature transform(SIFT) feature [4], and the speed-up robust features (SURF) feature [5]. In 2001, the American company ENSCO developed the Visual Identity

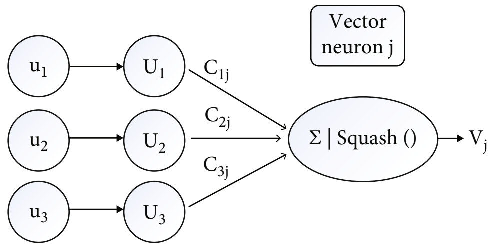


Figure 1: Capsule neuron model.

System (VIS) track video detection system to realize real-time detection of the working status of PandaPal fasteners [6]. In 2005, the German railway engineering company GBM Wiebe developed the GeoRail-Xpress comprehensive inspection vehicle that was able to perform a real-time inspection of the entire railway electrical equipment system [7]. However, because it is necessary to extract and classify multiple regions of the image, the recognition speed was slow, and it was difficult to meet the requirements of real-time detection.

In recent year, the target recognition algorithms based on deep learning have made significant progress compared with the traditional target detection algorithms. The representative algorithms include R-CNN [8], Fast-RCNN [9] and Faster-RNN [10]. However, the detection steps of these methods are more complicated, and the real-time effect is poor. The AlexNet [11] and YOLO [12] that have appeared one after another can meet the requirements of real-time detection, but often require a lot of data for neural network training, and the training weights are easy to overfit. Moreover, the technical requirements for the equipment are relatively high. In 2017, Hinton proposed that the Capsule Network, referred to as CapsNet, would possibly replace the traditional CNN network, bringing new opportunities to the field of deep learning [13]. For example, the literature [14] used the capsule network to classify handwritten digits. Due to the single characteristics of the digits, the recognition rate is high. However, the radar targets are generally weapons and equipment with complex structures and are easily affected by the conditions such as illumination and angle.

This paper proposes a radar target recognition and location algorithm based on CapsNetv2. For the characteristics of weapons and equipment, light intensity, position, deformation, angle, texture, and position information should be considered. Therefore, these six features are selected as the input vectors, and then a 1×1 reduction layer combined with a 3×3 convolutional layer is used to simplify the traditional Capsule 9×9 capsule neurons. Then the MSTAR dataset is trained and learned through the double-layer capsule network, and two prediction boxes are obtained, one large and one small, to complete the recognition under different complex backgrounds. Finally, the improved CV model and the infrared Lidar uses edge detection to accurately locate the location of hazardous enemy weapons and equipment.2 Structure.

A paper for publication can be subdivided into multiple sections: title, list of all the authors and their affiliations, a concise abstract, keywords, main text (including figures, equations, and tables), acknowledgement, references, and appendix.

2. Capsnetv2

In 2011, Hinton proposed the concept of capsule [15]. Unlike the traditional scalar neurons, the capsule network is a vector composed of many neurons. The vector length of the capsule neuron model indicates the possibility of the existence of the target passed by the upper network, and its direction represents the actual state of the entity, that is, "Instance parameters", as shown in Figure 1 [16].

The dynamic routing algorithm (Squash) solves the problem by the output value of the capsule.

The update formula is:

$$c_{ij} = \frac{\exp\left(b_{ij}\right)}{\sum_k \exp(b_{ik})} \tag{1}$$

$$b_{ij} = b_{ij} + U_i \cdot V_j \tag{2}$$

where c_{ij} is the dynamic routing coupling coefficient and k is the number of initial similarity weights b_{ij}.

The capsule output S_j is obtained from the lower-level capsule inputs U_i and c_{ij}:

$$S_j = \sum_i c_{ij} U_i \tag{3}$$

where U_i is derived from $U_i = W_{ij} u_i$,

W_{ij} is the weight of the capsule network.

The output V_j should be expressed as a probability. Thus, the output value should be controlled between [0, 1], which can be obtained by nonlinear compression:

$$V_j = \frac{\left\|S_j\right\|^2}{1 + \left\|S_j\right\|^2} \times \frac{S_j}{\left\|S_j\right\|} \tag{4}$$

The principle of CapsNetv2 is roughly the same as that of the capsule network. The image is first input to the convolutional layer (ReLu), and a basic capsule layer is obtained through the convolution operation. Then the data of the basic capsule layer is transmitted to the image through the dynamic routing algorithm (squash). The capsule layer then transfers the image capsule layer data to the feature capsule layer, and finally uses the fully connected layer to reorganize and model the feature capsule layer data. However, the CapsNetv2 consists of two image capsule layers and two feature capsule layers. If training is performed when the data of one layer of the capsule layer has over-fitting, it can ensure the success of the training of the other capsule layer. The structure of the CapsNetv2 is shown in Figure 2.

The radar target image is composed of 3 categories, which are set as BTR70(armored transport vehicle), BMP2(infantry fighting vehicle), and T72(tank). The moduli of the three types of target vectors are calculated and the

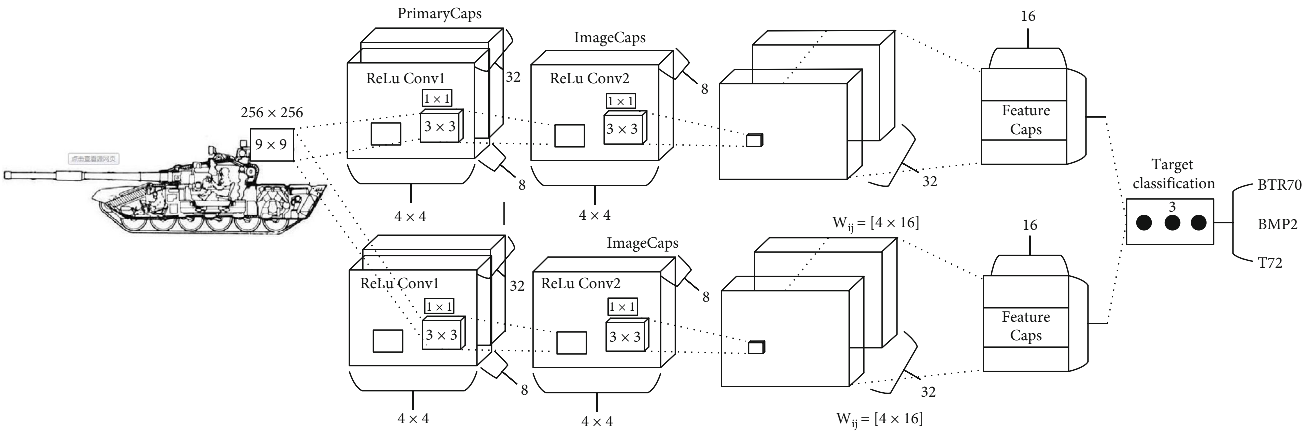


Figure 2: Schematic diagram of the CapsNetv2 structure.

Table 1: Performance comparison of different models.

Model	Top-1%	Top-5%	GPU/ms	CPU/S
AlexNet	57.0	90.0	2.5	0.30
Yolov4	63.0	92.0	2.7	0.26
CapsNet	70.0	92.5	1.8	0.66
CapsNetv2	72.0	93.5	1.3	0.18

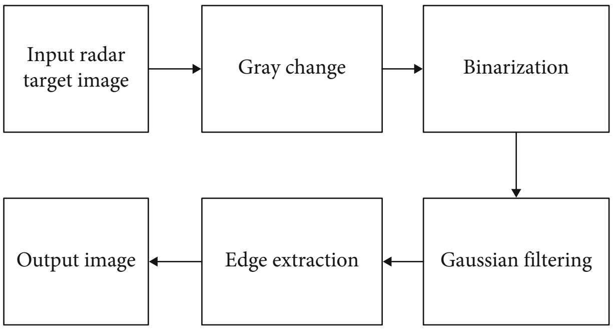


Figure 3: Flow chart of hanging string image pre-processing.

vector with the largest modulus value is the category with the highest possible target probability.

The AlexNet, Yolov4 and the traditional capsule network are used to compare the performance of the CapsNetv2 and classify the image dataset. Table 1 compares the Top-1% and Top-5% classification performance of each model for the same dataset, where the GPU model is Titan X, and the CPU model is Intel I7-10700(4GHz).

As shown in Table 1, CapsNetv2 has higher classification accuracy in Top-1% and Top-5% compared with AlexNet, Yolov4 and CapsNet. Moreover, the recognition time of GPU and CPU is less, indicating that the CapsNetv2 has better performance.

3. Principles of Radar Target positioningSubheadings

3.1. Radar Image Preprocessing. In the radar target recognition technology, the collected radar image contains various disadvantages such as noise, jitter, and weak light due to its complex background, weather and other factors that will affect the model training and recognition results. Therefore, it is necessary to pre-process and correct the collected original image and then extract the feature value of the target and separate the target from the background.

The pre-processing steps include grayscale change, binarization, noise reduction, filtering and edge extraction. The specific flow chart of pre-processing is shown in Figure 3.

(1) Perform grayscale processing on images of different categories in the mSTAR dataset. The results are shown in Figure 4(a);

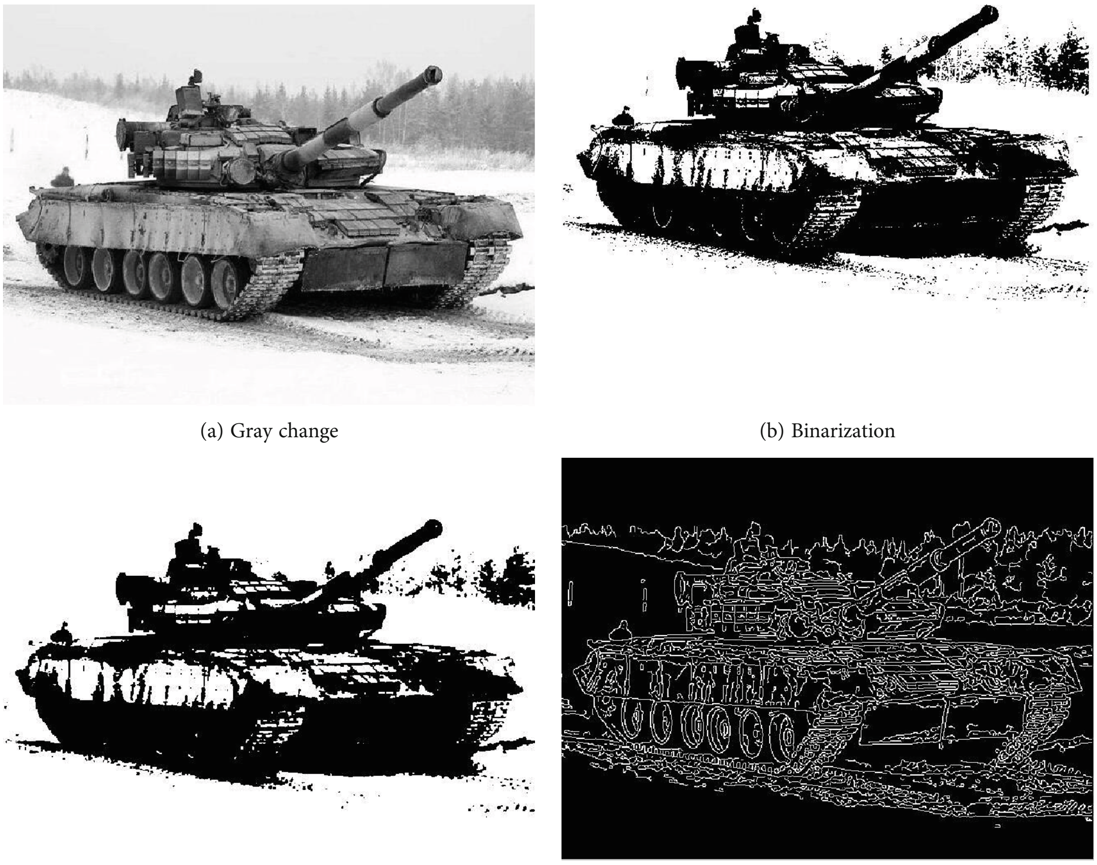
(2) Binarize the grayscale processed image to remove the influence of complex background, that is, set the pixel point to 0 or 255, where the target gray value is 255, and the other background is 0 as shown in Figure 4(b);

(3) Noise will reduce the quality of the image, and the collected radar target image is usually accompanied by auxiliary equipment and anti-jamming equipment that contain a lot of Gaussian noise. Therefore, this article uses Gaussian filtering to process the image, as shown in Figure 4(c);

(4) By comparing the radar target recognition effect with Robert, Sobel or LOG operator, the edge of the target detected by the Canny algorithm is more complete. Therefore, this paper uses the Canny algorithm to extract the edge of the radar target as shown in Figure 4(d).

3.2. Improved CV Model. The Chan-Vese(CV) model is used to divide the fuselage and barrel of the T72 tank. The energy function of the CV model is [17]:

$$E(C, c_1, c_2) = \mu \cdot L(C) + \lambda_1 \int_{i(C)} (\mu_0(x, y) - c_1)^2 + \lambda_2 \int_{i(C)} (\mu_0(x, y) - c_2)^2 \tag{5}$$



(a) Gray change

(b) Binarization

(c) Gaussian filtering

(d) Edge extraction

Figure 4: Processing results of the T72 tank.

$$c_1 = \frac{\int_{\Omega}\mu(x,y)H_{\varepsilon}(\phi)dxdy}{\int_{\Omega}H_{\varepsilon}(\phi)dxdxy} \tag{6}$$

$$c_2 = \frac{\int_{\Omega}\mu(x,y)[1-H_{\varepsilon}(\phi)]dxdy}{\int_{\Omega}[1\text{-}H_{\varepsilon}(\phi)]dxdxy} \tag{7}$$

In the formula, μ is the CV model constant; $L(C)$ is the arc length of the curve C; $\mu \cdot L(C)$ is the length term, which can smooth the evolution curve; λ_1 and λ_2 are the weight coefficients, both greater than 0; $\mu_0(x,y)$ is the image pixel gray value; c_1 and c_2 are the average gray values of the pixels outside and inside the image evolution curve, respectively, and $H_{\varepsilon}(\phi)$ is the regularization step function.

Considering that targets such as armored vehicles or tanks are regular models with regular shapes and horizontal symmetry, adding the level set method can better correct the contour topological changes. The level set evolution Euler-Lagrangian equation is:

$$\begin{aligned}\partial\phi/\partial t = \delta_1 &\left[\mu div(\nabla\phi/|\nabla\phi|) - \lambda_1\int_{i(C)}(\mu_0(x,y)-c_1)^2\right] \\ &+ \lambda_2\int_{i(C)}(\mu_0(x,y)-c_2)^2\end{aligned} \tag{8}$$

where δ_1 is the global function, as the impulse function of the CV model; div is the divergence operator and div $(\nabla\phi/|\nabla\phi|)$ is the curvature of the evolution curve.

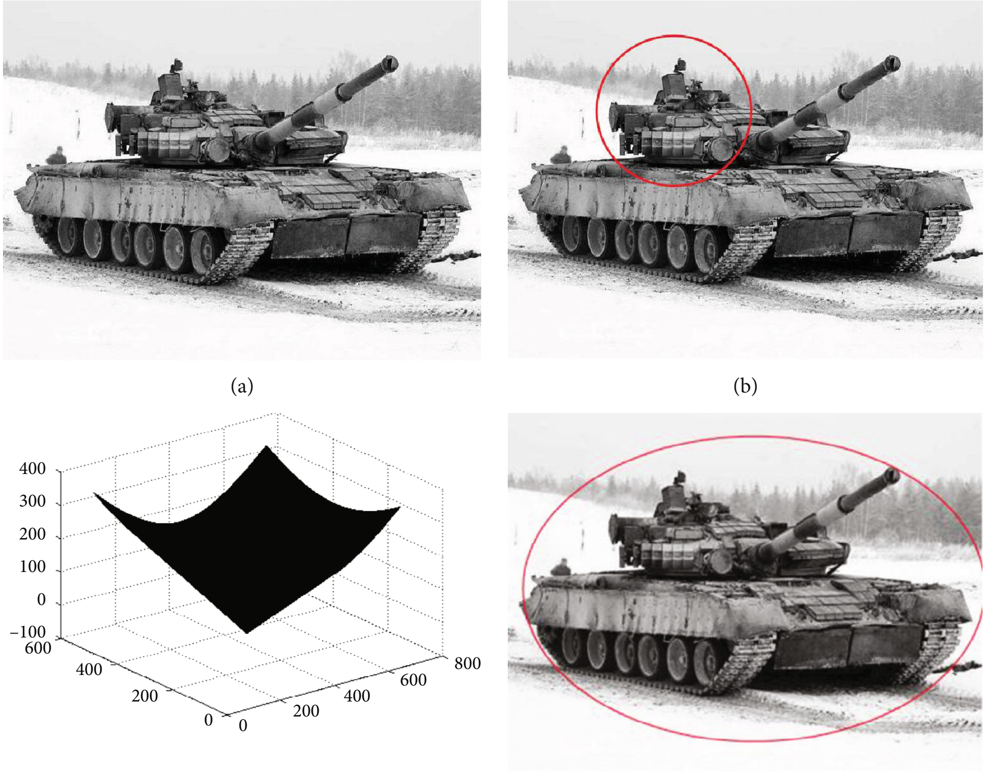
This paper selects the T72 heavy tank with obvious barrel characteristics as the segmentation object. The rectangle is set as the initial contour line through the CV model. The image needs to be corrected by Hough to accurately locate the damaged location.

Figure 5 shows the original image (a), the initial circular contour (b), the level set function (c) and the ellipse contour positioning result corrected by Hough transform (d). It can be seen that for the barrel with obvious characteristics on the tank, the rectangular initial contour of the CV model is modified by Hough change. The elliptical contour can better locate the whole part of the tank compared with the circular initial contour. Therefore, this article adopts Hough change. The revised CV model is used for the positioning of the T72 tank.

3.3. Infrared Lidar Positioning. Suppose T_{lidar}^{camera} is calibrated as the conversion matrix from the lidar coordinate system to the camera coordinate system, and the formula is as follows:

$$P_{camera}(X,Y,Z) = T_{lidar}^{camera}p_{lidar}(X,Y,Z) \tag{9}$$

According to formula (9), the relative three-dimensional coordinates of the target in the camera can be obtained.



(a)

(b)

(c)

(d)

Figure 5: Accurate positioning results of T72.

Let $O-x-y-z$ be the coordinate system of the camera, where O is the optical origin, and a point $P(X, Y, Z)$ in space corresponds to a point $P'(X, Y, Z)$ on the image plane, then:

$$\frac{Z}{f} = \frac{X}{X'} = \frac{Y}{Y'} \tag{10}$$

where f is the focal length of the camera.

Let $O-u-v$ be the pixel coordinate system, u and v axes are parallel to x axis to the right and y axis down, respectively. If the pixel coordinates are scaled α times on the u axis and β times on the v axis, the relationship between the coordinate P' and the pixel coordinates $[u, v]^T$ is:

$$\begin{cases} u = \alpha X' \\ V = \beta y' \end{cases} \tag{11}$$

Let $\alpha f = f_x$, $\beta f = f_y$, be rewritten into a matrix form through a homogeneous linear equation as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \overset{def}{=} \frac{1}{Z} KP \tag{12}$$

where K is the parameter matrix in the camera.

By formula (12), the real coordinates of the target in space can be obtained, and then it can be combined with the CapsNetv2 to realize the recognition and positioning of the radar target.

4. Algorithm Implementation

Figure 6 shows the specific process of the proposed radar target image recognition and positioning model based on the CapsNetv2.

(1) Input target images such as BTR70, BMP2 and T72 in the MSTAR dataset as different output vectors into the CapsNetv2

(2) The primary capsule layer is obtained through the convolution operation of the 1×1 reduction layer and the 3×3 convolution layer. Then the two image capsule layers are, respectively, trained and predicted to obtain $8 \times 8 \times 255$ and $16 \times 16 \times 255$ two prediction boxes

(3) The dynamic routing algorithm iterative formula (1) is updated to obtain the characteristic capsule layer

(4) According to formula (4), the maximum probability of the vector output modulus can be obtained, and the classification with the maximum probability of the radar target is obtained, and the two prediction boxes are mutually verified

(5) A clear and complete edge line is obtained through edge extraction of the identified target classification image

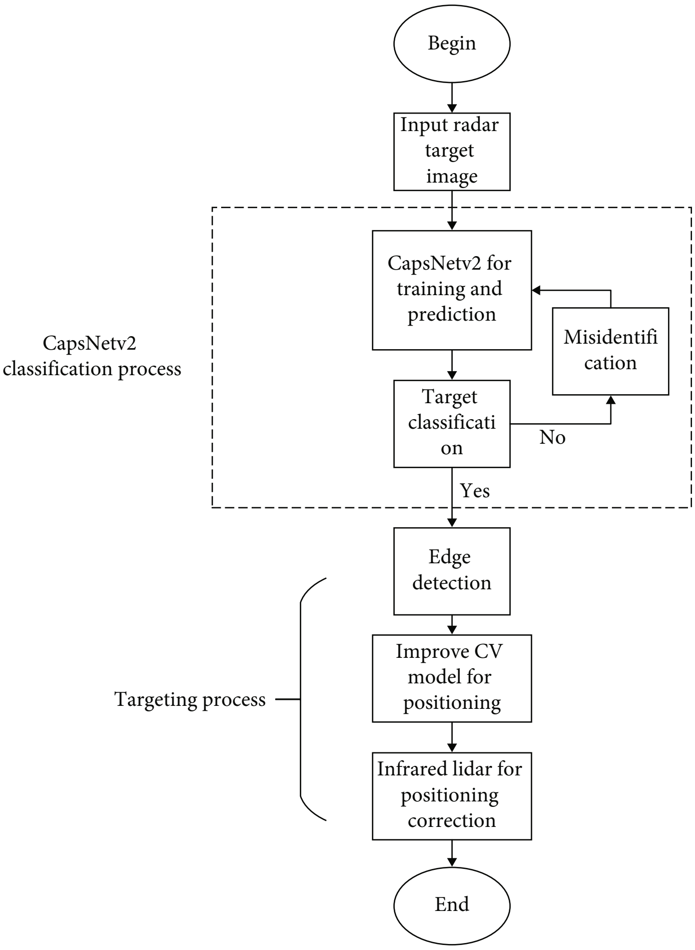

Begin
Input radar target image
CapsNetv2 for training and prediction
Misidentifi cation
CapsNetv2 classification process
Target classificati on
No
Yes
Edge detection
Improve CV model for positioning
Targeting process
Infrared lidar for positioning correction
End

Figure 6: Flowchart of the proposed radar target recognition and positioning.

(6) By improving the CV model and the precise positioning of the infrared lidar, the real three-dimensional coordinates of the target can be obtained

5. Experimental Verification

The method proposed in this paper is implemented using MATLAB R2014b and TensorFlow software. The 6000 radar target images with complex backgrounds in the MSTAR dataset are used as the training set, and 20% of the training set is randomly selected as the test set to verify the accuracy of the classification.

Example target images are shown in Figure 7. Among them, (a) and (b) are the T72 tanks in the sand and forest environments, (c) and (d) are the BTR70 armored vehicles in the sand and forest environments, and (e) and (f) are the BMP2 tank in the sand and forest environments.

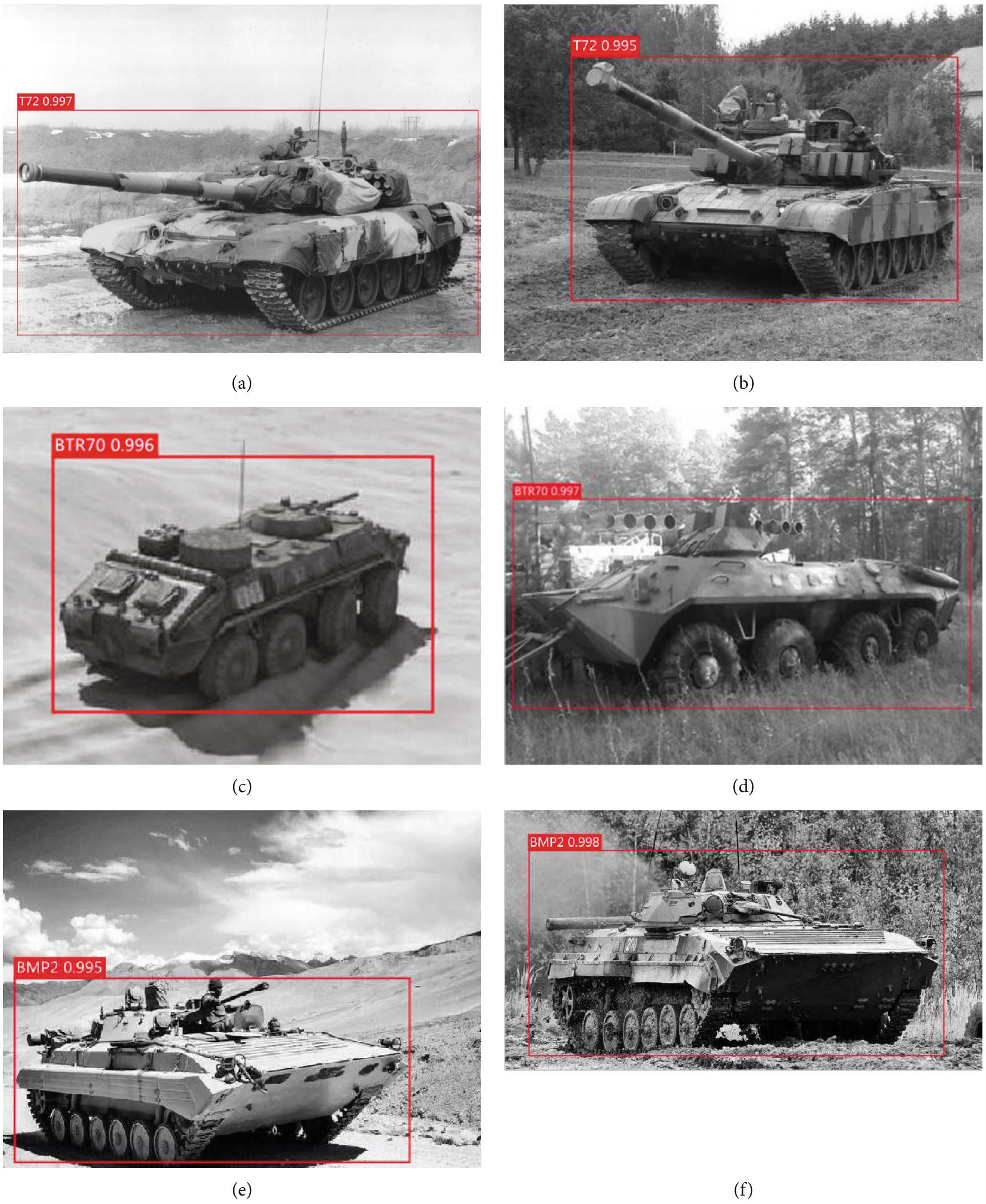
Figure 7 shows that the CapsNetv2 can accurately identify radar targets in different complex backgrounds and has good robustness.

5.1. Different Network Training Effects. To verify the practicability and recognition accuracy of the CapsNetv2, the radar target images with different complex backgrounds were used for training, and the performance of the CapsNetv2 was compared with that of the deep learning models of AlexNet and YOLOv4. The learning rate and the step length were changed and their performances were compared to select the best value of the parameter. The learning rate was 0.5, and the total number of steps was equal to 3000 as the optimal parameter. The training results are represented by the loss values, as shown in Figure 8.

Figure 8 shows the loss functions of AlexNet, YOLOv4 and CapsNetv2. The following conclusions can be drawn:

(1) The Loss value has shown an overall downward trend in the training of the three networks, and the first half of the decline is very fast. However, compared with the AlexNet and YOLOv4 networks, the initial loss value of the CapsNetv2 is only 0.9. This is because the AlexNet needs to scramble every time it reads the data, while in YOLOv4, the MSE loss itself has certain problems and needs to be replaced by IOU loss

(2) After the step size reaches 3000, the final loss function value of the CapsNetv2 is equal to 0.00015,



(a) (b) (c) (d) (e) (f)

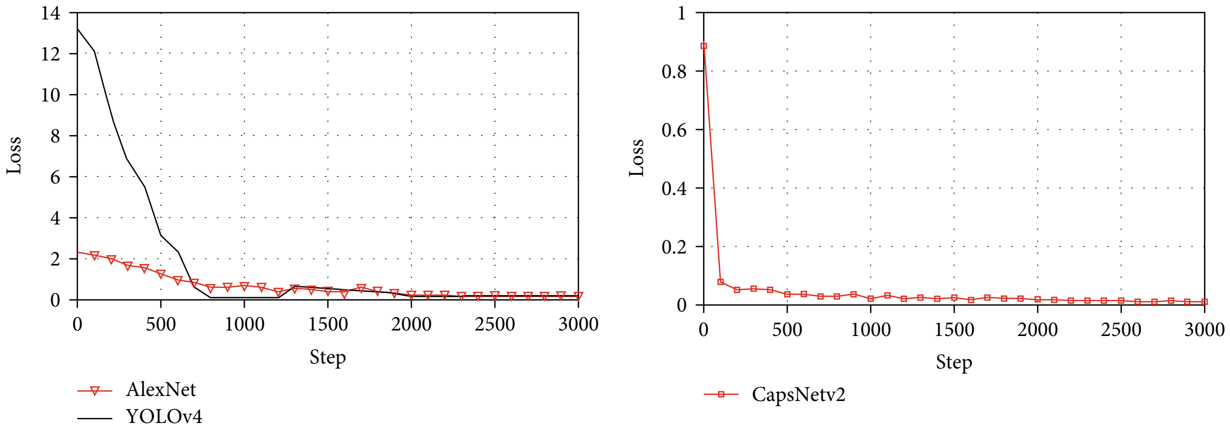
Figure 7: Identification effect.

which is about ten times smaller than the loss value of AlexNet. This is because the CapsNetv2 uses a simpler convolutional layer and a protocol layer, and two image capsule layers for training. At the same time, a model can be selected that has not been trained over-fitting, so that the model has good robustness

(3) Since there are no corner feature points in the AlexNet, and training on a small sample dataset cannot make the model more stable, once the loss value reaches 0.5, the model training becomes jittered and the training is terminated early. However, the CapsNetv2 retains different features information and training is more stable, which highlights the superior performance of the CapsNetv2

In order to verify and improve the learning performance of the capsule network, a database was used to randomly select the image data and compared with several other different algorithms. The results are shown in Figure 9.

It can be seen from Figure 9 that the recognition rate of the CapsNetv2 is higher than that of AlexNet and YOLOv4.



(a) The loss of AlexNet and YOLOv4

(b) The loss of CapsNetv2

Figure 8: Loss values of different training models.

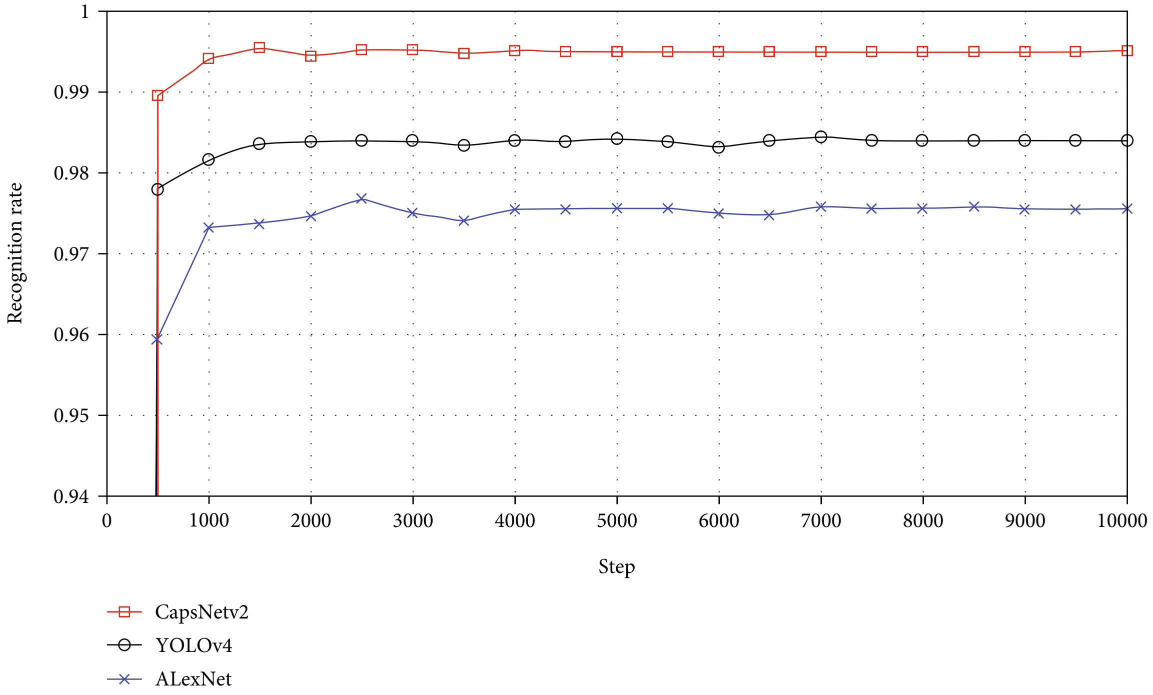


Figure 9: Contrasts of identification rates of different networks.

Table 2: Identification times of different algorithms.

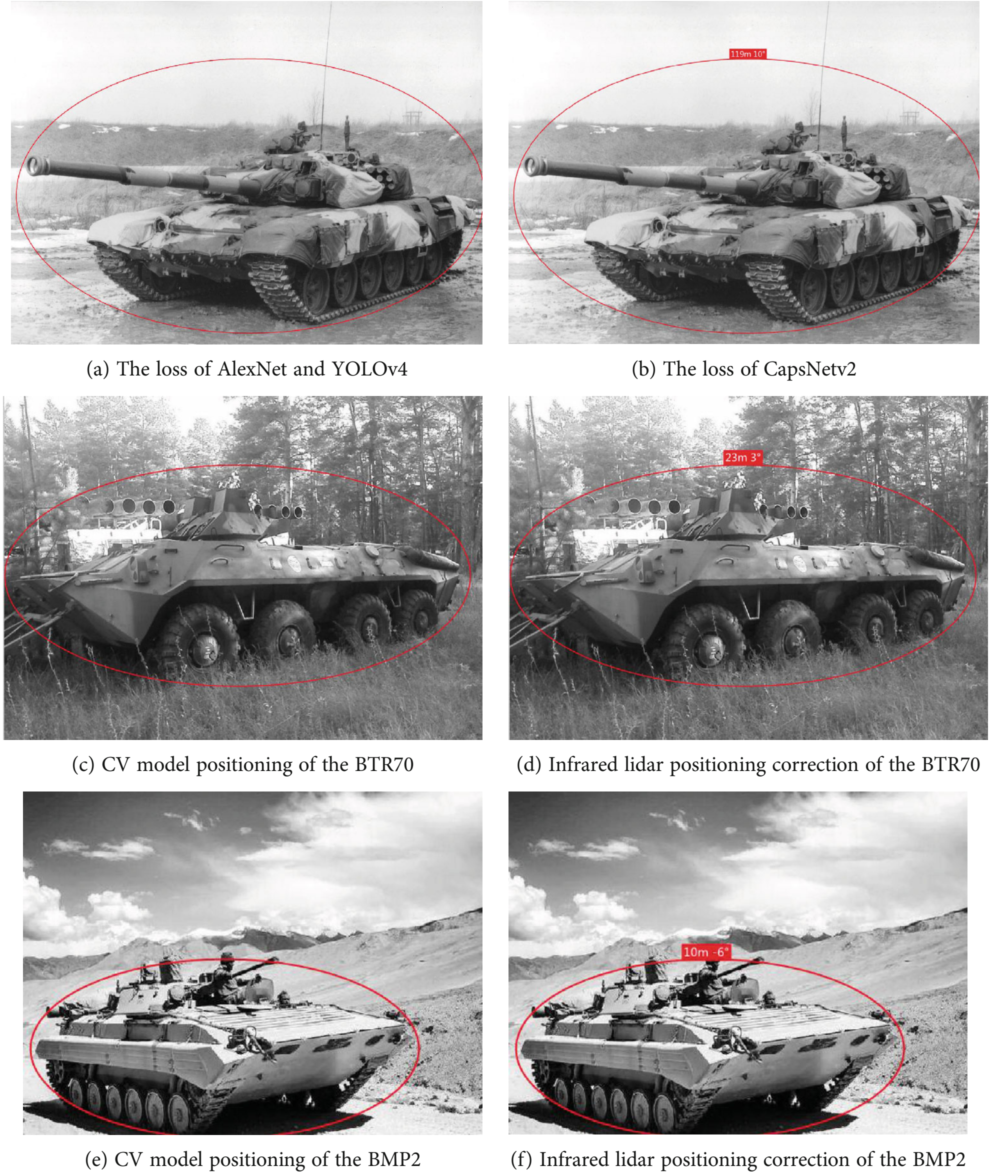
	400*600	600*800	800*100
AlexNet	7.6	8.2	12.4
Yolov4	8.4	8.9	11.6
CapsNetv2	2.5	2.9	4.3

Through continuous learning, the recognition accuracy reaches 99.5%. This is the result of learning by multiple vector capsules and retaining different feature vectors (such as amplitude and angle). At the same time, the two image capsule layers can be predicted separately, which reduces the phenomenon of over-fitting and the possibility of misclassification.

Table 2 compares the recognition times of different algorithms. It can be seen from the table that compared with AlexNet and YOLOv4, the CapsNetv2 has a shorter classification time and is more suitable for detecting radar targets in different complex backgrounds.

5.2. Target Positioning in a Complex Background. The improved CV model is used to locate the radar target image identified in the CapsNetv2. From Steps 4 to 6 in Figure 6, the precise positioning of the radar target includes edge extraction, CV model positioning and infrared lidar positioning correction. The positioning results are shown in Figure 10.

It can be seen from Figure 10 that the improved CV model proposed in this paper and then corrected by the



(a) The loss of AlexNet and YOLOv4

(b) The loss of CapsNetv2

(c) CV model positioning of the BTR70

(d) Infrared lidar positioning correction of the BTR70

(e) CV model positioning of the BMP2

(f) Infrared lidar positioning correction of the BMP2

Figure 10: The results of radar target positioning.

Table 3: Comparison of positioning accuracies of different methods.

Positioning method	Target precise positioning accuracy rate/%
The CV model+ infrared lidar positioning	97.5%
Infrared lidar imaging	92.6%
Local feature analysis	90.2%

infrared lidar can accurately locate the radar target images in different backgrounds.

From the different types of radar targets identified by the CapsNetv2, a group of 20 images were randomly selected for precise positioning of the radar target, and compared with the infrared imaging method and the local feature analysis method [18]. Table 3 compares the positioning accuracies of different methods.

Table 3 shows that the proposed radar target recognition and positioning method that combines the CapsNetv2 and the CV model is more suitable for small sample learning and has better training effects. Thus, the proposed method has higher positioning accuracy than the other methods and is suitable for different complex backgrounds.

6. Conclusion

With the continuous improvement of military warfare technology, the real-time detection of different radar targets under different complex backgrounds is particularly important. This paper proposes a radar target detection model based on the CapsNetv2 and the improved CV model modified by infrared lidar. The proposed model can identify radar targets in complex backgrounds and accurately locate their positions. The target positioning algorithm is simulated and experimentally verified. The following conclusions can be drawn:

(1) The CapsNetv2 has strong self-learning and adaptive capabilities, and has a good training effect for small sample sets. It can effectively detect different types of radar targets and suppress interference caused by complex backgrounds. The recognition rate reaches as high as 99.5%. The reason is that the input of CapsNetv2 is a vector, which retains the feature information of the target to the greatest extent, and through the double-layer image capsule layer for training, it effectively reduces the over-fitting phenomenon and can more accurately classify different radar targets

(2) The radar target image identified and classified by CapsNetv2 is segmented by the improved CV model, and finally corrected by the infrared laser mine, which can accurately locate the position of the target. The accuracy rate of the proposed method reaches 97.5%, which is more suitable for the precise positioning of radar targets than the other methods

The method proposed in this paper can better realize the radar target recognition under complex background and can provide accurate location information to meet the requirements of real-time inspection. However, the training time of the CapsNetv2 for a large number of images is relatively long. Thus, reducing the training time of the capsule network will be the focus of future research.

Data Availability

The [MSTAR] data and the [CaspNetv2 solution] data used to support the findings of this study were supplied by [Jiaxing Hao] under license and so cannot be made freely available. Requests for access to these data should be made to [Jiaxing Hao, 873328461@qq.com].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Y. Shen and X. Y. Huang, "Doppler weather radar wave echo identification and effect verification based on Bayesian classifier," *Marine Sciences*, vol. 44, no. 6, pp. 83–90, 2020.

[2] J. R. Liu, "Weather radar rapid identification and early warning of heavy rainfall based on probability statistical model," *Information Recording Materials*, vol. 20, no. 4, pp. 157-158, 2019.

[3] Z. Zhang, C. Zou, P. Han, and X. Lu, "A runway detection method based on classification using optimized Polarimetric features and HOG features for PolSAR Images," *IEEE Access*, vol. 8, pp. 49160–49168, 2020.

[4] Z. Zhou, Q. M. J. Wu, S. Wan, W. Sun, and X. Sun, "Integrating SIFT and CNN feature matching for partial-duplicate image detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 5, pp. 593–604, 2020.

[5] Y. Feng, "Mobile terminal video image fuzzy feature extraction simulation based on SURF virtual reality technology," *IEEE Access*, vol. 8, no. 99, pp. 156740–156751, 2020.

[6] W. Q. Liu, Z. G. Liu, A. Núñez et al., "Multi-Objective Performance Evaluation of the Detection of Catenary Support Components Using DCNNs," *IFAC PapersOnLine*, vol. 51, no. 9, pp. 98–105, 2018.

[7] S. Huang, Y. Zhai, M. Zhang, and X. Hou, "Arc detection and recognition in pantograph-catenary system based on convolutional neural network," *Information Sciences*, vol. 501, pp. 363–376, 2019.

[8] Z. Li, L. Yang, and L. Wang, "Detection approach based on an improved faster RCNN for brace sleeve screws in high-speed railways," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 4395–4403, 2019.

[9] Y. Gan, H. Wu, and N. Xiao, "Cross-modal attentional context learning for RGB-D object detection," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1591–1601, 2019.

[10] Z. Zhou, Y. Cao, M. Wang, E. Fan, and Q. M. J. Wu, "Faster-RCNN based robust coverless information hiding system in cloud Environment," *IEEE Access*, vol. 7, pp. 179891–179897, 2019.

[11] Y. Guo, Z. Pang, and J. Du, "An improved AlexNet for power edge transmission line anomaly detection," *IEEE Access*, vol. 47, no. 8, pp. 97830–97838, 2020.

[12] S. Du, P. Zhang, and B. Zhang, "Weak and occluded vehicle detection in complex infrared environment based on improved YOLOv4," *IEEE Access*, vol. 9, pp. 25671–25680, 2021.

[13] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," *Advances in neural information processing systems*, vol. 30, pp. 90–96, 2017.

[14] J. Zh and X. Chen, "An improved overlapping handwritten digit recognition method for capsule networks," *Laser Journal*, vol. 40, no. 7, pp. 43–46, 2019.

[15] C. Xiang, Z. Lu, and W. Zou, "MS-CapsNet: A Novel Multi-Scale Capsule Network," *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1850–1854, 2018.

[16] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion Recognition from Multiband EEG Signals Using CapsNet," *Sensors*, vol. 19, no. 9, p. 2212, 2019.

[17] H. Y. Zhao, Z. J. Liu, and H. Zhang, "Multi-moving target segmentation based on improved CV model," *Chinese Journal of Scientific Instrument*, vol. 31, no. 5, pp. 1082–1089, 2010.

[18] Q. F. Lai, J. Yang, and L. W. Han, "Insulator automatic identification and defect diagnosis model based on YOLOv2 network," *Chinese Power*, vol. 9, no. 22, pp. 1–10, 2019.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 9608394, 10 pages
https://doi.org/10.1155/2022/9608394

Review Article

Smart Agriculture for Sustainable Food Security Using Internet of Things (IoT)

Taimoor Qureshi,[1] Muhammad Saeed ,[1] Kamran Ahsan,[2] Ashfaq Ahmad Malik,[3] Emaduddin Shah Muhammad ,[4] and Nasir Touheed[5]

[1] *Department of Computer Science, University of Karachi, Karachi 75600, Pakistan*
[2] *Department of Computer Science, Federal Urdu University for Arts, Science and Technology, Karachi 75600, Pakistan*
[3] *Bahria Enterprise Systems and Technologies (BEST), Bahria Complex II, MT Khan Road, Karachi, Pakistan*
[4] *College of Computing and Information Sciences, Karachi Institute of Economics and Technology, Karachi 75600, Pakistan*
[5] *Department of Mathematics and Computer Science, Institute of Business Administration (IBA), Karachi, Pakistan*

Correspondence should be addressed to Emaduddin Shah Muhammad; shahmuhammademad@gmail.com

Received 28 January 2022; Revised 26 April 2022; Accepted 6 May 2022; Published 25 May 2022

Academic Editor: SK Hafizul Islam

Internet of Things (IoT) is being used in various parts of human life (domestic and commercial) to provide ease in living, safety, increase productivity, monitoring, and resource optimization in various industries. Agriculture is one of them, where IoT and robots are being used before and after the cultivation process, from preparing land for cultivation to supplying them to the consumer market. These domains include crop monitoring, smart irrigation, pest monitoring, and smart pest control, harvesting, and safely supplying them in the consumer market by maintaining the quality and integrity of the final product. Pakistan is an agricultural country, where it stands in terms of advanced agriculture technology. In this review, we discussed the major IoT ecosystem components. What are the most practiced smart agriculture techniques and their benefits and some widely used applications of IoT in agriculture? Through this overview, we are trying to highlight the potential of IoT in agriculture for sustainable food security for Pakistan.

1. Introduction

Food security is becoming the biggest challenge to the world as by 2050 the world's population is predicted to reach 9.7 billion which is 20.6% of the current population [1]. Also, the rate of urbanization is accelerated, with 68% of the world population expected to be urbanized in 2050 [2], which was 54% until 2018 [3] which will reduce available arable land. On the other hand, not all the land on earth is arable because of some factors like soil quality, climate, topography, and high variability factors within the homogeneous land. Furthermore, the rate of arable land declining surpasses the rate of recovery because of pollution, soil erosion, and land degradation [4]. All these issues are inclined toward the adoption and advancement of agriculture. Pakistan is an agricultural country, and agriculture contributes 19.5% to the country's Gross Domestic Product (GDP) and provides 38.5% of employment of the national labor force [5]. Agriculture in Pakistan is declining because of water scarcity, old practices, and uneducated farmers. Unfortunately, Pakistan follows downward in agriculture, mostly focusing on chemical fertilizers and pesticides and genetically modified organism (GMO) seed, although GMO crops proved to be beneficial as cash crops for Pakistan; it is harmful to biodiversity, especially in Pakistan producing 95% of yield as Bt-cotton and hybrid maize [6]. Pakistan is taking several initiatives for food security [5]. But there is no plan found for adaptation for technology; we did not even find any concrete interagriculture-information technology research in the last decade. Even the most viable solution could be the adoption of IoT systems in the processes of agriculture. The reasons could be Pakistan being a developing country, upfront cost, lack of budget for research and development (R&D), and undefined agriculture policies. After the emergence of the IoT, it changed the dimensions of different sectors and industries like health care [7], car monitoring systems [8], smart agriculture [9], smart cities [10], and smart homes [11]. The IoT is a system that does not require any machine-to-machine and human-to-machine interaction to perform any task, possesses the ability to transfer data over a network, and consists of multiple interrelated computing devices and digital and mechanical machines with unique identifiers (UIDs), which is relatively cheap as compared to previous existing technologies but still required researchers and industry attention.

IoT-based systems provide some major capabilities such as data acquisition and communication infrastructure (used to connect smart objects to end-user applications through the Internet), cloud-based intelligent data analysis, decision-making, end-user interface, and operation automation. These capabilities are opening new dimensions in the field of agriculture. This paper is a brief overview of the application of the Internet of Things in smart agriculture for sustainable food security. The overview consists of some practical case studies, white papers, and articles about how IoT could provide sustainable food production with minimal resources and what are challenges to begin with and what could be the possible approach to implementing the IoT-based ecosystem.

2. IoT Ecosystem's Equipment and Technology

IoT ecosystem is a combination of several technologies and equipment that are embodied by integrated systems that work seamlessly in their operations (Figure 1), representing the IoT system architecture. Data acuisition is done from the sensors and the data is transferred to the cloud architecture where desicions are taken to perform operations in the field based to provide insight for the end-user application. All components of the system work independently without having any human-to-human or human-to-machine interaction. Here are some common components that make this whole process seamless and integrated.

2.1. IOT Sensor Components/Technology. Smart agriculture cannot be possible without the sensor's technology. Sensors are used to gather and measure different factors and variables of environments that could affect crop yield. The success of precision agriculture is based on accurate sensor data acquisition for crop- and soil-specific management [12]. Almost all the equipment and vehicles (i.e., tractor, harvester, unmanned aerial vehicle, and sensor device) are equipped with remote sensing facilities like Geographic Information System (GIS) and Global Positioning System (GPS) for precise and autonomous site-specific operations. A wide range of IoT sensors available for monitoring applications can be classified into two categories. The first one is intelligent multipurpose imagery sensors, which could be embedded on unmanned aerial vehicle (UAV), rails, and fixed position components and could involve remote sensing [13]. Combined with deep learning, these sensors can reach their full potential and are capable of soil and vegetation/crop mapping, crop phenology, crop height, estimation of yields, fertilizers' effect and biomass, plants water stress detection and drought conditions, pest detection and management, weed detection, and greenhouse monitoring [14]. The second type of sensor is more commonly used and specific to their use case and can be deployed at various locations on the field. The most common sensors are airflow, soil moisture, electrochemical, capacitive humidity, position, mechanical, optical, and temperature sensors. Table 1 represents sensor working and their use cases. Furthermore, there are some worth mentioning factors that make IoT sensors suitable for smart agriculture: (1) computational efficiency, (2) cost, (3) coverage, (4) durability, (5) memory, (6) portability, (7) power efficiency, and (8) reliability.

2.2. Unmanned Aerial Vehicles (UAVs). Apart from the IoT ecosystem, UAVs is itself an emerging and self-existing technology that is a combination of various other technology stacks such as robotics, on-board computing, artificial intelligence (AI) [21], information and communication technology (ICT), IoT, and battery. The reason behind the popularity of UAVs is that it is filling the gap of limitation of remote sensing imaging through satellite because of weather and cloud penetration and on-ground limitation of robots because of uneven plains, obstacles, and speed. UAVs provide imaging with high resolution using hyperspectral, multispectral, and Red Green Blue (RGB) cameras [22]. It entails more accurate details of the field at a much cheaper cost.

UAVs are workable in monitoring as well as in the action phase of the application. Common usage of UAVs in two major phases of precision agriculture is as follows. First one is monitoring, where applications are soil and crop mapping and sampling [23], yield forecasting [24], weed detection [25], pest and disease detection [26], and soil and crop stress assessment [27, 28]. The second one is an action phase where applications are sowing seed [29], spraying herbicides [30], pesticides [31], and fertilizer [32].

UAVs have two main types shown in Figure 2: fixed-wing UAVs and rotary-wing UAVs. Fixed-wing UAVs are more similar to airplanes and more or less follow similar phenomena for flying; fixed-wing UAVs are more favorable to work on large areas because of the capability of long range, high speed and altitude, and crash tolerance. Rotary-wing UAVs have further classification such as helicopter and multirotary type; commonly, multirotary UAVs are named after their number of rotaries, i.e., four-rotary UAVs as quadcopter [33, 34], six-rotary UAVs as hexcopter [35], and eight-rotary UAVs as octocopter [36]. Fly in a hovering manner similar to a helicopter. Rotary-wing UAVs have more advantages than fixed-wing UAVs such as being easy to set up and operate, low altitude flight, precise location operation ability, no wind planning required, and being fully autonomous for daily agriculture operations.

With all the ease, UAVs have some limitations as well. The technical limitations of UAVs are low battery time and efficiency, payload, communication distance, and low flight time. Fixed-wing UAVs can communicate up to 100

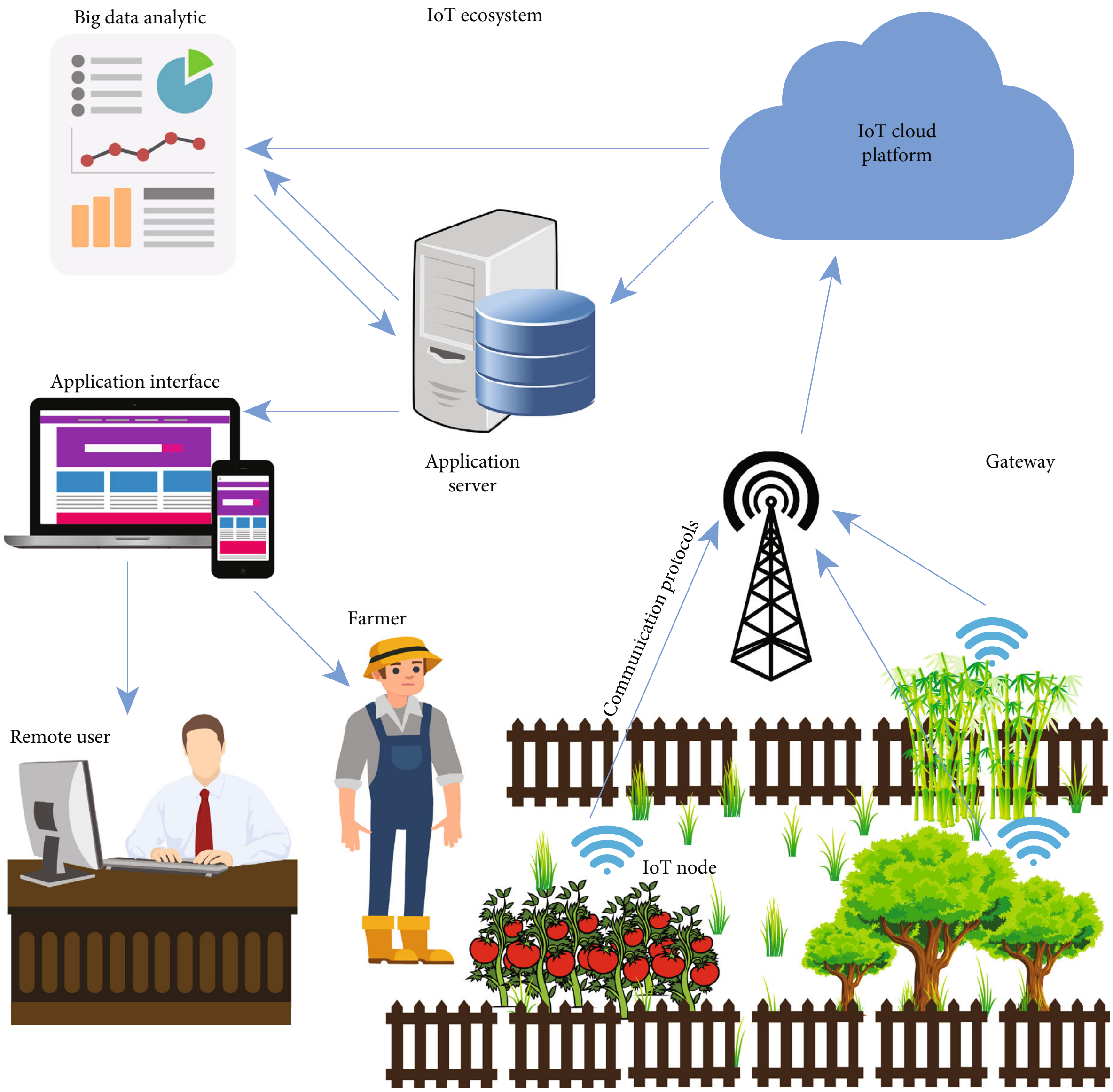


Figure 1: IoT ecosystem for agriculture.

kilometers, and the average flight time is about 5 h [37]. Battery efficiency researchers are working to develop a more efficient hybrid battery and battery management and optimization techniques [38, 39].

2.3. Communication Technologies. Smart agriculture is impossible without the inclusion of ICT. Data has no purpose if it cannot be sent to some database or cloud for further computing and analysis; it is considered the backbone of smart agriculture. There are several classifications in communication technologies based on their communication range, data rate bandwidth, power consumption, licensed or unlicensed spectrum, frequency band, and subscription prices. Every communication technology works better than others in different application scenarios which depend on what aspect is most important in that particular application. For each application scenario, some work best or some work worst. For example, Zigbee communication technology is more suitable for greenhouse agriculture monitoring, and Narrowband Internet of Technology (NB-IoT) and long range (LoRa) are more suitable for field precision agriculture [40].

Choosing communication technology for smart agriculture depends on multiple factors. Some factors are more prominent than the other; e.g., unlicensed spectrum technologies could have better bandwidth but come with some pitfalls such as radio frequency interference, insecure communication, infrastructure setup cost, and low range connectivity. Radio frequency identification (RFID), Bluetooth, Wi-Fi, and Zigbee are examples of unlicensed spectrum technologies. On the other hand, licensed spectrum technologies are reliable, provide accessibility for large areas, are secure, and have less infrastructure cost but have a subscription for data transmission and low data rate bandwidth as compared to the unlicensed ones. A survey suggests that ZigBee, Wi-Fi, and cellular technologies are

Table 1: Agriculture sensors and use cases.

Type of sensors	Functionality	Use case
Capacitive humidity [15]	Use of electrodes with hygroscopic dielectric material, to detect air moisture by electrical permittivity	Monitor humidity of soil and air in a controlled environment for irrigation and fertilization
Electrochemical [16]	Use of electrodes to detect specific ions in the soil	Properties like the macro and micro nutrients in the soil, salinity, and pH are measured
Imagery and remote sensing [13]	Use of multispectral cameras, hyperspectral cameras, IR cameras, and digital cameras to generate a digital image	Monitor anomaly, weed, disease, pest, and crop mapping with the help of spatial, spectral, and temporal resolution
Mechanical [17]	Use to measure soil mechanical resistance to indicate the soil compaction	To detect the force used by the roots in water absorption and useful for irrigation and soil inspection
Optical [18]	Use of light to measure soil properties	Light reflectance phenomena used to determine clay content, color, minerals and their composition, organic matter, and moisture content of soil
Position [19]	Use of Global Positioning System (GPS) satellites to determine the latitude, longitude, and altitude	The GPS provides precise positioning and backbone for GIS that is used for geospatial analysis
Soil moisture [20]	Use of electrodes to assess moisture levels by measuring the dielectric constant in the soil	Time-domain reflectometry (TDR) for nondestructive continuous monitoring of soil water content

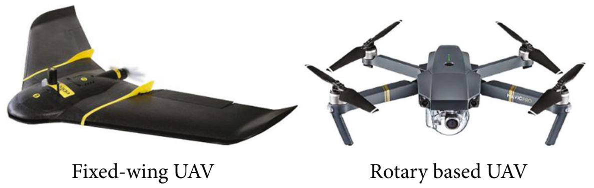


Figure 2: Two main types of UAVs.

more popular among researchers for agriculture applications. About 45% of Zigbee, 25% of Wi-Fi, and 20% of cellular or multihopping technologies are utilized by the researcher for their agriculture-related experiments [41]. Furthermore, NB-IoT and Long-Term Evolution Machine (LTE-M) are relatively new Low-Power Wide-Area (LPWA) technologies and could capture more attention as the 3rd Generation Partnership Project (3GPP), the standard group specifying the 5th generation mobile network (5G) and other wireless networking standards, has affirmed that these technologies are going to be a part of 5G and will be the only LPWA-supported 5G technology [42]. Table 2 shows some widely used communication technologies in smart agriculture.

3. Smart Agriculture Methods and Techniques

Humans have been trying to improve food production for centuries to meet food requirements. To achieve this task, they are adopting and applying different advanced agriculture techniques. After the emergence of IoT, advanced agriculture techniques like vertical farming, hydroponics, and phenotyping significantly improve their performance by utilizing IoT and becoming an essential part of them. It is cost-effective and can help us in the efficient management of resources ranging from input resources, labor resources, and operational resources and also provide a high yield. Applications of technology are geospatial and temporal sampling and mapping, disease and pest monitoring, smart irrigation, and fertilization. Commonly used technologies and equipment are sensors, UAVs, IoT-based machinery and communication, etc.

3.1. Precision Agriculture. Precision agriculture existed long ago but it was not viable for small and medium farmers and even not viable for large farmers in developing countries like Pakistan, the challenges ahead like climate change, a gap in demand and supply of food, urbanization, and declining arable land are unavoidable. In this situation, the emergence of IoT is enabling a new dimension in precision agriculture, consisting of several already existing technologies such as WSN, RFID Gateways, cloud computing, communication protocols, middleware components and end-user inferface [49]. Communication protocol, middleware components, and end-user interface [49].

Precision agriculture is focused on the utilization of natural resources efficiently and protecting the natural environment. There are four steps to implement precision agriculture: characterizing the extent and scale of variability in soil and crop attributes, interpreting the significance and causes of variability, managing variability on a spatial and temporal basis, and monitoring the outcomes resulting from the variability management practices [50] that could only be done efficiently using the IoT. To wind up the discussion, Figure 3 illustrates the common hurdle in the adoption and implementation of technology in precision agriculture; on the other hand, Figure 4 indicates the key advantages of IoT in precision agriculture (PA). The precision agriculture adoption starting point could be yield monitoring by gathering data to develop spatial and temporal feature databases for management of land for interpretation and yield mapping.

3.2. Greenhouse Farm. Greenhouse farming is more or less similar to precision farming with some subtle differences and purposes. The major difference is that greenhouse

Table 2: Communication technology.

Name	Spectrum	Transmission range	Network	Frequency bands	Data rate	Pros	Cons
SigFox [43]	Licensed	Rural: 30-50 km Urban: 3-10 km	LPWA	868 or 902 MHz	100 bps (UL) 600 bps (DL)	Consumes low power, wide coverage area	Mobility is difficult, low data rate
LoRaWAN [44]	Licensed	<20 km	LPWA	Various, sub—GHz	0.3–37.5 kbps	Device works well even in motion, longer battery life	Low data rates, long latency time
3GPP NB-IoT [45]	Licensed (cellular)	<35 km	LPWA	450 MHz, 3.5 GHz	250 kbps	Coverage, quality of service	Network and tower handoffs, difficulty in sending large amounts of data
3GPP LTE-MTC (Cat-M1) [46]	Licensed (cellular)	<5 km	WWAN	1.4 MHz	200 kbps	Coverage, connectivity to any service, faster data rates	Costs, low data rate, no high speed as other cellular technology
Wi-Fi [47]	Unlicensed	6–50 m, 1000 m	WLAN	2.4/5 GHz, various, sub—1 GHz	2 Mbps–7 Gbps, 78 Mbps	Access and availability, flexibility, cost savings	Security issue, radio frequency interference, coverage
Bluetooth [48]	Unlicensed	<100 m	WPAN	2.4 GHz	2 Mbps–26 Mbps	Low power consumption, cost	Low bandwidth, not secure
Zigbee [48]	Unlicensed	<1 km	WHAN	2.4 GHz	250 kbps	Easy and simple setup, long battery, scalable	Not secure, low coverage

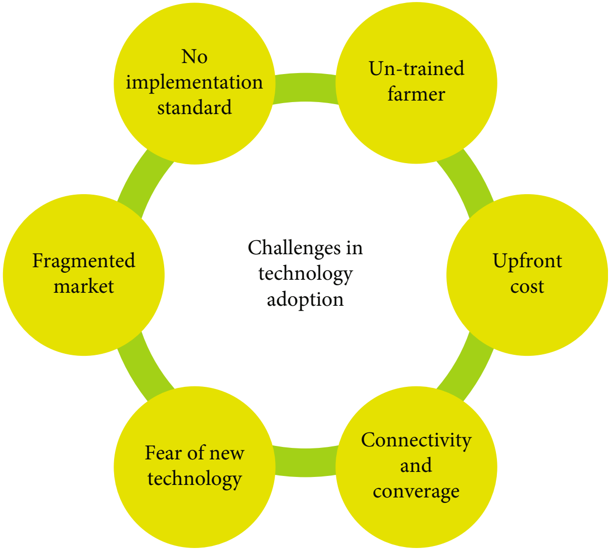

No implementation standard
Un-trained farmer
Fragmented market
Challenges in technology adoption
Upfront cost
Fear of new technology
Connectivity and converage

Figure 3: Challenges in technology adoption.

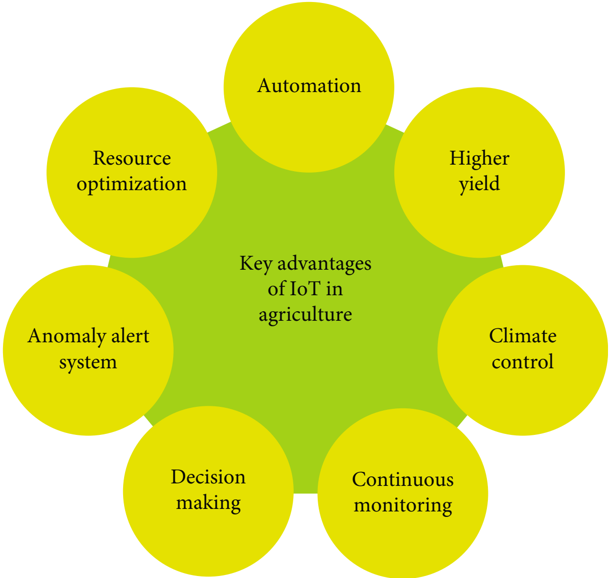


Figure 4: Key advantages of IoT in agriculture.

farming is done in a closed or isolated environment or space where environmental parameters are controlled and managed by smart systems. Although greenhouse farming is not new, the information technology and IoT found applications that are aligned with greenhouse techniques like temperature, humidity control, and monitoring in the shed. The precise and continuous monitoring and controlling cannot be achieved without IoT and smart systems. Greenhouse space is comparatively smaller than open-field agriculture but yields more productivity than traditional methods. The Netherlands is one of the small countries and the second largest exporter of agricultural goods utilizing greenhouse farming and hydroponics variety of crops [51]. Through greenhouse farming, we can even utilize desert space for sustainable farming [52].

3.3. Urban Farming. Urban farming is another revolutionary idea and a relatively new concept, with the fusion of different methods such as rooftop farming, indoor farming, vertical farming, hydroponics, aquaponics, and aeroponics; as already mentioned in the article, more population is shifting toward urban areas, and urban areas are the large consumer of food products. On the other hand, climate change and water scarcity affect agriculture harshly and are serious challenges for sustainable food security. Also, some other challenges are long distance between food producers and food consumers so there is a transportation and chain supply expense that impacts food quality, causing extra pollution by transport vehicles. Urban farming is a solution where people can grow food in their proximity to have fresh and cheap food. Urban farming is completely dependent on the precise control environment and system to grow day and night and a whole year without any season and weather impact. We can grow food in a closed box without any sunlight the whole year [53]. By realizing the potential of urban farming, Paris is taking a shift with its largest rooftop farm with an expectation to have 30 different types of plants, furthermore aiming to have more than 100 hectares of rooftops farms [54]. Similarly, the 100 ft below the ground abandoned space is well utilized for vertical farming in London [55].

4. Applications of IoT in Agriculture

In the smart era of agriculture, almost all the agriculture processes are data-driven that are acquired by continuous monitoring of on-site IoT devices. Applications of smart agriculture that can only be done by utilizing IoT are geospatial and temporal mapping and sampling [56], smart drip and sprinkler irrigation, pest and pathogen monitoring and controlling, yield assessment, precision fertilization, and environment maintenance. All these applications are briefly discussed below:

4.1. Geospatial and Temporal Mapping and Sampling. The simple and crucial application of precision agriculture can be used for crop field assessment and mapping. Applications that depend on the geospatial and temporal sampling are weed management systems, water stress assessments, and vegetation indexes. Also, it can be used for spatial variability assessment using GIS [57].

Geospatial can be done using remote sensing, aerial surveys using planes, and remote imagery using UAV. Initially, it was expensive and not as efficient as today. Because of satellite remote sensing and cloud distortion, UAV is cost-effective and way more efficient and could be adopted as the first step for precision farming, even by farmers in developing countries like Pakistan.

4.2. Smart Irrigation. As the world has been facing the challenge of water scarcity, Pakistan is becoming water scarce from a water-stressed country [58]. By using IoT and smart systems, weather adaptive smart irrigation systems can be implemented and reduce the usage of precious resource water [59]. Smart irrigation is designed to irrigate only if necessary, depending on the crop and soil stress level.

UAV is a great tool to deal with the variability factor of water stress; sprinkler irrigation can be done using UAV for precise irrigation on the spot [60].

4.3. Pest and Weed Management System. Pest, weeds, and pathogens can affect the crop harshly and may reduce productivity by up to 30% only by weeds [61]. On the other hand, pesticides and herbicides also reduce the profit and degrade the product quality as well which is a big concern for the consumer. IoT and smart systems can assess the disease, pest, and weed in the crop in the early stages and can inform the farmer, also capable of eradicating the pest and pathogens by precise targeting with pesticides and herbicides; smart vehicles [62] can also be used for this purpose.

4.4. Yield Assessment. Yield assessment is the most essential part of smart agriculture. For any type of assessment, data acquisition is the first step. Precise and continuous monitoring for the biotic and abiotic factors is only possible by IoT, WSNs, and UAV imagery. All these devices generate enormous amounts of unstructured data. The acquired data can be utilized for the early prediction of disease [63], crop prediction [64], and harvest planning [65]. Through these applications, farmers can reduce their labor cost and operation cost, can do the error-free assessment for diseases and pests, estimate the revenue and profit, and schedule and plan a more suitable harvesting period that results in less input cost and more profitability in the long run.

4.5. Precision Fertilization. Another most important application of IoT for agriculture is that it can save money and the environment at the same time. Imbalance fertilization can cause multi-impact damage; i.e., sometimes plants require fewer nutrients; thus, excessive fertilizer may drain away or cause salinity in the soil which may rotten the plant, decrease productivity, cost you extra, and also cause climate change by evaporation. On the other side, if the plant required more nutrients, but was provided less, that also caused a decline in productivity and growth. Furthermore, fertilizer proportions of different elements such as nitrogen (N), potassium (K), and phosphorus (P) and water also matter because proportion depends on plant type, soil type, and weather; otherwise, crops cannot be productive. One more aspect is the variability which can only be handled through precision monitoring and mapping of land and crop. Smart IoT-based agriculture systems provide an optimal estimation of nutrient requirement [66] and reduce the labor cost and input costs.

5. Challenges and Solutions of Using IoT Devices in Smart Agriculture

The most significant limitation of using IoT devices is battery life and especially when using UAVs in agriculture is the flight time along with the battery. A thorough study has been performed, and this one is an open area for various solutions [67]. Many researchers have worked to reduce these hurdles and proposed and tested their solutions.

To enhance the UAVs' fly time, ultralightweight WPT systems were proposed [68]. The system is flexible enough to handle air-gap geometrical changes. The system is capable of charging UAVs in midair and extending the flight time to around 7 minutes. Their system can charge drones wirelessly with 10 W. In another work [69], a wireless charging system for UAVs was developed using capacitive power transfer (CPT) technology. This system can charge UASs on wide charging areas. Their system's emitting side is comprised of a circuit, transformar and inductors.The receiving side is comprised of all the small devices using semiconductor elements for a DC-DC converter and charge controlling IC. Their prototype system works on around 12 W and provides more than 50% efficiency.

While considering the magnetic resonant coupling technique due to its efficiency and capability of high power transfer, [70] has proposed and developed a wireless charging system for UAVs used in agriculture fields. In their experiments, they achieved maximum transfer power and efficiency by using FSC coil with 150 coil turns in the transmitter circuit and the MTC comprising 60 coil turns in the receiver UAVs.

Another major hurdle of using UAVs in smart agriculture is path loss while communicating wirelessly due to the surrounding environment, and an accurate path loss model is essential for smart agriculture applications to make sure wireless data communication without unnecessary packet loss among each component of the system. [71] has proposed and tested two improved models. Their simulation results show that the hybrid exponential and polynomial and particle swarm optimization models noticeably improved the coefficient of determination (R^2) of the regression line, with the mean absolute error (MAE) found to be 1.6 and 2.7 dBm for both algorithms. The Wireless Underground Sensor Network (WUSN) faces the same path loss issues, and [72] has proposed and developed a system based on an accurate prediction of the Complex Dielectric Constant (CDC) to handle the path loss for precision agriculture known as WUSN-PLM. Their results show that the WUSN-PLM outperforms the existing path loss models in different communication types and provides 87.13% precision and 85% balanced accuracy on real cheap sensors.

6. Conclusion

In this review, the importance of IoT and its successful applications in agriculture is presented along with challenges and solutions. IoT's ecosystem and use of UAVs and their various types and benefits, different communication protocols, and their pros and cons for agriculture applications are also covered. We have also discussed how IoT can be applied in different smart agriculture techniques such as precision agriculture, greenhouse farming, and urban farming with some case studies of food product leaders. Furthermore, we have discussed some widely used applications of IoT in agriculture. Consider geospatial and temporal mapping and sampling of crops as the first step for smart agriculture for any developing country like Pakistan where the upfront cost is a big issue for farmers. After the overview, we conclude that Pakistan as agriculture taking several initiatives to cope with climate change, water scarcity, food insecurity.The usage of advanced information technology, artificial intelligence which is somewhat missing in local developed projects. Thus, Pakistan and other developing countries should bear the upfront cost and R&D in the inter-agriculture and information technology field; it will help them in the long run for sustainable food security irrespective of climate conditions.

Data Availability

No such data is required.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] "Growing at a slower pace, world population is expected to reach 9.7 billion in 2050 and could peak at nearly 11 billion around 2100|UN DESA|United Nations Department of Economic and Social Affairs," April 2020, https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html.

[2] "68% of the world population projected to live in urban areas by 2050, says UN|UN DESA|United Nations Department of Economic and Social Affairs," April 2020, https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

[3] "Key facts-world urbanization prospects-population division-United Nations," April 2020, https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf.

[4] "Earth has lost a third of arable land in past 40 years, scientists say|Environment|The Guardian," https://www.theguardian.com/environment/2015/dec/02/arable-land-soil-food-security-shortage.

[5] "Ministry of Finance|Government of Pakistan|Pakistan Economic Survey 2018-19-Agriculture," May 2020, http://finance.gov.pk/survey_1819.html.

[6] "In Pakistan, how smart can agriculture be|Political Economy," May 2020, https://www.thenews.com.pk/tns/detail/576184-how-smart-can-agriculture-be.

[7] S. B. Baker, W. Xiang, and I. Atkinson, "Internet of things for smart healthcare: technologies, challenges, and opportunities," *IEEE Access*, vol. 5, pp. 26521–26544, 2017.

[8] E. Husni, G. B. Hertantyo, D. W. Wicaksono, F. C. Hasibuan, A. U. Rahayu, and M. A. Triawan, "Applied Internet of things (IoT): car monitoring system using IBM BlueMix," in *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 417–422, Lombok, Indonesia, 2017.

[9] M. S. Mekala, V. Perumal, M. Shareef Mekala, and P. Viswanathan, "A survey: smart agriculture IoT with cloud computing," in *2017 international conference on microelectronic devices, circuits and systems (ICMDCS)*, pp. 1–7, Vellore, India, 2017.

[10] H. Arasteh, V. Hosseinnezhad, V. Loia et al., "Iot-based smart cities: a survey," in *2016 IEEE 16th international conference on environment and electrical engineering (EEEIC)*, pp. 1–6, Italy, 2016.

[11] B. L. Risteska Stojkoska and K. V. Trivodaliev, "A review of Internet of things for smart home: challenges and solutions," *Journal of Cleaner Production*, vol. 140, pp. 1454–1464, 2017.

[12] Q. F. Hassan, *Internet of things A to Z: technologies and applications-concepts and perspectives*, Wiley-IEEE Press, 2018.

[13] M. P. Wachowiak, D. F. Walters, J. M. Kovacs, R. Wachowiak-Smolíková, and A. L. James, "Visual analytics and remote sensing imagery to support community-based research for precision agriculture in emerging areas," *Computers and Electronics in Agriculture*, vol. 143, pp. 149–164, 2017.

[14] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: a survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[15] S. A. Imam, A. Choudhary, and V. K. Sachan, "Design issues for wireless sensor networks and smart humidity sensors for precision agriculture: a review," *International Conference on Soft Computing Techniques and Implementations, ICSCTI*, vol. 2016, pp. 181–187, 2015.

[16] D. J. Cocovi-Solberg, M. Rosende, and M. Miró, "Automatic kinetic bioaccessibility assay of lead in soil environments using flow-through microdialysis as a front end to electrothermal atomic absorption spectrometry," *Environmental Science & Technology*, vol. 48, no. 11, pp. 6282–6290, 2014.

[17] A. Hemmat, A. R. Binandeh, J. Ghaisari, and A. Khorsandi, "Development and field testing of an integrated sensor for on-the-go measurement of soil mechanical resistance," *Sensors and Actuators A: Physical*, vol. 198, pp. 61–68, 2013.

[18] S. C. Murray, "Optical sensors advancing precision in agricultural production," *Photonics Spectra*, vol. 51, no. 6, pp. 48–56, 2017.

[19] M. R. Yousefi, "Application of GIS and GPS in precision agriculture (a review)," May 2020, http://www.ijabbr.com/.

[20] H. Sharma, M. K. Shukla, R. Steiner, M. K. Shukla, and P. W. Bosland, "Soil moisture sensor calibration, actual evapotranspiration, and crop coefficients for drip irrigated greenhouse chile peppers," *Agricultural Water Management*, vol. 179, pp. 81–91, 2017.

[21] B. H. Y. Alsalam, K. Morton, D. Campbell, and F. Gonzalez, "Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture," in *2017 IEEE Aerospace Conference*, pp. 1–12, Big Sky, MT, USA, 2017.

[22] W. H. Maes and K. Steppe, "Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture," *Trends in Plant Science*, vol. 24, no. 2, pp. 152–164, 2019.

[23] G. Sona, D. Passoni, L. Pinto et al., "UAV multispectral survey to map soil and crop for precision farming applications," https://air.unimi.it/handle/2434/451326.

[24] M. A. Hassan, M. Yang, A. Rasheed et al., "A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform," *Plant Science*, vol. 282, pp. 95–103, 2019.

[25] H. Huang, J. Deng, Y. Lan, A. Yang, X. Deng, and L. Zhang, "A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery," *PLoS One*, vol. 13, no. 4, article e0196302, 2018.

[26] F. Vanegas, D. Bratanov, K. Powell, J. Weiss, and F. Gonzalez, "A novel methodology for improving plant pest surveillance in vineyards and crops using UAV-based hyperspectral and spatial data," *Sensors*, vol. 18, no. 1, p. 260, 2018.

[27] K. Ivushkin, H. Bartholomeus, A. K. Bregt et al., "UAV based soil salinity assessment of cropland," *Geoderma*, vol. 338, pp. 502–512, 2019.

[28] S. Park, D. Ryu, S. Fuentes, H. Chung, E. Hernández-Montes, and M. O'Connell, "Adaptive estimation of crop water stress in nectarine and peach orchards using high-resolution imagery from an unmanned aerial vehicle (UAV)," *Remote Sensing*, vol. 9, no. 8, p. 828, 2017.

[29] S. Diwate, V. Nitnaware, and K. Argulwar, "Design and development of application specific drone machine for seed sowing," *International Research Journal of Engineering and Technology*, vol. 5, no. 5, pp. 4003–4007, 2018.

[30] F. Castaldi, F. Pelosi, S. Pascucci, and R. Casa, "Assessing the potential of images from unmanned aerial vehicles (UAV) to support herbicide patch spraying in maize," *Precision Agriculture*, vol. 18, no. 1, pp. 76–94, 2017.

[31] B. S. Faiçal, H. Freitas, P. H. Gomes et al., "An adaptive approach for UAV-based pesticide spraying in dynamic environments," *Computers and Electronics in Agriculture*, vol. 138, pp. 210–223, 2017.

[32] M. N. Muhammad, A. Wayayok, A. R. Mohamed Shariff, A. F. Abdullah, and E. M. Husin, "Droplet deposition density of organic liquid fertilizer at low altitude UAV aerial spraying in rice cultivation," *Computers and Electronics in Agriculture*, vol. 167, article 105045, 2019.

[33] J. Senthilnath, M. Kandukuri, A. Dokania, and K. N. Ramesh, "Application of UAV imaging platform for vegetation analysis based on spectral-spatial methods," *Computers and Electronics in Agriculture*, vol. 140, pp. 8–24, 2017.

[34] J. Navia, I. Mondragon, D. Patino, and J. Colorado, "Multispectral mapping in agriculture: terrain mosaic using an autonomous quadcopter UAV," in *2016 International Conference on Unmanned Aircraft Systems, ICUAS 2016*, pp. 1351–1358, Arlington, VA, USA, 2016.

[35] B. Y. Suprapto, M. A. Heryanto, H. Suprijono, J. Muliadi, and B. Kusumoputro, "Design and development of heavy-lift hexcopter for heavy payload," in *2017 International Seminar on Application for Technology of Information and Communication (iSemantic)*, pp. 242–246, Semarang Indonesia, 2017.

[36] B. Dai, Y. He, F. Gu, L. Yang, J. Han, and W. Xu, "A vision-based autonomous aerial spray system for precision agriculture," in *2017 IEEE International Conference on Robotics and Biomimetics, ROBIO*, pp. 1–7, Macau, Macao, 2017.

[37] "Fixed-wing drone, UAV for long-range commercial applications|Applied Aeronautics," April 2020, https://www.unmannedsystemstechnology.com/company/applied-aeronautics/.

[38] M. N. Boukoberine, Z. Zhou, and M. Benbouzid, "A critical review on unmanned aerial vehicles power supply and energy management: solutions, strategies, and prospects," *Applied Energy*, vol. 255, article 113823, 2019.

[39] G. Sierra, M. Orchard, K. Goebel, and C. Kulkarni, "Battery health management for small-size rotary-wing electric unmanned aerial vehicles: an efficient approach for constrained computing platforms," *Reliability Engineering and System Safety*, vol. 182, pp. 166–178, 2019.

[40] X. Feng, F. Yan, and X. Liu, "Study of wireless communication technologies on Internet of things for precision agriculture," *Wireless Personal Communications*, vol. 108, no. 3, pp. 1785–1802, 2019.

[41] S. Verma, A. Bhatia, A. Chug, and A. P. Singh, "Recent advancements in multimedia big data computing for IoT applications in precision agriculture: opportunities, issues, and challenges," in *Multimedia Big Data Computing for IoT Applications*, pp. 391–416, Springer, Singapore, 2020.

[42] "The Path to 5G with LTE-M and NB-IoT|Sierra Wireless," May 2020, https://www.sierrawireless.com/iot-blog/iot-blog/2018/05/lte-m-nb-iot-5g-networks/.

[43] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Z. Kovács, P. Mogensen, and M. Sorensen, "Coverage and capacity analysis of sigfox, lora, gprs, and nb-iot," in *2017 IEEE 85th vehicular technology conference (VTC Spring)*, pp. 1–5, Sydney, NSW, Australia, 2017.

[44] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the limits of LoRaWAN," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, 2017.

[45] A. Hoglund, X. Lin, O. Liberg et al., "Overview of 3GPP release 14 enhanced NB-IoT," *IEEE Network*, vol. 31, no. 6, pp. 16–22, 2017.

[46] F. Ghavimi and H.-H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: architectures, service requirements, challenges, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 525–549, 2014.

[47] S. R. Pokhrel and C. Williamson, "Modeling compound TCP over WiFi for IoT," *IEEE/ACM Transactions on Networking*, vol. 26, no. 2, pp. 864–878, 2018.

[48] N. Baker, "ZigBee and Bluetooth: strengths and weaknesses for industrial applications," *Computing & Control Engineering Journal*, vol. 16, no. 2, pp. 20–25, 2005.

[49] J. A. Manrique, J. S. Rueda-Rueda, and J. M. T. Portocarrero, "Contrasting Internet of things and wireless sensor network from a conceptual overview," in *2016 IEEE international conference on Internet of Things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData)*, pp. 252–257, Chengdu, China, 2017.

[50] H. S. Mahmood, M. Ahmad, T. Ahmad, M. A. Saeed, and M. Iqbal, "Potentials and prospects of precision agriculture in Pakistan-a review," *Pakistan Journal of Agricultural Research*, vol. 26, no. 2, pp. 151–167, 2013.

[51] "The Netherlands is a leader in sustainable agriculture|World Economic Forum," May 2020, https://www.weforum.org/agenda/2019/11/netherlands-dutch-farming-agriculture-sustainable.

[52] "Restorative farm produces food, fresh water, and clean energy in the desert - Science," May 2020, https://pk.mashable.com/science/1003/restorative-farm-produces-food-fresh-water-and-clean-energy-in-the-desert.

[53] "How to grow four tons of food a year in a metal box without sunlight|MIT Technology Review," May 2020, https://www.technologyreview.com/2018/03/08/3213/how-to-grow-four-tons-of-food-a-year-in-a-metal-box-without-sunlight/#Echobox=1582745134.

[54] "Paris is opening the world's largest urban rooftop farm|World Economic Forum," May 2020, https://www.weforum.org/agenda/2019/08/vertical-urban-farm-city-paris.

[55] "Inside London's first underground farm|The Independent," May 2020, https://www.independent.co.uk/Business/indyventure/growing-underground-london-farm-food-waste-first-food-miles-a7562151.html.

[56] J. Torres-Sánchez, J. M. Peña, A. I. de Castro, and F. López-Granados, "Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV," *Computers and Electronics in Agriculture*, vol. 103, pp. 104–113, 2014.

[57] O. A. Denton, V. O. Aduramigba-Modupe, A. O. Ojo et al., "Assessment of spatial variability and mapping of soil properties for sustainable agricultural production using geographic information system techniques (GIS)," *Cogent Food & Agriculture*, vol. 3, no. 1, 2017.

[58] "Making every drop count: Pakistan's growing water scarcity challenge|IISD," May 2020, https://www.iisd.org/library/making-every-drop-count-pakistan-s-growing-water-scarcity-challenge.

[59] B. Keswani, A. G. Mohapatra, A. Mohanty et al., "Adapting weather conditions based IoT enabled smart irrigation technique in precision agriculture mechanisms," *Neural Computing and Applications*, vol. 31, no. S1, pp. 277–292, 2019.

[60] "Tasmanian farmers use drones to make irrigation more efficient - ABC News," May 2020, https://www.abc.net.au/news/2017-07-27/drone-technology-on-farms/8746272.

[61] "Yield losses due to pests," May 2020, https://blog.agrivi.com/post/yield-losses-due-to-pests.

[62] H. Obasekore, M. Fanni, and S. M. Ahmed, "Insect killing robot for agricultural purposes," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, vol. 2019, pp. 1068–1074, 2019.

[63] A. Khattab, S. E. D. Habib, H. Ismail, S. Zayan, Y. Fahmy, and M. M. Khairy, "An IoT-based cognitive monitoring system for early plant disease forecast," *Computers and Electronics in Agriculture*, vol. 166, article 105028, 2019.

[64] D. Sinwar, V. S. Dhaka, M. K. Sharma, and G. Rani, *AI-based yield prediction and smart irrigation*, Springer, Singapore, 2020.

[65] S. Athani, C. H. Tejeshwar, M. M. Patil, P. Patil, and R. Kulkarni, "Soil moisture monitoring using IoT enabled Arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka-India," *Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC*, vol. 2017, pp. 43–48, 2017.

[66] E. Suganya, S. Sountharrajan, S. K. Shandilya, and M. Karthiga, "IoT in agriculture investigation on plant diseases and nutrient level using image analysis techniques," in *Internet of Things in Biomedical Engineering*, pp. 117–130, Academic Press, 2019.

[67] M. Lu, M. Bagheri, A. P. James, and T. Phung, "Wireless charging techniques for UAVs: a review, reconceptualization, and extension," *IEEE Access*, vol. 6, pp. 29865–29884, 2018.

[68] S. Aldhaher, P. D. Mitcheson, J. M. Arteaga, G. Kkelis, and D. C. Yates, "Light-weight wireless power transfer for mid-air charging of drones," in *2017 11th European Conference on Antennas and Propagation (EUCAP)*, pp. 336–340, Paris, France, 2017.

[69] T. M. Mostafa, A. Muharam, and R. Hattori, "Wireless battery charging system for drones via capacitive power transfer," in *2017 IEEE PELS Workshop on Emerging Technologies: Wireless Power Transfer (WoW)*, pp. 1–6, Chongqing, China, 2017.

[70] A. M. Jawad, H. M. Jawad, R. Nordin, S. K. Gharghan, N. F. Abdullah, and M. J. Abu-Alshaeer, "Wireless power transfer with magnetic resonator coupling and sleep/active strategy for a drone charging station in smart agriculture," *Access*, vol. 7, pp. 139839–139851, 2019.

[71] H. M. Jawad, A. M. Jawad, R. Nordin et al., "Accurate empirical path-loss model based on particle swarm optimization for wireless sensor networks in smart agriculture," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 552–561, 2020.

[72] D. W. Sambo, A. Forster, B. O. Yenke, I. Sarr, B. Gueye, and P. Dayang, "Wireless underground sensor networks path loss model for precision agriculture (WUSN-PLM)," *IEEE Sensors Journal*, vol. 20, no. 10, pp. 5298–5313, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 9572992, 14 pages
https://doi.org/10.1155/2022/9572992

Research Article

Trusted Blockchain-Based Signcryption Protocol and Data Management for Authentication and Authorization in VANETs

Jinqi Su,[1] Runtao Ren ,[2] Yinghao Li,[2] Raymond Y. K. Lau,[3] and Yikuan Shi[4]

[1] *School of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710061, China*
[2] *School of Cyberspace Security, Xi'an University of Posts and Telecommunications, Xi'an 710121, China*
[3] *Department of Information Systems, City University of Hong Kong, Hong Kong*
[4] *AVIC Jonhon Optronic Technology Co., Ltd., Luoyang 471000, China*

Correspondence should be addressed to Runtao Ren; 760473028@qq.com

Received 6 April 2022; Accepted 30 April 2022; Published 21 May 2022

Academic Editor: Maode Ma

Vehicular Ad hoc Networks (VANETs) are the industrial cornerstone of intelligent transportation system (ITS), which are widely used in traffic management, automatic driving, and road optimization. With the expansion of the scale of the mobile ad hoc networks (MANETs) and smart vehicles (SV), VANETs will produce a large amount of data. In the open access environment of VANETs, the security of information transmission and the authenticity of user identity need to be considered when different vehicles communicate. In order to solve the cybersecurity risks of large-scale deployment of VANET, this paper proposes a trusted blockchain-based signcryption protocol and data management (TB-SCDM) for authentication and authorization (A&A) in VANETs. In the existing attack model, TB-SCDM can ensure the confidentiality and undeniability of information, as well as can effectively resist 51% attacks, eclipse attacks and double-spending attacks, etc. Through benchmark analysis, this scheme has higher computing efficiency and lower storage cost compared with other existing schemes.

1. Introduction

The VANETs have stimulated interest in both academic and industry, thanks to their intelligence and networking that assist vehicle driving and promote the application and development of ITS (e.g., automatic driving) [1–3]. At the same time, the VANETs have also become one of the most promising and fastest-growing subsets of the MANETs [4]. The VANETs are distributed and self-organized networks which communicate through wireless media, built up by SV, roadside units (RSUs), global positioning system (GPS), trusted authority (TA), and on-board units (OBUs). SV could communicate with each other as well as with roadside units (RSU) (e.g., electric toll collection of highways), which provide a good dedicated short-range communication (DSRC) by IEEE 802.11p standard for automatic driving technology to identify real-time traffic conditions [5–7]. TA is a third-party certification center used by the RSU and OBU that is responsible for controlling the whole network. RSU is a base station (e.g., Wi-Fi or WiMAX) that keeps as a central hub between the TA and the OBU and performs different authentications. The OBU is introduced on the vehicle to acquire procedure and exchange data identified with different vehicles and RSUs through DSRC.

With the main goal of improving road safety and driving conditions, VANETs are established with five types of communications: the vehicle-to-vehicle (V2V), vehicle-to-roads (V2R), vehicle-to-infrastructure (V2I), roads-to-roads (R2R), and the roads-to-infrastructure (R2I) [8]. The architecture of VANETs is appeared in Figure 1. Due to the open nature of VANETs and lacking infrastructure, these delays establish reliable end-to-end communication paths and have efficient data transfer [9–10]. In particular, automatic driving technology has many system problems and security difficulties in obtaining availability, securing communication, and accessibility of exchange. In VANETs, SV are strangers who do not trust each other [11]. Without authentication and authorization, the attacker may impersonate any vehicle to broadcast forged messages to easily track the target vehicle by analyzing

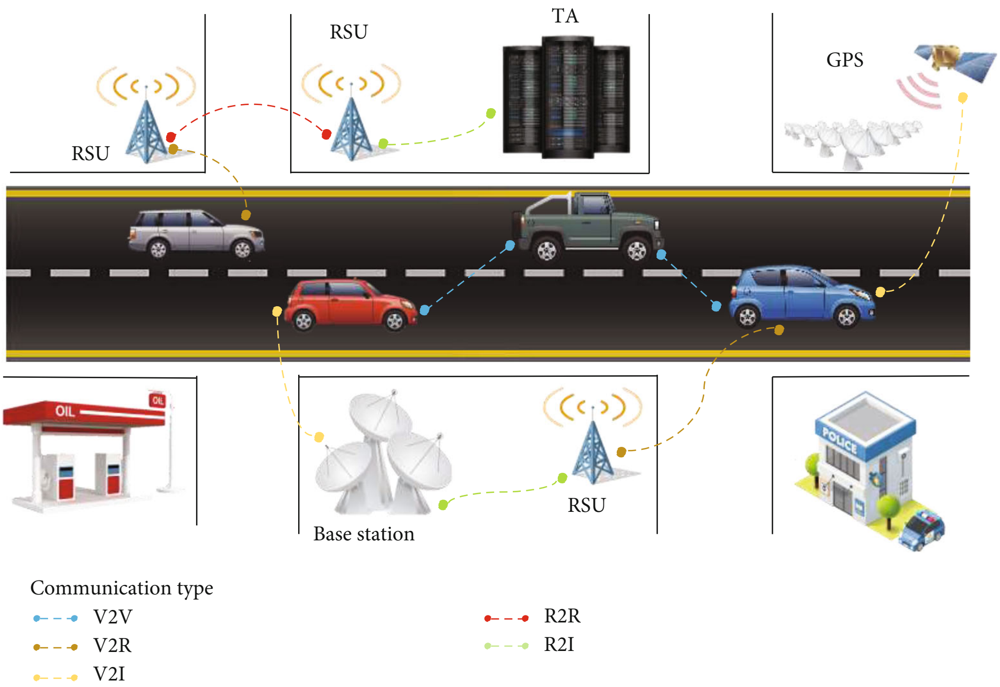


Figure 1: The architecture of VANETs.

the broadcast messages, which will pose a serious threat [12]. Therefore, when the users of SV use automatic driving, they need to authenticate and authorize the identity of vehicle in VANETs.

In the conventional A&A schemes, public key infrastructure- (PKI-) based solutions need a certificate authority (CA), while identification-based solutions require a key generation center (KGC) to provide vehicles with secure authentication [13–15]. However, there is a high computational overhead and large storage capacity on the CA and TA in the case of large number of certificates.

Considering the above limitations, blockchain has the function of distributed storage, which can effectively realize decentralization [16–17]. The methods of automatically injecting trust, checking reliability, monitoring interentity communication, and analyzing behavior can be implemented in the blockchain. It forms a distributed database by using digital signature, encryption technology, hash function, and timestamp [18]. Blockchain assigns the responsibility of maintaining privacy and security to all entities in VANET instead of centralized operation [19–22]. In addition, identity-based signcryption protocol has shorter ciphertext and less computational overhead, which can sign and encrypt the data to ensure the confidentiality and nonrepudiation of the information [23–25].

Our contribution: in automatic driving, since VANETs consist of a large number of SV at high speed, the security of information transmission must be satisfied A&A efficiently. In this paper, we propose a scheme that combined blockchain and signcryption to realize the A&A when the SV using automatic driving interacts with other media. Figure 2 shows the physical process of TB-SCDM when the SV use automatic driving. The contributions of this article are as follows.

(1) This article is an SV management system built on the consortium chain, which can upload the relevant data of SV to the blockchain to realize distributed storage

(2) The A&A function of SV users in ITS can be effectively realized in TB-SCDM scheme. The A&A mechanism we designed can ensure the trusted identity and effective authorization of SV users in VANETs

(3) The TB-SCDM scheme combines blockchain and signcryption. The data on the consortium chain cannot be tampered with arbitrarily. This mechanism provides a stronger security level for signing and encrypting the data that needs to be verified. Therefore, the confidentiality and unforgeability of SV information transmission in VANETs can be realized through TB-SCDM

2. Preliminaries

2.1. Consortium Chains. Consortium chain has the advantages of weak concentration, high controllability, and great scalability [26]. Thanks to the number of nodes and organizational structure being relatively limited, consortium chain is mainly applied in systems built by specific organizations (e.g., data interaction of ITS). The rights of each participating node in the consortium chain are completely equal, and they can realize the trusted exchange of data. Each node of the consortium chain has a corresponding entity that wants to join and exit only to be executed after authorization. In the consortium chain, data transactions do not need the consensus of the whole blockchain network. Therefore, the consortium chain satisfies the data management requirements of VANETs through controlled access, efficient storage and trusted storage.

2.2. Smart Contract. Smart contract refers to a computer program that can be executed by a network of mutually untrusted nodes without any trusted authority. Compared with traditional programming source code, smart contracts utilize blockchain immutable distributed storage. In the initial stage of building the data storage system of SV, the vehicle management system can write triggers to realize the functions according to the actual needs. Once the system is put into operation, when the trigger conditions are met, the content of the smart contract can be executed to complete data upload, network access, and other processing functions. Finally, smart contracts can be developed to achieve smaller permission control granularity.

2.3. Practical Byzantine Fault Tolerance. The practical byzantine fault tolerance (PBFT) means a kind of fault tolerance of distributed network (i.e., the network can still make honest nodes reach a consensus. The PBFT mechanism will specify that one node in the system is the master node, and the other nodes are secondary nodes [27]. The process of PBFT is shown in Figure 3. When the primary node fails, all legal nodes in the system are eligible to upgrade from the secondary node to the primary node and follow the principle of the minority obeying the majority to ensure that honest nodes can reach a consensus. However, in order for the PBFT to operate normally, the number of malicious nodes must be less than 1/3 of the total number of nodes in the network. For example, in order to ensure the normal operation of the whole system, assuming that the number of invalid or malicious nodes tolerated by PBFT is F and the total number of nodes of the system is $|R| = 3F + 1$, then $2F + 1$ normal nodes are required. Hence, the PBFT algorithm can tolerate less than 1/3 invalid or malicious nodes.

2.4. Meaning of Symbols. The specific meaning of the symbols is contained in Table 1.

3. Formation Definition

3.1. Syntax. The algorithm definition of TB-SCDM is as follows.

Initialize $(1^{\theta}) \rightarrow$ *Table*: the initialize algorithm is executed by an administrator in the securable environment. Firstly, the administrator has a query for system and takes as input a security parameter θ, then return a local table named management and output 0 otherwise.

BlockUp (Table) $\rightarrow$ *1 or 0*: the BlockUp algorithm is run by the administrator as well. For this algorithm, administrator sends each primary key $(N_i, I_i, ID_i, IK_i, IR_i, PK_i, C_i, \mu_i, \sigma_i, \sigma_{IR_i})$ to table for achieving consensus among nodes then output 1 or 0.

Signcrypt $(N_i, ID_i, IK_i) \rightarrow PK_i, SK_i$: the Keygen algorithm is performed by one user who tries to register a new account in the system. The user sends N_{i}, $\mathrm{ID_i}$, and $\mathrm{IK_i}$ to the system to generate $\mathrm{PK_i}$, $\mathrm{SK_i}$, and σ_{i}. Then, $\mathrm{PK_i}$, $\mathrm{SK_i}$,C_{i}, δ_{i}, μ_{i}, and σ_{i} will be saved in the table for connecting blockchain.

Authentication $(N_i, IK_i, SK_i) \rightarrow$ *1 or 0*: this user sends N_{i}^*, $\mathrm{IK_i}^*$, and $\mathrm{SK_i}^*$ to the system to produce digest δ_{i}^* and δ_{i}^{**} for validation. There are two cases in this process.

Case 1. If $\delta_{\mathrm{i}}^* = \delta_{\mathrm{i}}^{**}$, the user can realize the login process to show that the user's identity information is reliable.

Case 2. If $\delta_{\mathrm{i}}^* \neq \delta_{\mathrm{i}}^{**}$, the authentication of this user with identity is failed and output 0.

Update $(IR_i, SK_i) \rightarrow \boldsymbol{\sigma_{IR_i}}$: this algorithm is executed by the user who needs to update the resource in the system. Assume the identity of user is valid, the $\mathrm{IR_i}$ and $\mathrm{SK_i}$ can get input by this user to output signcryptedUserResourceσ_{IR_i}on the block.

Authorization $(N_i, PK_i) \rightarrow IR_i$ *or 0*: this Authorization algorithm is to realize the authorization of users. Initially, the user should send the target account N_{i}^* and the corresponding public key $\mathrm{PK_i}^*$ to platform for verification. There are two cases in this algorithm.

Case 1. If $\delta_{\mathrm{i}}^* = \delta_{\mathrm{i}}^{**}$, the user can be authorized and gain the part access for userResource $\mathrm{IR_i}$.

Case 2. If $\delta_{\mathrm{i}}^* \neq \delta_{\mathrm{i}}^{**}$, this user failed to authorization and output 0.

Conversation $(N_i, PK_i) \rightarrow$ *1 or 0*: the algorithm is used to establish dialogue between different users. First, the user can send N_{i} and $\mathrm{PK_i}^*$ to platform for communication. There are two situations in this algorithm.

Case 1. If $\delta_{\mathrm{i}}^* = \delta_{\mathrm{i}}^{**}$, the user can be authorized and gain a conversation.

Case 2. If $\delta_{\mathrm{i}}^* \neq \delta_{\mathrm{i}}^{**}$, instant messaging channel cannot establish and output 0.

Transaction (Chain) $\rightarrow$ *transactionHash*: this algorithm is run by administrator in order to obtain the information on the blockchain. The administrator can query the main parameters of the blockchain to get buildTime, buildType, genesisBlockHash and contractAddress, etc.

4. Concrete Scheme

There are eight parts in the TB-SCDM: Initialize, BlockUp, Signcrypt, Authentication, Update, Authorization, Conversation, and Transaction. The steps of Authentication, Update, and Authorization are described in Figure 4.

4.1. Initialize. This algorithm is to register a table named management on the blockchain so that later users' information can be registered on the consortium chain.

4.2. BlockUp. This algorithm is executed by the administrator. Its purpose is to create each primary key in the table

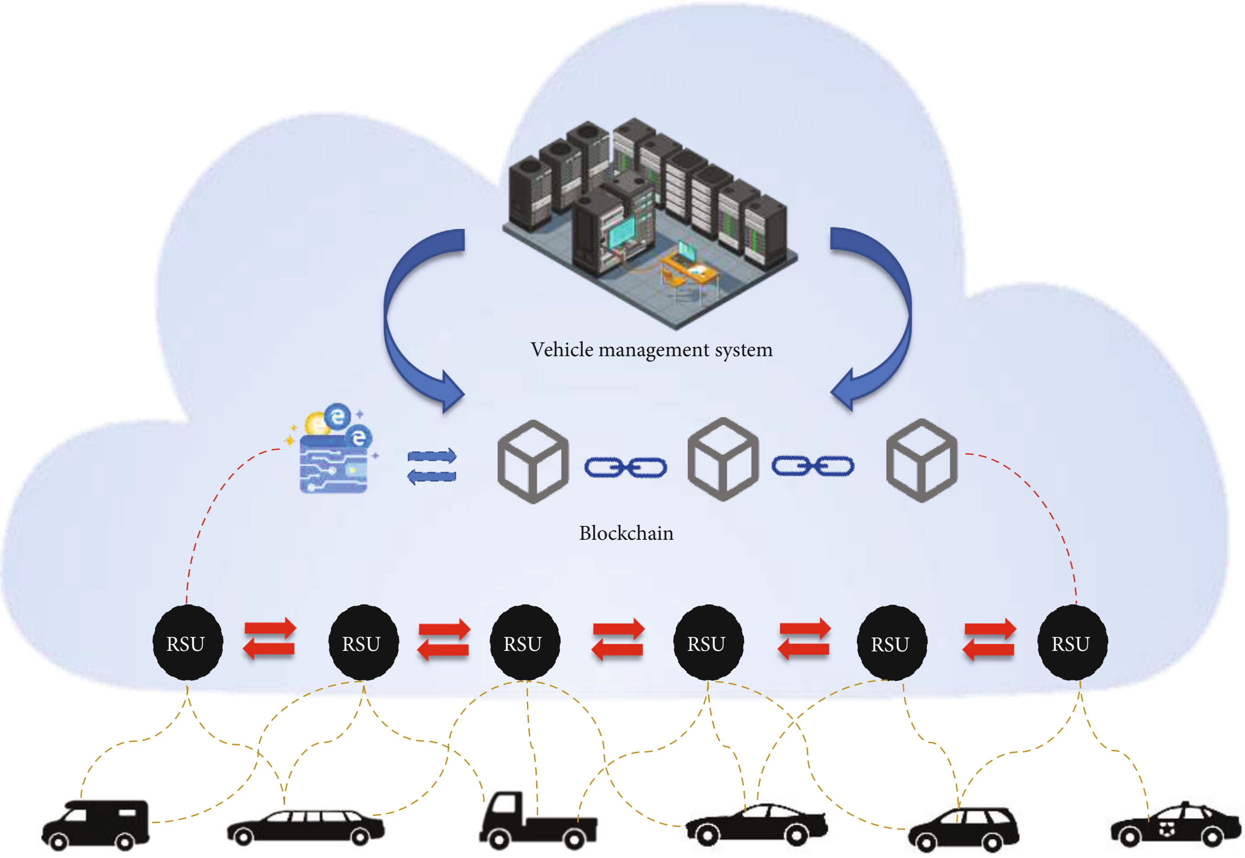


Figure 2: The process of TB-SCDM.

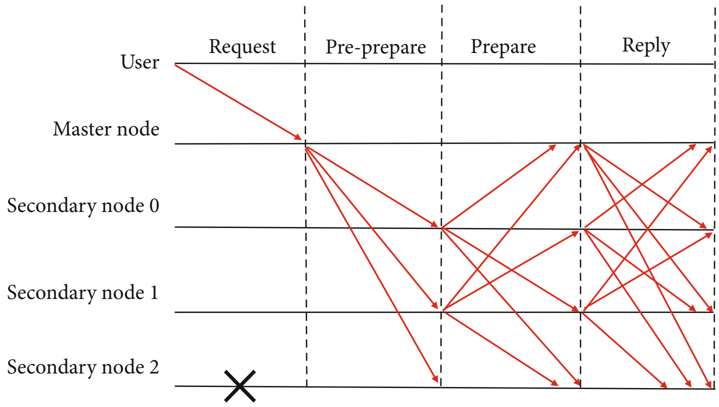


Figure 3: The process of PBFT.

generated in algorithm 1 and then upload the data of each primary key to the blockchain.

4.3. Signcrypt. Firstly, the system will first give a public-private key pair to the user with identity I_i. Accordingly, the user will deposit the I_i, ID_i, and IK_i in the plainText M_i to generate the hash value δ_i. Then, this user utilizes the private key SK_i to produce the signature μ_i and utilizes the public key PK_i to encrypt M_i for getting the ciphertext C_i. Finally, μ_i and C_i will be merged to return the signcryption σ_i.

4.4. Authentication. For Authentication algorithm, the user of identity I_i can input $N_i{}^*$, $\mathrm{IK}_i{}^*$, and $\mathrm{SK}_i{}^*$ in the system and then query whether there exists the account named $N_i{}^*$ in the table. If exist N_i $*$, this user will enter the authentication stage.

Table 1: Specific meaning of symbols.

Notations	Meaning
N_{i}	The accountName of user
I_{i}	The identity of arbitrary user
IK_{i}	The userKey of I_{i}
ID_{i}	The userData of I_{i}
IR_{i}	The userResource of I_{i}
PK_{i}	The publicKey of I_{i}
SK_{i}	The secretKey of I_{i}
δ_{i}	The digest of I_{i}
M_{i}	The plainText of I_{i}
C_{i}	The cipherText of I_{i}
μ_{i}	The signature of I_{i}
$?_{\mathrm{i}}$	The signcryption of I_{i}
σ_{IR_i}	The signcryptedUserResource of I_{i}

On the client side, the (I_{i}, ID_{i}, and $\mathrm{IK}_{\mathrm{i}}^*$) will deposit in plainText M_{i} to produce the hash value δ_{i}^*. Accordingly, the signature μ_{i}^* can be generated by the private key $\mathrm{SK}_{\mathrm{i}}^*$.

On the blockchain side, the signature μ_{i}^* can be unsigned by the trusted public key PK_{i} stored by previous user of identity I_{i}. Accordingly, the trusted signcryption σ_{i} can be unsigncrypted to get the hash value δ_{i}^{**}. After obtaining the above data, the next step will verify the user's identity. There are two cases in this process.

Case 1. If $\delta_{\mathrm{i}}^* = \delta_{\mathrm{i}}^{**}$, the user can realize the login process to show that the user's identity information is reliable.

Case 2. If $\delta_{\mathrm{i}}^* \neq \delta_{\mathrm{i}}^{**}$, the authentication of this user with identity is failed and output 0.

4.5. Update. The user of identity I_{i} can update the resource in the system through this algorithm. The user can input IR_{i} and SK_{i} in the system. Then, I_{i}, PK_{i}, and IR_{i} will be merged into the plainText M_{i}. The following queries are same as those in Algorithm 1.

Finally, the updated information of these users will be uploaded to consortium chain.

4.6. Authorization. This algorithm is designed to authorize the legitimacy of user's behavior. In the authorization process, we add the token technology. In this mechanism, we first set the upper limit of the user's single query time to 300 s.

After exceeding the time, the user's access rights will disconnected, and his identity needs to be verified newly. Within legal time, account N_{i}^* will be first verified for existence. If account N_{i}^* exists, then the user of identity I_{i} will enter the authentication stage for authorization.

On the blockchain side, the trusted signcryption σ_{i} can be unsigncrypted to return the signature μ_{i}^* and ciphertext $\mathrm{C}_{\mathrm{i}}^{**}$. Accordingly, the signature μ_{i}^* can be unsigned to get

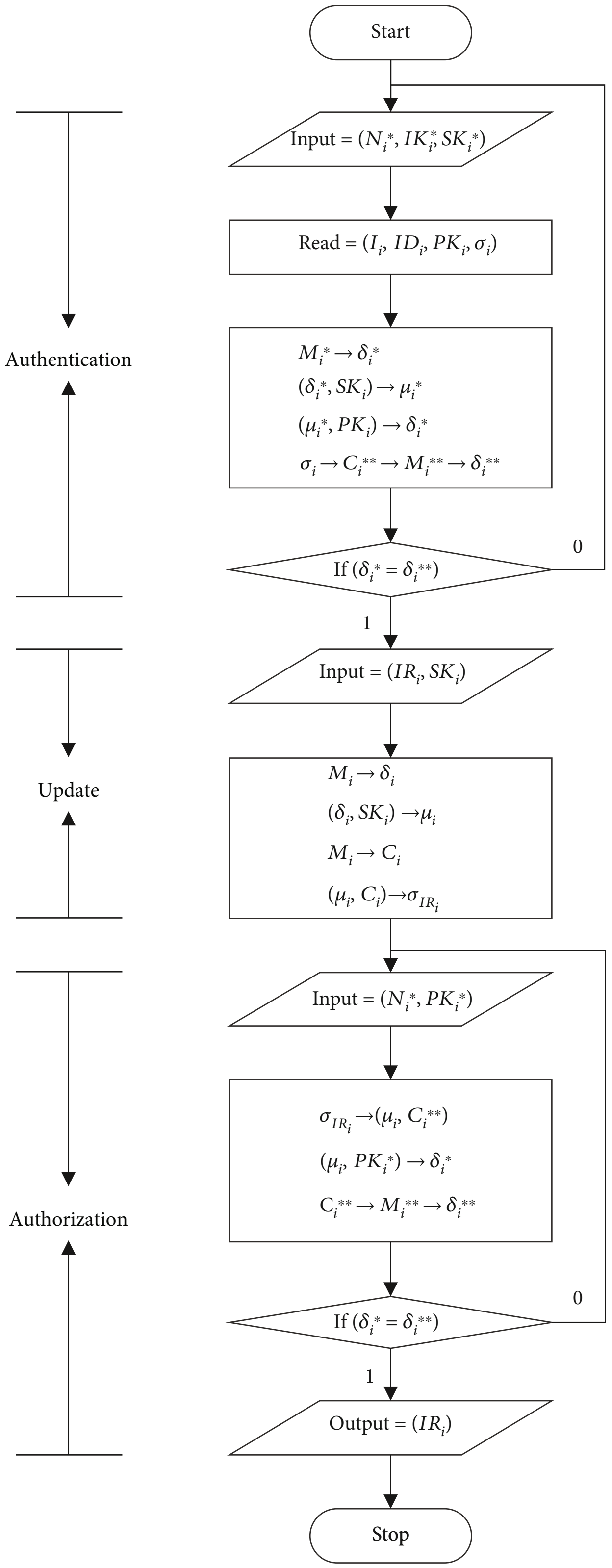

Start
Authentication
Input = (N_i^*, IK_i^*; SK_i^*)
Read = (I_i, ID_i, PK_i, σ_i)
$M_i^* \rightarrow \delta_i^*$
$(\delta_i^*, SK_i) \rightarrow \mu_i^*$
$(\mu_i^*, PK_i) \rightarrow \delta_i^*$
$\sigma_i \rightarrow C_i^{**} \rightarrow M_i^{**} \rightarrow \delta_i^{**}$
If ($\delta_i^* = \delta_i^{**}$)
0
1
Update
Input = (IR_i, SK_i)
$M_i \rightarrow \delta_i$
$(\delta_i, SK_i) \rightarrow \mu_i$
$M_i \rightarrow C_i$
$(\mu_i, C_i) \rightarrow \sigma_{IR_i}$
Authorization
Input = (N_i^*, PK_i^*)
$\sigma_{IR_i} \rightarrow (\mu_i, C_i^{**})$
$(\mu_i, PK_i^*) \rightarrow \delta_i^*$
$C_i^{**} \rightarrow M_i^{**} \rightarrow \delta_i^{**}$
If ($\delta_i^* = \delta_i^{**}$)
0
1
Output = (IR_i)
Stop

Figure 4: The process of Authentication, Update, and Authorization.

```
Input:$1^{\theta}$
Output: Table
   createTable() private
{
   tf.createTable(“management”, “$I_i$”, “$N_i$”, “$ID_i$”, “$IK_i$”, “$PK_i$”, “$\mu_i$”, “$C_i$”, “$\sigma_i$”, “$IR_i$”, “$\sigma_{IR_i}$”);
}
openTable() private returns(table)
{
   TableFactory tf = TableFactory(0x1001) ;
   Table table = tf.openTable(“management”);
   return table;
}
```

Algorithm 1: Initialize.

```
Input:$I_i$, $N_i$, $ID_i$, $IK_i$, $PK_i$, $\mu_i$, $C_i$, $\sigma_i$, $IR_i$, $\sigma_{IR_i}$
Output: true or false
   statu = select(management)
   if(statu !==0) {
      Table table = openTable();
      Entry entry = table.newEntry();
      entry.set("$I_i$, $N_i$, $ID_i$, $IK_i$, $PK_i$, $\mu_i$, $C_i$, $\sigma_i$, $IR_i$”", $I_i$, $N_i$, $ID_i$, $IK_i$, $PK_i$, $\mu_i$, $C_i$, $\sigma_i$, $IR_i$);
      return True;
   } else {
      return false;
   }
```

Algorithm 2: BlockUp.

```
Input:$N_i$, $ID_i$, $IK_i$
Output: $PK_i$, $SK_i$
Function SignCrypted Input($M_i$, $SK_i$) Output($\delta_i$, $\mu_i$, $C_i$, $\sigma_i$){
      $\delta_i$ = Method.hash($M_i$);
      $\mu_i$ = Method.sign($\delta_i$, $SK_i$);
      $C_i$ = homomorphicEncryption.Enc($M_i$);
      $\sigma_i$ = signcrypt($\mu_i$ || $C_i$);
   }
   statu = select($N_i$);
   if(statu !==0) {
      Get cryptographic KeyPair = new createKeyPair();
      Get cryptographic Method = new cryptographic (CryptoType.SCHNORRTYPE);
      $PK_i$ = KeyPair.get$PK_i$ ();
      $SK_i$ = KeyPair.get$SK_i$ ();
      $M_i$ = $I_i$ || $ID_i$ || $IK_i$;
      $\delta_i$ || $\mu_i$ || $C_i$ || $\sigma_i$ = function.SignCrypted($M_i$, $SK_i$);
      entry.set("$N_i$, $ID_i$, $IK_i$, $PK_i$, $\mu_i$, $C_i$, $\sigma_i$”", $N_i$, $ID_i$, $IK_i$, $PK_i$, $\mu_i$, $C_i$, $\sigma_i$);
      count = table.insert($N_i$, entry);
      if (count ==1) {
         statu_code = true;
      } else {
         statu_code = false;
      }
   } else {
      statu_code = false;}
   return statu_code $PK_i$$SK_i$;
```

Algorithm 3: Signcrypt.

```
**Input:** $N_i$, $IK_i$, $SK_i$
**Output:** true or false
Function unSignCrypted Input($\sigma_i$, $PK_i$) Output(true or false)
    {
        '$\mu_i$`' || '$C_i$**' = unsigncrypt($\sigma_i$);
        if('$\mu_i$*' == 0) {
            '$\mu_i$*' = '$\mu_i$`';
        }
        '$\delta_i$*' = Method.unsign('$\mu_i$*', $PK_i$);
        '$M_i$**' = homomorphicEncryption.Dec('$C_i$**');
        '$\delta_i$**' = Method.hash('$M_i$**');
        if('$\delta_i$*' == '$\delta_i$**') {
            statu_code = true;
        } else {
            statu_code = false;
        }
    }
    statu = select($N_i$);
    if(statu !==0) {
        Get cryptographic Method = new cryptographic (CryptoType.SCHNORRTYPE);
        '$IK_i$*' = result.getValue2();
        '$SK_i$*' = result.getValue3();
        '$M_i$*' = $I_i$ || $ID_i$ || '$IK_i$*';
        '$\delta_i$*' = Method.hash('$M_i$*');
        '$\mu_i$*' = Method.sign('$\delta_i$*', '$SK_i$*');
        statu_code = function.unSignCrypted($\sigma_i$, $PK_i$);
    } else {
            statu_code = false;
        }
    return statu_code;
```

Algorithm 4: Authentication

the hash value δ_i^*. And the ciphertext C_i^{**} can be decrypted to acquire δ_i^{**}. After obtaining the above data, the next step will enter to the validation. There are two cases in this process.

Case 1. If $\delta_i^* = \delta_i^{**}$, the user can be authorized and gain the part access for userResource IR_i.

Case 2. If $\delta_i^* \neq \delta_i^{**}$, this user failed to authorization and output 0.

4.7. Conversation. Before two users establish a session, the system will set the maximum time limit for a single query to 300 s. After exceeding the time, it will be disconnected automatically and need to be verified again. During the verification process, account N_i^* will be queried whether exist. The following queries are same as those in Algorithm 5.

4.8. Transaction. The administrator can query the main parameters of the blockchain to get buildTime, buildType, genesisBlockHash and contractAddress, etc. These data are unique and cannot be tampered with arbitrarily.

```
**Input:** $IR_i$, $SK_i$
**Output:** true or false
    Get cryptographic Method = new cryptographic
(CryptoType.SCHNORRTYPE);
    $M_i$ = $I_i$ || $PK_i$ || $IR_i$;
    $\delta_i$ || $\mu_i$ || $C_i$ || $\sigma_{IR_i}$ = function.SignCrypted($M_i$, $SK_i$);
    enter.set("$IR_i$, $\sigma_{IR_i}$”", $IR_i$, $\sigma_{IR_i}$);
    count = table.insert($IR_i$, $\sigma_{IR_i}$, entry);
    if (count ==1) {
        return true;
    } else {
        return false;
    }
```

Algorithm 5: Update.

5. Theoretical Analysis

5.1. Security Proof of Blockchain

5.1.1. Eclipse Attack. The multinode consortium blockchain system of TB-SCDM is built based on the FISCO BCOS platform. The system has a node access mechanism, so it is

```
Input:$N_i$, $PK_i$
Output:$IR_i$
    timeStamp = System.TimeSeconds();
    expireTime = System.TimeSeconds() - timeStamp;
    If(expireTime <300) {
        Statu = select($N_i$);
        If(statu! =0) {
            ‘$PK_i$*’ = result.getValue2();
            statu_code = function.unSignCrypted($\sigma_{IR_i}$, ‘$PK_i$*’);
            If(statu_code ==1) {
                Return $IR_i$;
            }
        }
    }
```

Algorithm 6: Authorization.

```
Input:$N_i$, $PK_i$
Output: true or false
timeStamp = System.TimeSeconds();
expireTime = System.TimeSeconds() - timeStamp;
if(expireTime < 300) {
    statu = select($N_i$);
    if(statu !==0) {
        '$PK_i$*' = result.getValue2();
        statu_code = function.unSignCrypted($\sigma_i$, '$PK_i$*');
        if(statu_code ==1) {
            creat.Conversation($N_i$);
            return true;
            }
        }
}
```

Algorithm 7: Conversation

difficult for attackers to obtain legitimate nodes through normal channels. Therefore, it is difficult for attackers to obtain legal nodes through normal channels. The PBFT mechanism of the TB-SCDM determines that if one third of the nodes of the system operate normally, it will not affect the normal operation of the whole system. Even if the attacker obtains the permissions of multiple accounting nodes, then the attacked node will be quickly discovered and processed by the central node.

5.1.2. DOS/DDoS Prevention. TB-SCDM adopts the consensus algorithm mechanism of consortium blockchain and PBFT. Therefore, the attack on ordinary nodes without accounting permission cannot hinder the normal operation of the blockchain system. Due to the characteristics of PBFT consistency algorithm mechanism, as long as there are more than one-third of normal nodes in the system, the system can operate normally, which leads to a huge inverse ratio between the attack cost and benefit of DDoS/DOS. However, for the consortium blockchain of TB-SCDM, the time and cost of discovering and repairing accounting nodes are very small.

5.1.3. 51% Attack Prevention. For the consortium blockchain, the greater the computing power of all nodes, the more difficult to implement 51% attacks. It is hard for attackers to break more than 51% of nodes in a short time, and it is difficult to complete the destruction of the ledger before the central node takes corresponding countermeasures. Even if the ledger is attacked, the central node can repair the ledger in a very short time.

5.1.4. Sybil Attack Prevention. Each registered user will generate a unique public-private key pair. Each node needs a unique and unforgeable public key when uploading or updating the data on blockchain. Therefore, any attacker cannot use a single forged public key to disguise as multiple users and occupy all links of a billing node.

5.2. Security Proof of Signcryption

5.2.1. Identity Authentication. The system binds the user's public key with the user ID and then provides it to the user for safekeeping in the user registration stage. In addition, the signcryption method bound with the user's public key is adopted in the process of chaining or reading all information, which ensures the traceability of the system to the data and the authentication of the identity.

5.2.2. Confidentiality. Compared with the traditional digital signature, this paper adopts the signcryption technology based on Schnorr. Many literatures have verified the IND-CCA security (i.e., indistinguishability under the adaptive chosen-ciphertext attacks) based on Schnorr under the random oracles or standard oracles. Through the analysis of provable security theory, signcryption technology can effectively ensure the confidentiality of information in the process of transmission.

5.2.3. Unforgeability. TB-SCDM verifies whether the transmitted message comes from the real sender by verifying the message digest of the sender and receiver. We generate the compared message digest by storing the public key and trusted data in the blockchain. If the message digest is the same as the sender's message digest, verification can be realized to achieve UF-CMA security (i.e., existentially unforgeable under the adaptive chosen-message attacks). This article innovatively integrates signcryption, timestamp, and blockchain based on Schnorr to ensure the unforgeability of information.

6. Benchmark Test

6.1. Benchmark Test of Blockchain. In order to efficiently perform operations, we accessed the data on TB-SCDM using the CRUD interface supplied by FISCO BCOS 2.0. The hardware environment is an Intel i5-8265U 1.80 GHz computer, 16 GB of memory, and running Windows 10 operating system.

It is available to deploy several different nodes on the same server for a test chain, we used a Linux server to deploy

```
Input: getchainVersion
Output: buildTime, buildType, genesisHash, etc.
  [group:1]> getNodeVersion
  ClientVersion{
    version='2.8.0',
    supportedVersion='2.8.0',
    chainId='1',
    buildTime='20210830 12:52:15',
    buildType='Linux/clang/Release',
    gitBranch='HEAD',
    genesisHash='bf0e0242a8040ead7549de49423712233a36d1b51b056a1c20df5eb78a9613e5'
  }
  transaction hash: 0xe88c2b9bf6dec9fa10356fd75b3d5414a5bd48f7ca246a8134e7f877928c47fc
  contract address: 0x48102a5d29a6109384cb5a9c97d9fd07dd1a4416
  currentAccount: 0xb13d80305a847dd2160c71465b50a6a1c0506ee3
  [group:1]> getBlockNumber
  9
  [group:1]> getCurrentAccount
  0xb13d80305a847dd2160c71465b50a6a1c0506ee3
```

Algorithm 8: Transaction

Table 2: The performance metrics of send rate, latency, and throughput.

Name	Succ	Fail	Send rate (TPS)	Max latency(s)	Min latency(s)	Avg latency(s)	Throughput (TPS)
User	1000	0	606.2	2.16	0.32	1.52	371.6
Transfer	10000	0	976.6	18.35	1.33	12.25	509.7

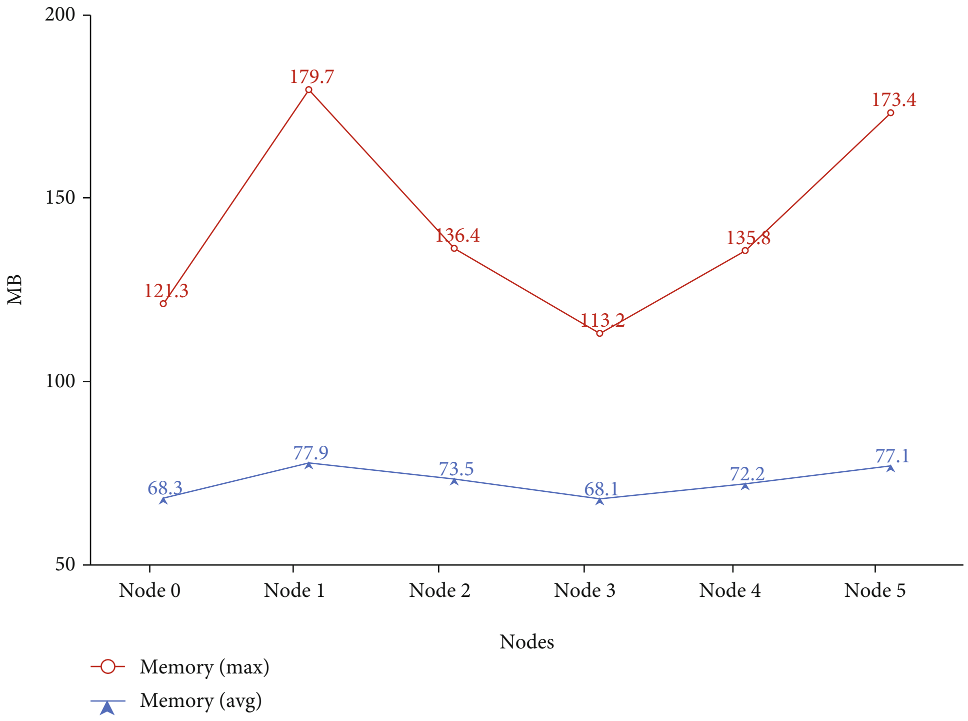


Figure 5: Memory usage of each node.

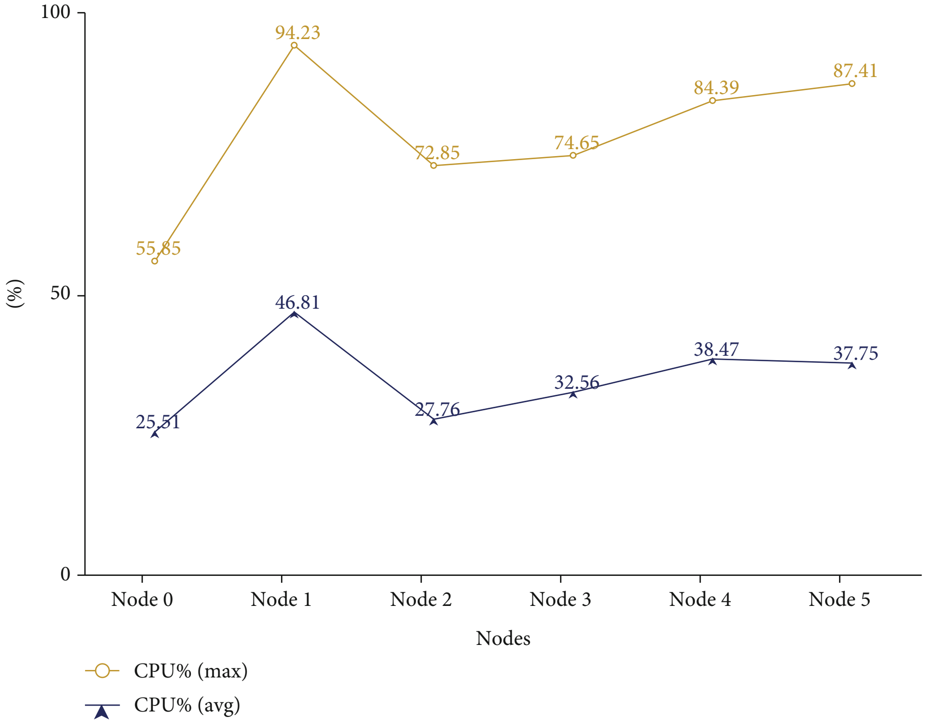


Figure 6: CPU usage of each node.

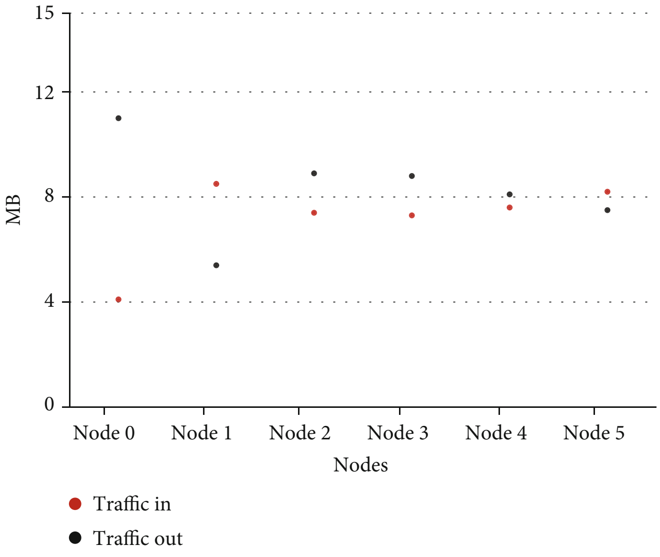


Figure 7: Traffic required for each node.

six nodes. For the smart contract of the blockchain, we chose the solidity language. This paper adopts Caliper as the test script to test the smart contract of consortium blockchain. The consortium blockchain is composed of a single group of six nodes. We select the scenario of 10000 concurrent transactions and 1000 new user registrations. The performance objects tested include memory usage, CPU usage, data traffic, disk read and write volume of each node, etc.

The performance metrics of send rate, latency, and throughput are described in Table 2. Figure 5 shows the memory usage of each node when processing data. Figure 6 shows the CPU usage of each node when verifying information. Figure 7 shows the traffic required for each node to form a consensus. Figure 8 shows the amount of traffic required by each node to form a consensus on the hard disk.

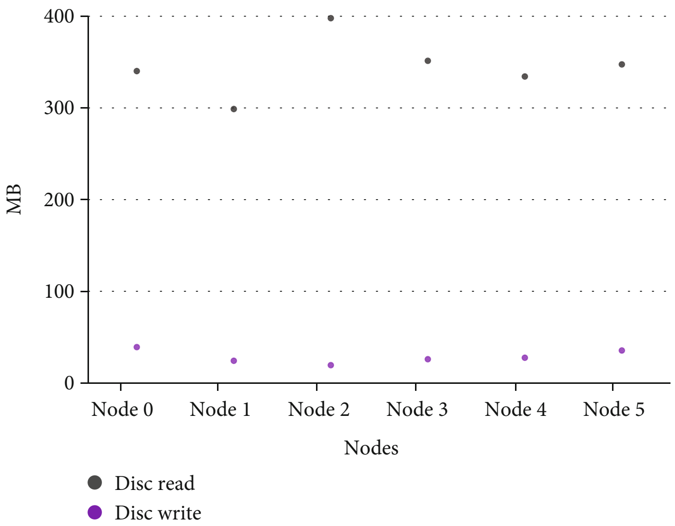


Figure 8: Disc read and write amount of each node.

Table 3: Symbols and descriptions of various operational times.

Symbols	Meaning
O_M	The time of a multiplication operation, 1 $O_M \approx 1.25$ ms
O_P	The time of a bilinear pairing operations, 1 $O_P \approx 32.23$ ms

Table 4: Performance comparison with other schemes.

Schemes	Signcryption cost	Unsigncryption cost	Execution time/($n = 1000$)	Confidentiality	Unforgeability
Iqbal et al. [28]	$4nO_M$	$nO_M + nO_P$	$5nO_M + nO_P$/71980 ms	√	√
Cui et al. [29]	nO_P	$2nO_M + 2nO_P$	$3nO_M + 2nO_{P/}$166090 ms	√	√
Hong et al. [30]	$n(3O_M + O_P)$	$n(3O_M + O_P)$	$6nO_M + 2nO_P$/8820 ms	√	√
Du et al. [31]	$4nO_M$	$4nO_M$	$8nO_M$/8820 ms	√	√
TB-SCDM	$2nO_M$	nO_M	$3nO_M$/2520 ms	√	√

6.2. Benchmark Test of Signcryption. The TB-SCDM and previous schemes [28–31] are exploited by the jPBC library on a laptop, where the configuration is a Windows 11 operating system, 2.60 GHz Intel(R) Core(TM) i7-9750H CPU with 16-GB RAM.

The meaning of the operation symbols is described in Table 3. The performance comparison of different schemes is described in Table 4.

A simple and intuitive method can be adopted in order to estimate the computation efficiency of the computational of several schemes. In terms of overall cryptographic operations, we can find that Iqbal et al. [28] is $5nO_M + nO_P$, Cui et al. [29] is $3nO_M + 2nO_P$, Hong et al. [30] is $6nO_M + 2n$ O_P, Du et al. [31] is $8nO_M$, and TB-SCDM is $3nO_M$. From the perspective of formula, the cost efficiency of TB-SCDM is the highest.

Figures 9 and 10 describe the execution time of different schemes when n changes from 100 to 1000. From the perspective of change range, it can be seen that when the number of users gradually increases, the computational efficiency of TB-SCDM is more obvious than other schemes.

In Figure 11, in order to compare various schemes more clearly, we specially select the execution time of signcryption, unsigncryption, and total operations when the number of users n equals 1000. The execution times of signcryption operations are as follows: the running time of Iqbal et al. [28] is $4 \times 1000 \times 1.25 = 5000$ ms, the running time of Cui et al. [29] is $1000 \times 32.23 = 32230$ ms, the running time of Hong et al. [30] is $1000 \times 3 \times 1.25 + 1000 \times 32.23 = 35980$ ms, the running time of Du et al. [31] is $4 \times 1000 \times 1.25 = 5000$ ms, and the running time of TB-SCDM is $2 \times 1000 \times 1.25 = 2500$ ms.

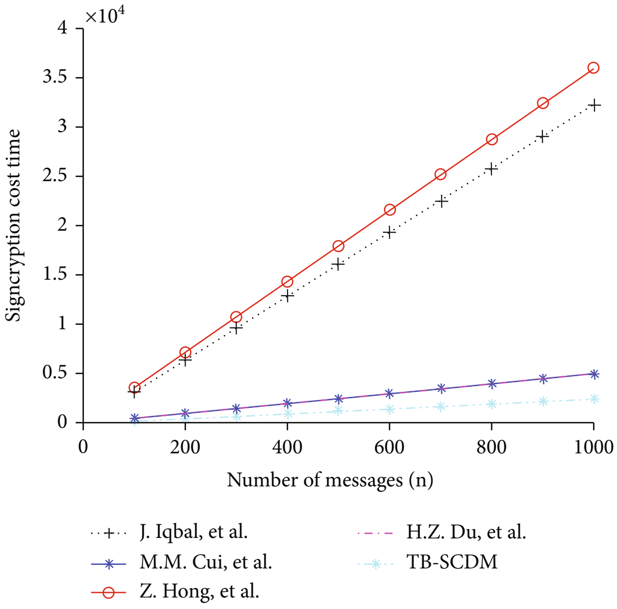


Figure 9: Comparison of signcryption cost.

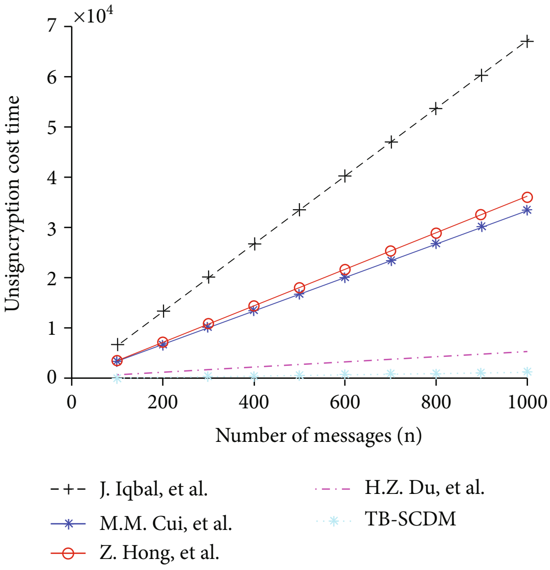


Figure 10: Comparison of unsigncryption cost.

The execution times of unsigncryption operations are as follows: the running time of Iqbal et al. [28] is 1000 × 1.25 + 1000 × 32.23 = 33480 ms, the running time of Cui et al. [29] is 2 × 1000 × 1.25 + 2 × 1000 × 32.23 = 66960 ms, the running time of Hong et al. [30] is 1000 × 3 × 1.25 + 1000 × 32.23 = 35980 ms, the running time of Du et al. [31] is 4 × 1000 × 1.25 = 5000 ms, and the running time of TB-SCDM is 1000 × 1.25 = 1250 ms.

The execution time of total operations are as follows: the running time of Iqbal et al. [28] is 5 × 1000 × 1.25 + 1000 × 32.23 = 38480 ms, the running time of Cui et al. [29] is 3 × 1000 × 1.25 + 2 × 1000 × 32.23 = 68210 ms, the running time of Hong et al. [30] is 6 × 1000 × 1.25 + 2 × 1000 × 32.23 = 71960 ms, the running time of Du et al. [31] is 8 × 1000 × 1.25 = 10000 ms, and the running time of TB-SCDM is 3 × 1000 × 1.25 = 3750 ms.

On the whole, the computational efficiency of TB-SCDM is faster than the other four schemes [28–31]. In terms of security and algorithm efficiency, TB-SCDM is very suitable for secure communication in VANETs.

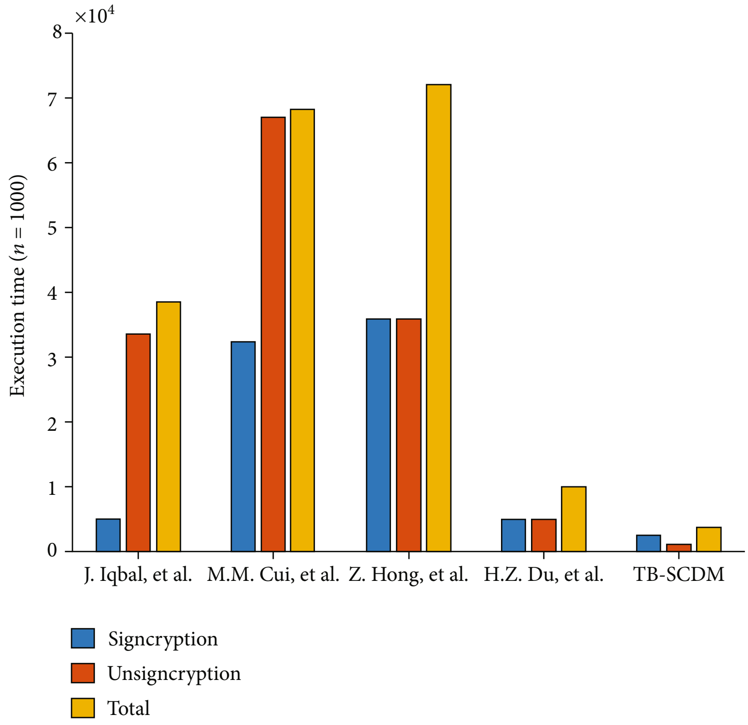


Figure 11: Comparison of total time ($n = 1000$).

7. Summary

In VANETs, SV using automatic driving need to access each other or RSU, GPS and other nodes to obtain reliable and stable data transmission services. Because VANET uses wireless communication, its openness allows attackers to easily obtain communication signals and further forge user nodes or Internet of Things nodes, which poses a greater security threat to SV. Based on the above reasons, this paper proposes a new trusted blockchain-based signaling protocol and data management for authentication and authorization. This scheme can effectively reduce the storage space occupied by information and the cost of signcryption verification.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jinqi Su and Runtao Ren have contributed equally to this work and should be considered co-first authors.

Acknowledgments

This work is supported by the Xi'an Soft Science Research Project under the grant 2021-0019; National Training Program of Innovation and Entrepreneurship for Undergraduates under the grant 202111664045; General Project of Humanities and Social Sciences Fund of the Ministry of Education (Youth Project) under the grant 20YJC630086; the Major Theoretical and Practical Research Project from Shaanxi Federation of Social Sciences Circles under the grant 20ZD195-144; the Research Grants Council of the Hong Kong Special Administrative Region, China (Project: CityU 11507219); the CityU SRG (Project: 7005780); Key Research Project on Major Theoretical and Practical Problems in Social Science Circles of Shaanxi Province under the grant SX-318; and Communication Soft Science Project of the Ministry of Industry and Information Technology under the grant R45.

References

[1] J. Feng, Y. Wang, J. Wang, and F. Ren, "Blockchain-based data management and edge-assisted trusted cloaking area construction for location privacy protection in vehicular networks," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2087–2101, 2021.

[2] F. Qu, Z. Wu, F. -Y. Wang, and W. Cho, "A security and privacy review of VANETs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2985–2996, 2015.

[3] R. Y. K. Lau, "Toward a social sensor based framework for intelligent transportation," in *2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 1–6, Macau, China, 2017.

[4] M. Arif, G. Wang, M. Zakirul Alam Bhuiyan, T. Wang, and J. Chen, "A survey on security attacks in VANETs: communication, applications and challenges," *Vehicular Communications*, vol. 19, article 100179, 2019.

[5] A. Festag, “Standards for vehicular communication—from IEEE 802.11p to 5G,” *e & i Elektrotechnik und Informationstechnik*, vol. 132, no. 7, pp. 409–416, 2015.

[6] M. Sepulcre, M. Gonzalez-Martín, J. Gozalvez, R. Molina-Masegosa, and B. Coll-Perales, “Analytical models of the performance of IEEE 802.11p vehicle to vehicle communications,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 713–724, 2022.

[7] A. T. Giang, A. Busson, A. Lambert, and D. Gruyer, “Spatial capacity of IEEE 802.11p-based VANET: models, simulations, and experimentations,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6454–6467, 2016.

[8] M. Dibaei, X. Zheng, Y. Xia et al., “Investigating the prospect of leveraging blockchain and machine learning to secure vehicular networks: a survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 683–700, 2022.

[9] C. Y. Yeung, L. C. K. Hui, T. W. Chim, S. Yiu, G. Zeng, and J. Chen, “Anonymous Counting Problem in Trust Level Warning System for VANET,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 34–48, 2019.

[10] S. Latif, S. Mahfooz, N. Ahmad et al., “Industrial Internet of Things based efficient and reliable data dissemination solution for vehicular ad hoc networks,” *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, pp. 1–16, 2018.

[11] A. Khalid, M. S. Iftikhar, A. Almogren, R. Khalid, M. K. Afzal, and N. Javaid, “A blockchain based incentive provisioning scheme for traffic event validation and information storage in VANETs,” *Information Processing & Management*, vol. 58, no. 2, article 102464, 2021.

[12] W. Luo and W. Ma, “Efficient and secure access control scheme in the standard model for vehicular cloud computing,” *IEEE Access*, vol. 6, pp. 40420–40428, 2018.

[13] Z. Lu, Q. Wang, G. Qu, H. Zhang, and Z. Liu, “A blockchain-based privacy-preserving authentication scheme for VANETs,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 12, pp. 2792–2801, 2019.

[14] Y. Yang, L. Zhang, Y. Zhao, K. K. R. Choo, and Y. Zhang, “Privacy-preserving aggregation-authentication scheme for safety warning system in fog-cloud based VANET,” *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 317–331, 2022.

[15] R. Guo, G. Yang, H. Shi, Y. Zhang, and D. Zheng, “O3-R-CP-ABE: an efficient and revocable attribute-based encryption scheme in the cloud-assisted IoMT system,” *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8949–8963, 2021.

[16] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, “Blockchain-based reliable and efficient certificateless signature for IIoT devices,” *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[17] Q. Wang, R. Y. K. Lau, and X. Mao, “Blockchain-enabled smart contracts for enhancing distributor-to-consumer transactions,” *IEEE Consumer Electronics Magazine*, vol. 8, no. 6, pp. 22–28, 2019.

[18] Y. Zhong and W. Wu, “A switching-based interference control for booster separation of hypersonic vehicle,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5560621, 9 pages, 2021.

[19] X. Liu, H. Huang, F. Xiao, and Z. Ma, “A blockchain-based trust management with conditional privacy-preserving announcement scheme for VANETs,” *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4101–4112, 2020.

[20] B. Luo, X. Li, J. Weng, J. Guo, and J. Ma, “Blockchain enabled trust-based location privacy protection scheme in VANET,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2034–2048, 2020.

[21] Z. Ma, J. Zhang, Y. Guo, Y. Liu, X. Liu, and W. He, “An efficient decentralized key management mechanism for VANET with blockchain,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 5836–5849, 2020.

[22] I. Dohare, K. Singh, A. Ahmadian, S. Mohan, and P. K. Reddy M, “Certificateless aggregated signcryption scheme for Cloud-Fog Centric Industry 4.0,” *IEEE Transactions on Industrial Informatics*, p. 1, 2022.

[23] L. Jiang, T. Li, X. Li, M. Atiquzzaman, H. Ahmad, and X. Wang, “Anonymous communication via anonymous identity-based encryption and its application in IoT,” *Wireless Communications and Mobile Computing*, vol. 2018, 8 pages, 2018.

[24] W. Luo and W. Ma, “Secure and efficient data sharing scheme based on certificateless hybrid signcryption for cloud storage,” *Electronics*, vol. 8, no. 5, pp. 590–601, 2019.

[25] X. Ye, G. Xu, X. Cheng, Y. Li, and Z. Qin, “Certificateless-based anonymous authentication and aggregate signature scheme for vehicular ad hoc networks,” *Wireless Communications and Mobile Computing*, vol. 2021, 16 pages, 2021.

[26] M. Lefebvre, S. Nair, D. W. Engels, and D. Horne, “Building a Software Defined Perimeter (SDP) for network introspection,” in *2021 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, pp. 91–95, Heraklion, Greece, 2021.

[27] K. Lei, Q. Zhang, L. Xu, and Z. Qi, “Reputation-based byzantine fault-tolerance for consortium blockchain,” in *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 604–611, Singapore, 2018.

[28] J. Iqbal, A. I. Umar, N. Amin, and A. Waheed, “Efficient and secure attribute-based heterogeneous online/offline signcryption for body sensor networks based on blockchain,” *International Journal of Distributed Sensor Networks*, vol. 15, no. 9, Article ID 155014771987565, 2019.

[29] M. Cui, D. Han, J. Wang, K. -C. Li, and C. -C. Chang, “ARFV: an efficient shared data auditing scheme supporting revocation for fog-assisted vehicular ad-hoc networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15815–15827, 2020.

[30] Z. Hong, F. Tang, and W. Luo, “Privacy-preserving aggregate signcryption for vehicular ad hoc networks,” in *Proceedings of the 2nd International Conference on Cryptography, Security and Privacy*, pp. 72–76, New York: ACM Press, 2018.

[31] D. U. Hongzhen, W. E. Qiaoyan, Z. H. Shanshan, and G. A. Mingchu, “A pairing-free certificateless signcryption scheme for vehicular ad hoc networks,” *Chinese Journal of Electronics*, vol. 30, no. 5, pp. 947–955, 2021.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1958984, 11 pages
https://doi.org/10.1155/2022/1958984

Research Article

Operation Stability Analysis of Basic Acupuncture Manipulation Based on Three-Dimensional Motion Tracking Data

Liu-Liu Xu,[1] Jian Xie,[2] Hua-Yuan Yang,[1] Fan Wang,[1] and Wen-Chao Tang[1]

[1]*School of Acupuncture-Moxibustion and Tuina, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[2]*Department of Acupuncture and Moxibustion, Yuhuan Hospital of Traditional Chinese Medicine, Tai-zhou, Zhejiang 317600, China*

Correspondence should be addressed to Fan Wang; 0000002681@shutcm.edu.cn and Wen-Chao Tang; vincent.tang@shutcm.edu.cn

Received 2 April 2022; Accepted 25 April 2022; Published 9 May 2022

Academic Editor: Maode Ma

Objective. To analyse the operational stability of basic acupuncture manipulation (AM) based on three-dimensional (3D) motion tracking. *Method*. Two quantitative indicators (spatial and temporal dispersions) and corresponding algorithms of operation stability were established based on the coordinate-time data derived from 3D motion tracking of basic AM. The differences in stability were compared between 20 acupuncture teachers and 20 acupuncture students. *Results*. The teachers and students had similar temporal stability, but the teachers were more stable in their spatial control, perhaps because of the teachers' better fingertip force and more practice with feedback. *Conclusion*. The spatial and temporal dispersions can be used to evaluate operational stability in basic AM. Repetitive training and finger force enhancement with more accurate feedback and rhythmic auditory stimulation are recommended for improving operation stability in basic AM.

1. Introduction

As a rhythmic medical skill, acupuncture manipulation (AM) is well known as an important basis for acupuncture treatment [1–3]. The performance of the acupuncturist's AM directly affects the patient's therapeutic effect so that AM has always been one of the difficulties in the education of traditional Chinese medicine (TCM) [4]. The basic AM includes lifting-thrusting and twisting [5, 6]; in clinical work, the acupuncturists are required to complete multiple cycles of the above skills in succession; and the finger movements in each cycle is required to be similar. This requirement is emphasized as operation stability during the teaching process of AM, and the main manifestation of poor operation stability is the obvious fluctuation of the operation frequency and amplitude. Therefore, operation stability is considered as a very important technical indicator [4] for the evaluation of AM. The traditional teaching mode of AM is instruction of teachers and self-training of students [7], and the operation stability is usually evaluated based on the observation and personal judgments of teacher. Quantitative evaluation data on this process to help students understand finger motion characteristics and how they differ from experts are lacking, resulting in a relatively small learning effects [8].

At present, quantitative evaluation research on AM mostly focuses on movement and force analysis during needling, including the motion amplitude, velocity, mechanical performance of the needle body and operational fingers [9], the distinctions of the above parameters in different AM skills [10], and the change in finger joint angles [11]. Some of the research achievements have applied to acupuncture education; for instance, the measured data of experts were used for training students, and the results showed a positive role of quantitative data in improving the AM performance of students [12]. However, few studies concern the operation stability of AM; the main reason is that the parameters that can be used to assess the stability have not been extracted from various existing measured data; and relevant comparative analysis based on these parameters need to be designed and conducted.

Throughout the current research progress, several kinematic and kinetic measurement technologies, including mechanical motion-electrical signal conversion [13], mechanical sensing [14], motion tracking [15], and ultrasonography [16, 17], have been applied to the quantitative study of AM. A large amount of kinematic and dynamic data, such as coordinate, velocity, acceleration, and force, has been obtained. Among them, motion tracking technology can provide the richest kinematic data without interfering with the operations of acupuncturists. Therefore, in this study, we defined the relevant quantitative indicators of operation stability from the coordinate-time data derived from three-dimensional (3D) motion tracking of basic AM. Moreover, the differences in the stability of basic AM between acupuncture professional teachers and students were compared. We hope that these efforts will provide new technical indicators and a quantitative reference for the evaluation of AM and improve the effects of classroom teaching and extracurricular training.

2. Materials and Methods

2.1. Participants. Twenty students and 20 teachers from the Acupuncture-Moxibustion and Tuina School of Shanghai University of Traditional Chinese Medicine (TCM) were recruited as the participants in this study. All the acupuncture teachers were required to have at least five years of clinical experience, and the students needed to have finished learning from the lifting-thrusting and twisting chapters in the course textbook, *Acupuncture and Moxibustion Techniques and Manipulations* [18], and to have hands-on needling experience with the human body. This study was approved by the ethics committee of Yueyang Hospital affiliated with Shanghai University of Traditional Chinese Medicine (reference no. 2021–062), and each participant signed an informed consent form.

In order to avoid experimental errors caused by the AM operation in different soft tissue environment of different human acupoints, this study uniformly used 0.30 × 40 mm acupuncture needles (Suzhou Medical Supplies Factory Co., Ltd.) to perform AM on a human tissue simulation model. All the participants were required to perform at least 10 cycles of the following three respective subtypes of lifting-thrusting and twisting skills.

2.1.1. Lifting-Thrusting Skill

(1) Mild reinforcing-attenuating: thrust and lift needle evenly

(2) Reinforcing: lift needle gently and slowly, and thrust needle forcefully and quickly

(3) Attenuating: thrust needle gently and slowly, and lift needle forcefully and quickly

2.1.2. Twisting Skill

(1) Mild reinforcing-attenuating: twist needle left and right evenly

(2) Reinforcing: twist needle left forcefully and quickly, and twist needle right gently and slowly

(3) Attenuating: twist needle right forcefully and quickly, and twist needle left gently and slowly

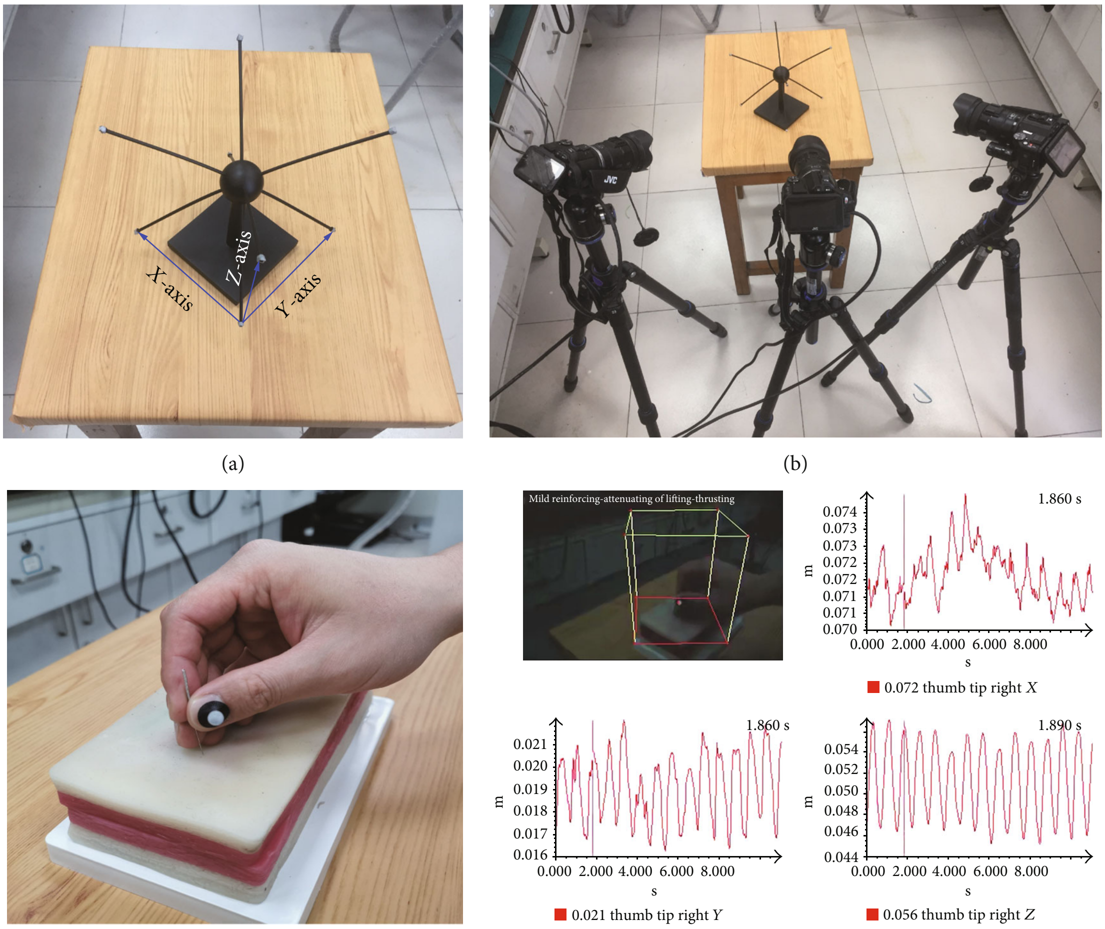
2.2. AM Measurement. The measurement of basic AM based on motion tracking technology was used for the stability evaluation. The experimental configuration was the same as that in our previous work [15]. Three sport cameras whose tripods were adjusted to the appropriate height were placed before the operation table; the shooting parameters of the cameras are as follows: resolution 1280 × 720 pixels, format MP4, full manual mode (M), aperture F1.2, shutter 1/1000s, ISO 6400, automatic white balance, and optical zoom 0 mm (Figure 1(a)).

Before the AM operation, a small 15 cm × 15 cm × 15 cm 3D calibration frame with 8 points was placed on the table for 3D calibration. Because the thumb tip is the main finger part for manipulating the needle (Figure 1(b)), a reflective ball with a diameter of 6.5 mm was attached on the center of thumb nail, used as the trace marker "thumb tip" (TT) for motion tracking (Figure 1(c)). The tracking point with the same name as the trace marker was also established in the motion analysis software Simi Motion 3D Ver 8.5 (Simi Motion, Simi Reality Motion Systems GmbH, Unterschleissheim, Germany). Simi Motion would automatically track the movement of the thumb tip during needling and record the 3D coordinate position of the tracking point at each sampling time node. After conducting analysis of the AM operation video of each participant, the X-, Y-, and Z-axis coordinate-time curves (Figure 1(d)) and related original data of TT of 9 operation cycles were exported for further processing. A video of the basic AM skills and their synchronized coordinate-time curves is also attached in the supplementary materials (Video 1).

2.3. Data Analysis. The operation stability of rhythmic skill movements such as AM evaluates whether the operator's performance of each action in every cycle is similar. It includes temporal and spatial stabilities. Temporal stability indicates whether a similar amount of time is used for each skill action in every cycle, and spatial stability determines whether each skill action has a similar operating trajectory. According to the action characteristics of basic AM and measured data exported by Simi Motion, the spatial stability can be judged according to whether each skill action in every cycle ends in a similar position. Therefore, two quantified stability parameters are established to evaluate the operator's stability performance:

(1) *Temporal dispersion*: the standard deviations of the time courses of thrusting and lifting actions or twisting-left and twisting-right actions were used to evaluate temporal stability (Figure 2(a))

(2) *Spatial dispersion*: the radius of the smallest sphere including all the end points reached by each skill action in the 9 operating cycles was used to evaluate spatial stability. In terms of its calculation idea, taking the twisting-left action during mild reinforcing-attenuating of twisting skill as an example, the red dots in Figure 2(b) show the 3D



(a) (b)

(c) (d)

Figure 1: Experimental configuration. (a) The positions of three cameras. (b) The placement of 3D calibration frame. (c) AM operation on human tissue simulation model with a trace marker on thumb tip. (d) The X- (upper right), Y- (left bottom), and Z-axis (right bottom) coordinate-time curves of TT, as well as the operation video (upper left) exported by Simi Motion.

distribution of the end points reached by each twisting-left action. If the smallest sphere includes all the end points, its radius can represent the dispersion of these points (Figure 2(c)). In order to calculate the sphere radius, the center of the sphere (blue point) should be located firstly by calculating the average values of the$X$, Y, and Z coordinates of each end point and then taking the maximum distance from each black point to the center as the radius. A video demonstrating this calculation idea is also attached in the supplementary material (Video 2)

The smaller the two types of dispersions were, the better the temporal and spatial stabilities. An original PHP script was used to calculate the above two dispersions based on the data exported by Simi Motion. All the source code has been shared in a GitHub repository (https://http://github.com/SHUTCM-tcme/AMA). The data process can be summarized as follows:

(1) Export the coordinate-time data from Simi Motion

(2) According to different operating skills, the coordinate-time data with significant motion characteristics along the corresponding axis was selected for the temporal dispersion calculation. In general, because TT mainly moves along the Z-axis during the lifting-thrusting skill and along the X- or Y-axis during twisting skill, thus, the Z-axis data was used in lifting-thrusting skill, and X- or Y-axis data was used in twisting skill

(3) Identify the inflection points of the coordinate-time curve for locating the crests and troughs. The interval between adjacent crest and trough is the operating time course of a skill action; then, record all the operating time courses of two skill actions in the operation cycle separately (Figure 2(a))

(4) Calculate the temporal dispersions (the standard deviations of the time courses) of different skill actions based on the corresponding operating time courses

(5) According to the sampling time nodes of the above crests and troughs, the 3D coordinate values of these

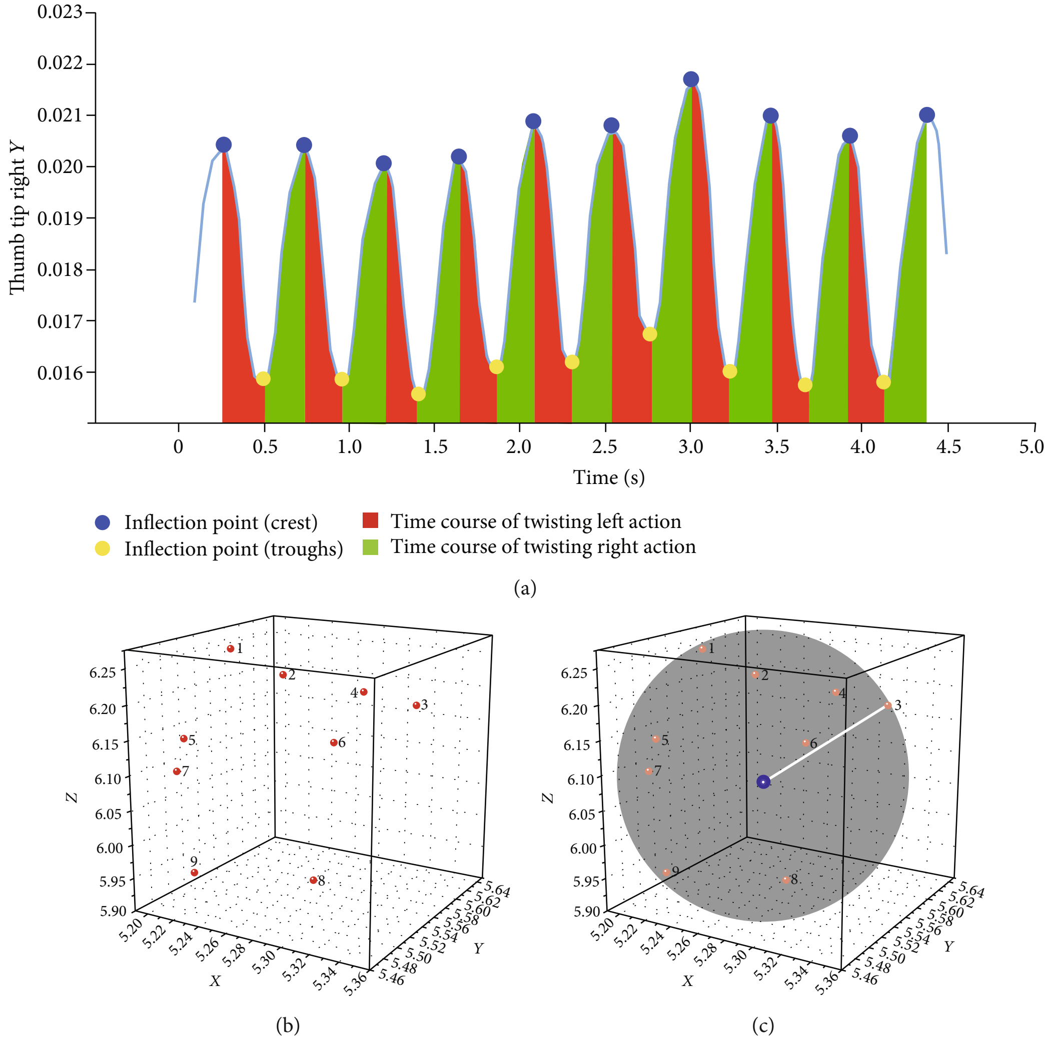


Figure 2: Illustration of the calculation idea of the temporal and spatial dispersions based on the twisting-left action of twisting skill. (a) The time dispersion is the standard deviation of the operation time course of each type of action. The calculation of the operation time course is based on the recognition of the inflection point of the curve. (b) The 3D distribution of the end points (red points) reached by each twisting-left action; the sequence number of the red dots is the cycle sequence of the twisting-left action. (c) The smallest sphere includes all the end points and its radius can be regarded as the spatial dispersion.

end points reached by each skill action in every operation cycle can be determined

(6) Calculate the spatial dispersions (the radius of the smallest sphere) of different skill actions based on the 3D coordinate values of the end points (Figures 2(b) and 2(c))

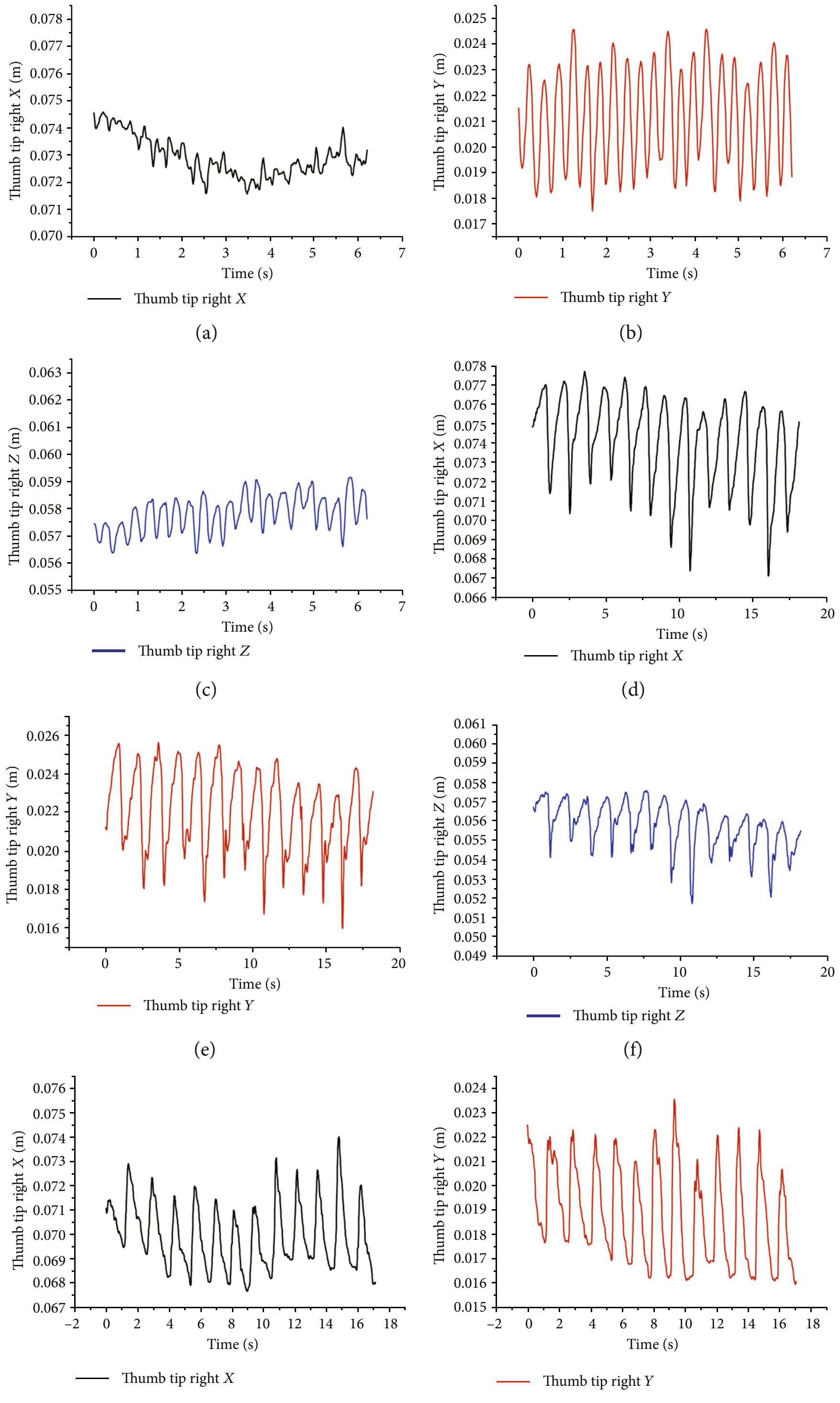
(7) Evaluate the operator's stability performance based on the results of temporal dispersions and spatial dispersions

2.4. Statistical Analysis. All outcomes were reported as the mean ± standard deviation. An analysis of independent-sample t-tests or rank-sum tests was used to assess differences between groups. The alpha level was established at $p < 0.05$ using the Statistical Package for the Social Sciences Ver.19 (SPSS, https://www.ibm.com/products/spss-statistics) to conduct all statistical analyses.

3. Results

The typical coordinate-time curves along three axes of all the subtypes of basic AMs are shown in Figures 3 and 4. The raw data of these curves was collected from the operation of a senior expert of Shanghai University of TCM. As one of the authoritative teachers of AM teaching in China, he is the judge of the AM event of the National Clinical Skill Competition of Acupuncture, Moxibustion & Tuina in Colleges and Universities of TCM held every year.

3.1. Lifting-Thrusting Skill. As shown in the lifting-thrusting part in Figures 5(a) and 5(b), the comparative analysis results between the two groups showed that except the temporal dispersion of thrusting action during reinforcing (63.20 ± 8.16 ms in teacher group vs. 99.03 ± 14.19 ms in student group, $p < 0.05$), teachers and students had similar temporal dispersion during the rest actions of three subtypes,



(a)

(b)

(c)

(d)

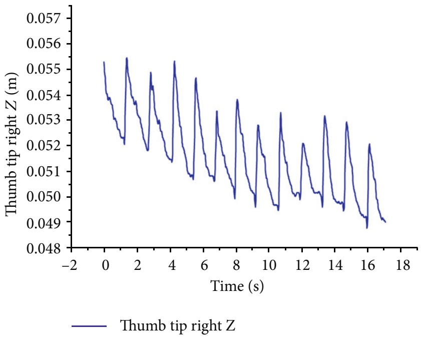
(e)

(f)

(g)

(h)

Figure 3: Continued.



(i)

Figure 3: Typical coordinate-time curves of TT during the twisting skill. The black, red, and blue curves are the typical X-, Y-, and Z-axis coordinate-time curves of TT during the twisting skill, respectively. (a–c), (d–f), and (g–i) show the corresponding curves of mild reinforcing-attenuating, reinforcing, and attenuating, respectively.

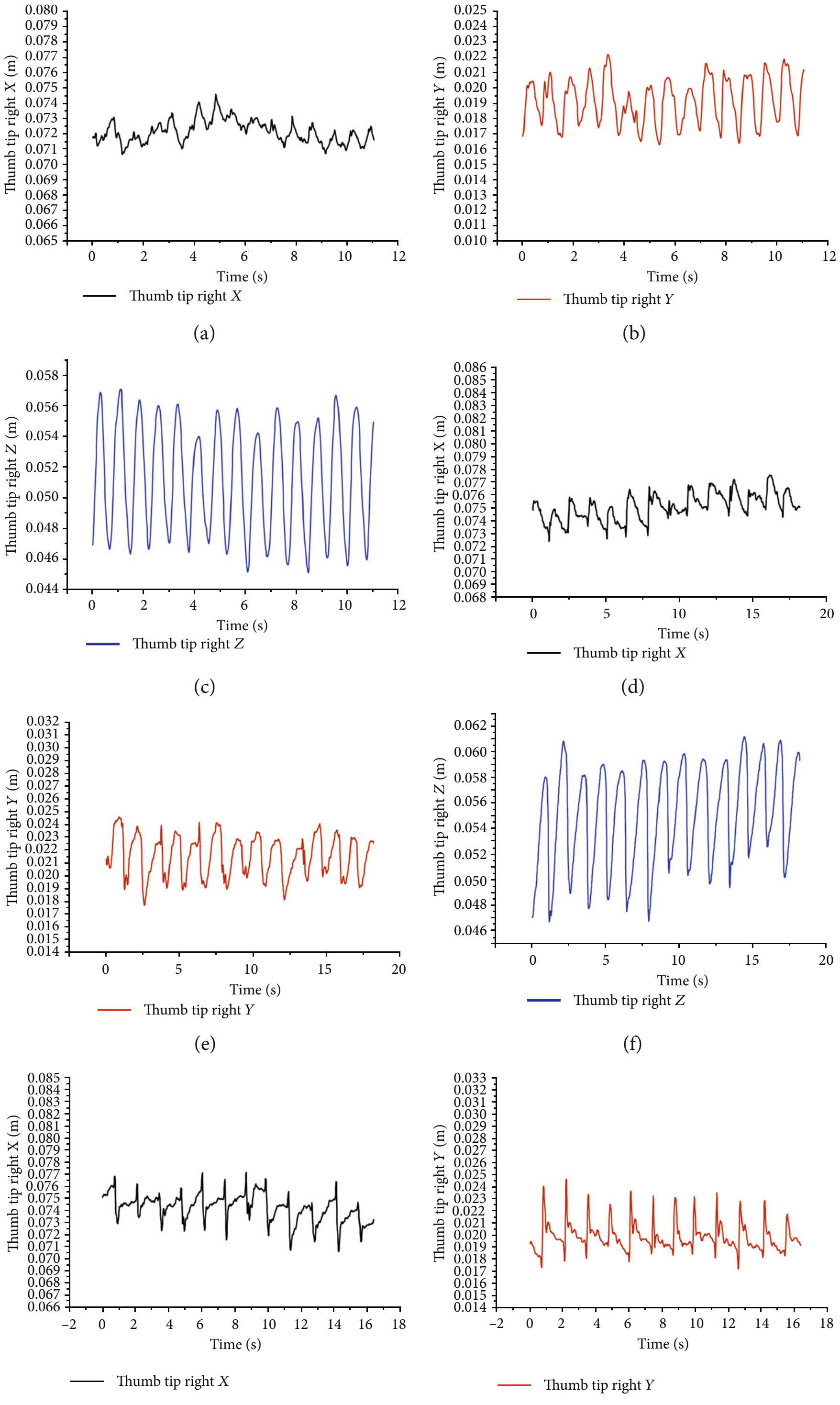
which suggested that the temporal stability of students was similar to that of teachers. In terms of the spatial stability, the teacher group had a lower spatial dispersion of lifting action during mild reinforcing-attenuating (1.99 ± 0.28 mm in the teacher group vs. 3.58 ± 1.15 mm in the student group, $p < 0.05$) and reinforcing (2.31 ± 0.29 mm in the teacher group vs. 3.07 ± 0.26 mm in the student group, $p < 0.05$). Hence, teachers had better ability to control the spatial position in the process of relatively rapid lifting.

3.2. Twisting Skill. The twisting part (Figures 5(c) and 5(d) shows that, similar to the results for the lifting-thrusting skill, there was no significant difference in the temporal dispersion during the most actions of three subtypes between the two groups; the only difference was found in the temporal dispersion of twisting-left action during attenuating (187.31 ± 20.01 ms in the teacher group vs. 401.00 ± 28.86 ms in the student group, $p < 0.05$). In terms of the spatial dispersion, lower spatial dispersion during the twisting-left action of three subtypes was found in the teacher group (2.12 ± 0.24 mm in teacher group vs. 3.45 ± 0.47 mm in student group during mild reinforcing-attenuating, $p < 0.05$; 2.23 ± 0.34 mm in the teacher group vs. 4.01 ± 0.36 mm in the student group during reinforcing, $p < 0.01$; and 4.22 ± 0.30 mm in the teacher group vs. 6.38 ± 0.41 mm in the student group during attenuating, $p < 0.01$). The teachers' better spatial stability is also maintained during the twisting-right action (1.88 ± 0.28 mm in the teacher group vs. 2.78 ± 0.28 mm in the student group during mild reinforcing-attenuating, $p < 0.01$; 1.87 ± 0.20 mm in the teacher group vs. 3.85 ± 0.33 mm in student group during reinforcing, $p < 0.01$; and 3.81 ± 0.40 mm in the teacher group vs. 5.70 ± 0.37 mm in the student group during attenuating $p < 0.01$). It suggested that the students' spatial control ability in the operation of twisting skill needed to be further improved.

4. Discussion

In the clinical application of manual acupuncture, the operation stability is one of the important factors affecting its therapeutic effect. During the process of rhythmic needling, excessive fluctuations in frequency and amplitude can easily cause discomfort to patients and even lead to fainting [19]. Poor spatial stability is also featured with too deep or too shallow operation amplitude, which may have resulted from the insufficient stimulation amount or the damage of important nerves or blood vessels, respectively [4]. Based on the coordinate data exported from motion analysis software, we established two types of parameters to analyse the operation stability of basic AM. According to the results of the comparative experiment of teachers and students, teachers generally had better stability than students, especially the spatial stability in rapid actions, which supports the general knowledge [20] and is in line with the results of some other comparative studies of AM [21, 22]. Therefore, these parameters can be regarded as technical indicators for evaluating the quality of AM. The better performance of teachers should mainly be attributed to their longer time spent practicing AM with abundant feedback. Lai et al. found that movement stability in generalized motor programs (GMPs) such as basic AM can be increased with constant practice and feedback, especially bandwidth knowledge of results (KR) [23]. Several studies also suggested a positive correlation between the amount of physical practice and motor performance [24] and the improvement of different types of feedback in motor control and learning [25, 26].

Another possible reason for this result is the greater force of the thumb and forefinger tips of teachers. Some mechanical sensor-based studies of AM have analysed finger force during needling, and their results showed that the force of teacher's fingertips on the needle handle has greater vertical and tangential components than the force of students' fingertips [14, 21]. As is known, force is the key to human motor control and is used to meet specific task goals [27]. During motor behaviour, humans interact with the environment using their various senses, such as the position and visual sense, and sensory feedback is constantly integrated into the central nervous system to coordinate the motion and force produced by arms or fingers [28]. Meanwhile, the control of body posture or limb position depends on muscle force. Studies have suggested that the spatial stability of limb movements can be improved by increasing muscle force through training [29]. Furthermore, the comparative



(a)

(b)

(c)

(d)

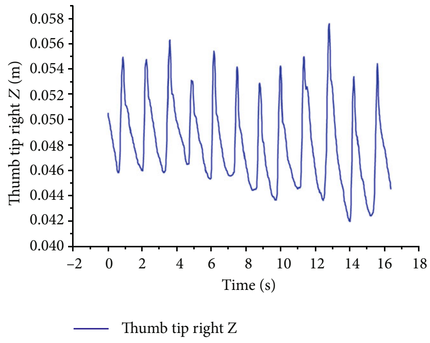
(e)

(f)

(g)

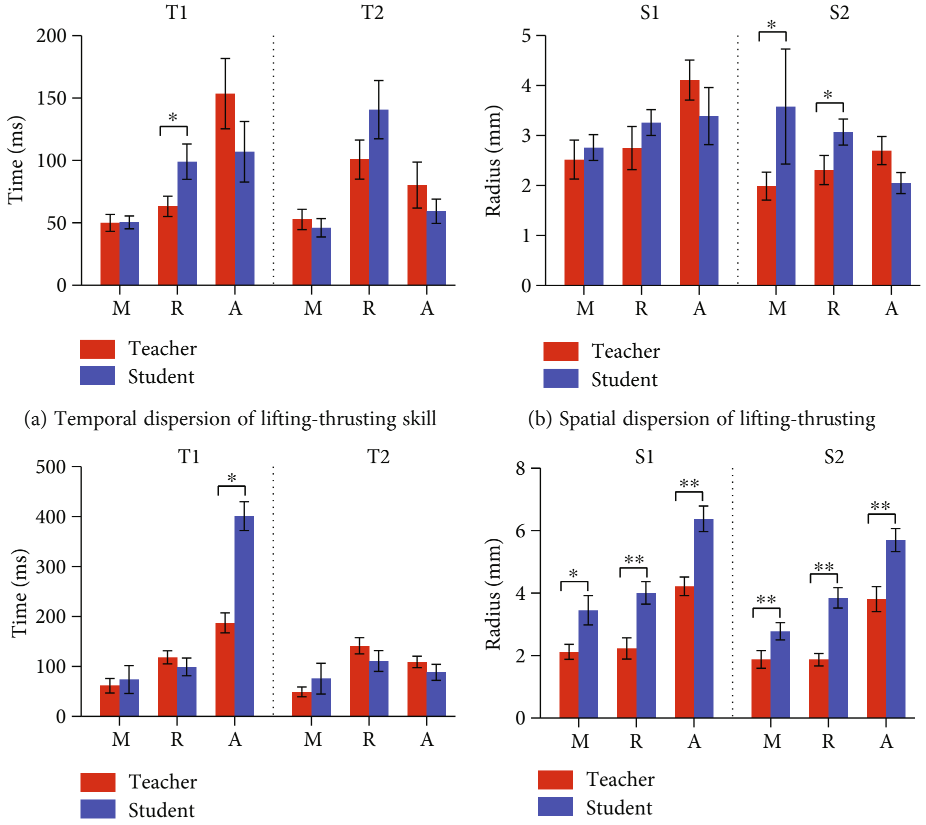
(h)

Figure 4: Continued.



(i)

Figure 4: Typical coordinate-time curves of TT during the lifting-thrusting skill. Figure legend refers to Figure 2.



(a) Temporal dispersion of lifting-thrusting skill

(b) Spatial dispersion of lifting-thrusting

(c) Temporal dispersion of twisting skill

(d) Spatial dispersion of twisting skill

Figure 5: Comparison of operation stability for basic AM between teacher and student. (a, b) and (c, d) showed the operation stability of lifting-thrusting skill and twisting skill, respectively. In each panel, T1 and S1 showed the temporal and spatial dispersions during thrusting action or twisting-left action; T2 and S2 showed the temporal and spatial dispersions during lifting action or twisting-right action. M, R, and A represented the subtype mild reinforcing-attenuating, reinforcing, and attenuating, respectively. * p <0.05; $^{**}p$ < 0.01.

results regarding muscle force and motor control ability for young and old people showed that older people have weakened control ability to control limb positions because of the decline in muscle force; this also provided evidence for the close relationship between spatial stability and muscle force during movement [30].

Another finding of this study is that no significant difference was found in the temporal dispersion during most skill actions between the two groups. Thus, students and teachers have a similar ability to control the cycle stability of rhythmic movement, and compared with spatial stability, consistent performance in temporal stability with experts can be achieved in a relatively short period of training. Many studies have explored the control mechanisms of vertebrate rhythmic movement and revealed that rhythmic activities are produced from the central pattern generators (CPGs) in the spine [31, 32]. The CPGs not only produce rhythms but also alter their frequencies and patterns; unlike spatial control, this control process does not require sensory signals from peripheral receptors for feedback regulation [33]. Moreover, an interesting numerical model developed in early research has demonstrated the importance of CPGs in human rhythmic movement, not only in maintaining stability against perturbations but also in controlling velocity [34]. Therefore, the feedback-independent characteristics of CPGs may contribute to the rapid improvement in temporal stability in the student group.

According to the above results, four approaches can be considered for improving the operational ability of basic AM. The first is repetitive training. Many studies in different fields, such as the sports education [35, 36] and music [37], have suggested that repetitive training is one of the key factors in the enhancement of motor performance, as well as the stability of continuous periodic skills. Our study also found that teachers with more practice have better stability than do students. Thus, repetitive training should be the priority option for improving stability in AM [38]. The second approach is providing appropriate feedback. It has been shown that feedback, including inherent and augmented feedback, can effectively enhance not only students' cognitive levels [25] but also their motor control [39, 40]. Furthermore, the more accurate quantified data in feedback are provided, the better the learning effects for [41]. Some studies also reported that students' self-efficacy is likely to increase when feedback is accurate [42]. Based on these findings, the data we measured can be used as feedback for enhancing motor control in basic AM. The third approach concerns finger force enhancement. Possible solutions include fingertip pressing and gripping training performed once or twice a week, for example, squeezing the Digi-Flex hand training device (IMC Products, Hicksville, NY) with the thumb and forefinger tips [43]. Arm strength training is also an option because finger force is inseparable from the support of the palm and arm, especially for spatial control. Studies have illustrated that a nonspecific upper-limb strength-training program may improve finger-pinch force control in older men [43, 44] and increase finger coordination in skill-specific training [45]. The fourth approach is rhythmic auditory stimulation (RAS), which is often used in training for music, dance, and sports, and rhythm perception and synchronization can help humans predict motion trajectories and improve spatial control sequentially [46]. Although students have relatively good temporal stability, RAS is still recommended for training to improve spatial control capabilities in basic AM.

AM training is a long-term process [47]. We believe that further analysis of more quantitative parameters based on existing measured data will be conducive to enhancing students' motor learning and control more quickly and promote innovation in education related to traditional Chinese medicine skills.

5. Conclusions

Spatial and temporal dispersions of coordinate data can be used to evaluate operation stability in basic AM. The comparison between teachers and students showed that the two groups have similar temporal stability in most skill actions of AM, but the teachers have more stable spatial control. The main reason for this result may lie in the teachers' greater practice with feedback and better fingertip force. Therefore, repetitive training and finger force enhancement with more accurate feedback and RAS are recommended for improving motor performance and control in basic AM.

Data Availability

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

Conflicts of Interest

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

Authors' Contributions

Liu-Liu Xu and Jian Xie contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Number 82174506).

Supplementary Materials

Video 1: video of the basic AM skills and their synchronized typical coordinate-time curves. A video includes three respective subtypes of lifting-thrusting and twisting skills (top left) and the corresponding synchronized typical coordinate-time along the X- (top right), Y- (bottom left), and Z- (bottom right) axis. Video 2: video of the calculation idea of spatial dispersion based on the twisting-left action of twisting skill. A video includes the acquisition of all the end points; the location of the center of the smallest sphere includes all the end points and the calculation of its radius. *(Supplementary Materials)*

References

[1] Y. J. Choi, J. E. Lee, W. K. Moon, and S. H. Cho, "Does the effect of acupuncture depend on needling sensation and manipulation?," *Complementary Therapies in Medicine*, vol. 21, no. 3, pp. 207–214, 2013.

[2] H. Yu, X. Li, X. Lei, and J. Wang, "Modulation effect of acupuncture on functional brain networks and classification of its manipulation with EEG signals," *IEEE Trans Neural Syst Rehabil Eng*, vol. 27, no. 10, pp. 1973–1984, 2019.

[3] G. Yang, G. Yongming, Z. Jiatai, X. Ying, and G. Yi, "Professor Guo Yi's acupuncture manipulation and clinical application," *World Chinese Medicine*, vol. 15, no. 11, pp. 1624–1628, 2020.

[4] R. Lyu, M. Gao, H. Yang, Z. Wen, and W. Tang, "Stimulation parameters of manual acupuncture and their measurement," *Evidence-Based Complementary and Alternative Medicine*, vol. 2019, Article ID 1725936, 2019.

[5] L. Fang-jie, Y. Hua-yuan, and W. Guan-tao, "Overview of mechanical research on basic acupuncture manipulation,"

Shanghai Journal of Acupuncture and Moxibustion, vol. 34, no. 1, pp. 85–87., 2015.

[6] W. Ya-lin and C. Bo, “Briefly mention in twisting acupuncture for the clinical application of catharsis,” *Asia-Pacific Traditionnal Medicine*, vol. 13, no. 19, pp. 79-80, 2017.

[7] L. Lanying, W. Hesheng, Z. Cong, W. Wenzhong, Z. Xia, and G. Yihuang, “Preliminary study 354 on the role of master-apprentice education of traditional Chinese medicine in the cultivation of 355 professional master postgraduates majoring in acupuncture and moxibustion,” *China Medicine and Pharmacy*, vol. 8, no. 18, pp. 24–27, 2018.

[8] B M Association, *Acupuncture: Efficacy, Safety and Practice*, Routledge, 2020.

[9] B. Jin-ling and Z. Chun-hong, “Conception and core of academician Shi Xuemin's acupuncture manipulation quantitative arts,” *Chinese Acupuncture & Moxibution*, vol. 5, pp. 38–40, 2003.

[10] T. Y. Liu, H. Y. Yang, L. Kuai, and G. Ming, “Classification and characters of physical parameters of lifting-thrusting and twirling manipulations of acupuncture,” *Acupuncture Research*, vol. 35, no. 1, pp. 61–66, 2010.

[11] Y. Peng, S. Xiao-wen, M. Ya-kun, Z. Chun-xin, and Z. Wenguang, “Quantification research on acupuncture manipulation based on video motion capture,” *Journal of Medical Biomechanics*, vol. 31, no. 2, pp. 154–159, 2016.

[12] P. Friedl and K. Wolf, “Proteolytic interstitial cell migration: a five-step process,” *Cancer and Metastasis Reviews*, vol. 28, no. 1-2, pp. 129–135, 2009.

[13] T. L. Yi, H. Y. Yuan, K. Le, G. Ming, and X. Gang, “Exploitation and application of acupuncture manipulation information analysis system,” *China Acupuncture*, vol. 22, no. 11, pp. 927–930, 2008.

[14] R. T. Davis, D. L. Churchill, G. J. Badger, J. Dunn, and H. M. Langevin, “A new method for quantifying the needling component of acupuncture treatments,” *Acupuncture in Medicine*, vol. 30, no. 2, pp. 113–119, 2012.

[15] W. C. Tang, H. Y. Yang, T. Y. Liu, M. Gao, and G. Xu, “Motion video-based quantitative analysis of the 'lifting-thrusting' method: a comparison between teachers and students of acupuncture,” *Acupuncture in Medicine*, vol. 36, no. 1, pp. 21–28, 2018.

[16] M. Q. H. Leow, S. L. Cui, M. T. B. Mohamed Shah et al., “Ultrasonography in acupuncture-uses in education and research,” *Journal of Acupuncture and Meridian Studies*, vol. 10, no. 3, pp. 216–219, 2017.

[17] M. Q. Leow, T. Cao, S. H. Lee, S. L. Cui, S. C. Tay, and C. C. Ooi, “Ultrasonography in acupuncture: potential uses for education and research,” *Acupuncture in Medicine*, vol. 34, no. 4, pp. 320–322, 2016.

[18] F.-c. Wang, *Acupuncture and Moxibustion Techniques and Manipulations*, Shanghai Science and Technology Press, Shanghai, 2009.

[19] J.-J. Wen and H. Chou, “Integration of Chinese and Western medicine in fainting during acupuncture treatment,” in *In Smart Science, Design & Technology*, pp. 109–112, CRC Press, 2019.

[20] C. Wang, “A randomized crossover trial and methodological discussion to evaluate the effect of acupuncture manipulation by acupuncturists with different qualifications,” in *Acupuncture and Moxibustion and Tuina of Chinese Medicine*, Beijing University of Chinese Medicine, 2016.

[21] J. Li, L. E. Grierson, M. X. Wu, R. Breuer, and H. Carnahan, “Perceptual motor features of expert acupuncture lifting-thrusting skills,” *Acupuncture in Medicine*, vol. 31, no. 2, pp. 172–177, 2014.

[22] S.-Y. Lee, Y.-N. Son, I.-h. Choi, K.-M. Shin, K.-S. Kim, and S.-D. Lee, “Quantitative study of acupuncture manipulation of lifting-thrusting using an needle insertion-measurement system in phantom tissue,” *Journal of Korean Medicine*, vol. 35, no. 4, pp. 74–82, 2014.

[23] Q. Lai, C. H. Shea, G. Wulf, and D. L. Wright, “Optimizing generalized motor program and parameter learning,” *Research Quarterly for Exercise and Sport*, vol. 71, no. 1, pp. 10–24, 2000.

[24] A. Urcuyo-Ovares, J. Ávila-Chaverri, J. Jiménez-Díaz Ph, and B. Montero-Herrera, “Effect of mental and physical practice in the motor performance and muscle electrical activity in healthy students,” *Pensar en Movimiento: Revista de ciencias del ejercicio y la salud*, vol. 18, no. 1, pp. 136–155, 2020.

[25] H. Hicheur, A. Chauvin, V. Cavin, J. Fuchslocher, M. Tschopp, and W. Taube, “Augmented-feedback training improves cognitive motor performance of soccer players,” *Medicine & Science in Sports & Exercise*, vol. 52, no. 1, pp. 141–152, 2020.

[26] R. Gilgen-Ammann, T. Wyss, S. Troesch, L. Heyer, and W. Taube, “Positive effects of augmented feedback to reduce time on ground in well-trained runners,” *International Journal of Sports Physiology and Performance*, vol. 13, no. 1, pp. 88–94, 2018.

[27] B. Lauber, M. Keller, C. Leukel, A. Gollhofer, and W. Taube, “Force and position control in humans-the role of augmented feedback,” *JoVE (Journal of Visualized Experiments)*, vol. 112, p. e53291, 2016.

[28] G. Ballardini, V. Ponassi, E. Galofaro et al., “Interaction between position sense and force control in bimanual tasks,” *Journal of Neuroengineering and Rehabilitation*, vol. 16, no. 1, pp. 1–13, 2019.

[29] E. Chu, Y. S. Kim, G. Hill, Y. H. Kim, C. K. Kim, and J. K. Shim, “Wrist resistance training improves motor control and strength,” *The Journal of Strength & Conditioning Research*, vol. 32, no. 4, pp. 962–969, 2018.

[30] S. H. Park, M. Kwon, D. Solis, N. Lodha, and E. A. Christou, “Motor control differs for increasing and releasing force,” *Journal of Neurophysiology*, vol. 115, no. 6, pp. 2924–2930, 2016.

[31] F. Haque and S. Gosgnach, “Mapping connectivity amongst interneuronal components of the locomotor CPG,” *Frontiers in Cellular Neuroscience*, vol. 13, p. 443, 2019.

[32] F. Dzeladini, J. Van Den Kieboom, and A. Ijspeert, “The contribution of a central pattern generator in a reflex-based neuromuscular model,” *Frontiers in Human Neuroscience*, vol. 8, p. 371, 2014.

[33] P. S. Katz, “Evolution of central pattern generators and rhythmic behaviours,,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1685, p. 20150057, 2016.

[34] G. Taga, “Emergence of bipedal locomotion through entrainment among the neuro-musculo-skeletal system and the environment,” *Physica D: Nonlinear Phenomena*, vol. 75, no. 1-3, pp. 190–208, 1994.

[35] X. Lu-lu, “Talking about the practical application of repetitive training method in high school sprint training,” *Track and Field*, vol. 41, no. 5, pp. 62–64, 2020.

[36] L. Liang, “Research on the technique of repetitive training to improve the acceleration after the start of 100 meters,” *Science*

& Technology of Stationery & Sporting Goods, vol. 41, no. 19, pp. 171-172, 2020.

[37] Z. Xu, "Probe into finger training methods and techniques in piano performance," *Northern Music*, vol. 40, no. 8, pp. 46-47, 2020.

[38] T. Y. Liu, H. Y. Yang, K. Le, M. Gao, Y. E. Hu, and G. Xu, "application of "acupuncture manipulation information analyzing system " in acupuncture manipulation education," *Chinese Acupuncture & Moxibution*, vol. 29, no. 11, pp. 927–930, 2009.

[39] M. Frikha, N. Chaâri, Y. Elghoul, H. H. Mohamed-Ali, and A. V. Zinkovsky, "Effects of combined versus singular verbal or haptic feedback on acquisition, retention, difficulty, and competence perceptions in motor learning," *Perceptual and Motor Skills*, vol. 126, no. 4, pp. 713–732, 2019.

[40] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 21–53, 2013.

[41] C. Krishnan, E. P. Washabaugh, and Y. Seetharaman, "A low cost real-time motion tracking approach using webcam technology," *Journal of Biomechanics*, vol. 48, no. 3, pp. 544–548, 2015.

[42] A. García-Dantas and E. Quested, "The effect of manipulated and accurate assessment feedback on the self-efficacy of dance students," *Journal of Dance Medicine & Science*, vol. 19, no. 1, pp. 22–30, 2015.

[43] H. B. Olafsdottir, V. M. Zatsiorsky, and M. L. Latash, "The effects of strength training on finger strength and hand dexterity in healthy elderly individuals," *Journal of Applied Physiology*, vol. 105, no. 4, pp. 1166–1178, 2008.

[44] J. W. Keogh, S. Morrison, and R. Barrett, "Strength training improves the tri-digit finger-pinch force control of older adults," *Archives of Physical Medicine and Rehabilitation*, vol. 88, no. 8, pp. 1055–1063, 2007.

[45] J. K. Shim, J. Hsu, S. Karol, and B. F. Hurley, "Strength training increases training-specific multifinger coordination in humans," *Motor Control*, vol. 12, no. 4, pp. 311–329, 2008.

[46] S. L. Bengtsson, F. Ullen, H. H. Ehrsson et al., "Listening to rhythms activates motor and premotor cortices," *Cortex*, vol. 45, no. 1, pp. 62–71, 2009.

[47] W. Xiao-song and W. Juan, "A discussion on the training methods for the stability of diving movement," *Journal of Capital University of Physical Education and Sports*, vol. 17, no. 4, pp. 84-85, 2005.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2947554, 7 pages
https://doi.org/10.1155/2022/2947554

Research Article

Research on the Fusion Model of Professional Vocal Music Performance Voice Care and Artificial Intelligence Technology in Intelligent Medical Treatment

Wei Qin and Qingna Lin

Chongqing University of Posts and Telecommunications, Chongqing South Bank, 400065, China

Correspondence should be addressed to Qingna Lin; linqn@cqupt.edu.cn

Received 21 March 2022; Accepted 13 April 2022; Published 9 May 2022

Academic Editor: Maode Ma

Intelligent medical treatment is an important research field in today's world. Artificial intelligence technology is the key factor to construct intelligent medical treatment. In the development of artificial intelligence technology, it is necessary to establish a scientific, systematic, and comprehensive system analysis model, inevitably with certain professional characteristics. At present, in the research of vocal health care in professional vocal music performance, the application of intelligent medical care and vocal health care in professional vocal music performance is studied. According to the DEMATEL-ISM research method, this paper constructs 4 internal and external factors and 16 influencing factors to build a comprehensive and systematic weight analysis model, which provides a theoretical and practical basis for the scientific construction of AI technology algorithms. The aim is to improve the value and research significance of intelligent medical artificial intelligence technology in professional vocal music performance sound care.

1. Introduction

For professional vocal singers, "human voice" is their most important natural "musical instrument," and how to protect and care for this "musical instrument" is of great value and significance to singers. According to Chen [1], "scientific sound health care is of great significance to comprehensively improve the technical level and singing quality of singers." How to get a beautiful singing voice depends on the following factors: healthy voice, scientific voice, and good sense of music. Good voice plays a very important role in it, and the larynx vocal cords are an important vocal organ of human beings to produce scientific vibration, promoting the production of sound (voice), and then, by the natural formation of various cavities in the body, resonance is produced to achieve the transmission of sound and ultimately achieve a beautiful voice. Professional vocal music performers master a method of controlling their body vocal function or their required timbre through their own scientific training. However, for a long time, voice health care relies more on their own knowledge, cognition, and traditional medical tests to carry out regular maintenance, which cannot well produce early warning and prompting effect for solving many chronic diseases such as voice pathological congestion, acute laryngitis, vocal nodules, and vocal polyps. More importantly, when the disease is often found, it is based on the very serious condition; singers in the traditional disease prevention and voice health management have a certain lag.

2. Research Status of Voice Health Care in Professional Vocal Music Performance

Chinese bel canto singers pointed out that the problems of white voice, tremolo, throat, nasal, and straight voice are the key to the voice problems of bel canto singers; Gunjawate et al. [2], through statistical research, analyze the influence of the popular folk arts in Karnataka, India, including singing and dancing, on the health of the voice. It is preliminarily concluded that both innate and acquired voice problems will affect the voice health of singers emotionally; Devadas et al. [3]

explored the relationship between voice problems and different health problems such as dental problems, frequent colds, hearing difficulties, occupational nature, and regular medication in the study of Western opera singing methods; in Baird et al. [4], acoustic and laryngoscopic evaluation of voice health care in a cappella choirs in colleges and universities shows that there is little difference between professional and nonprofessional singers in voice health care, but there are differences in access to health-related information, resulting in certain differences; Arunachalam et al. [5] evaluated the voice problems caused by the training of a classical music style. Through research, they found that voice change, high pitch difficulty, voice fatigue, pharyngitis, muscle tension, dysphonia, chronic pharyngitis, and so on are the key to voice problems; Tepe et al. [6] investigated the voice problems of chorus singers by questionnaire and found that the key factors causing chorus voice problems are morning hoarseness, chronic fatigue, insomnia, emotional tension, adolescent girls' physiological period, and so on; Flavia et al. [7] established a 20-item singer's singing ability assessment scale, which includes the singer's background, education, work experience, professional direction, etc. Quantitative analysis of physical health, singing level, and other aspects is performed; García and López [8] demonstrate that flamenco singers' vocal habits, behavioral differences, drinking, smoking, practice, speaking, and other habits are exposed to many voice health risk factors, which make them more prone to vocal fatigue. Mucosal dehydration, voice damage, and muscle stiffness are more common in classical singers; according to Irene and Wilson [9], the "daily life" of professional singers has an important impact on the health of singers' voices.

From the above analysis and research, we can see that the research on voice health care of professional vocal performers at home and abroad is very rich, but today, with the rapid development of artificial intelligence technology, the intelligent medical treatment system formed by combining artificial intelligence technology has not been studied for voice health care of professional vocal performers. There are significant differences between the research scope and artificial intelligence methods in the specific process of large data analysis and research, which are important for daily voice health care and professional vocal performers. We should study and analyze them separately in order to establish a reasonable artificial intelligence model. The authors will use the Decision-Making Trial and Evaluation Laboratory (DTEL), which is considered to be an effective method for identifying causal chains in complex systems. DEMATEL and Interpretive Structure Model (ISM) establish evaluation weights in a combination of qualitative and quantitative ways. At the same time, a multilevel structure model of professional vocal performers is obtained through data analysis. After independent analysis of the influencing factors, multiple factors are comprehensively integrated, and the relevance between factors and elements is considered, so as to establish a systematic sound health management model. In the intelligent medical treatment, the professional vocal performers can perceive the comprehensive interactive effect of artificial intelligence technology and realize the comprehensive management of voice health. This study is of great practical value and significance.

3. Extraction of Factors Affecting Vocal Health Care of Professional Vocal Actors

In order to comprehensively analyze the various reasons that affect the voice health of professional vocal music actors, 57 vocal music teachers, students, and actors were sampled to collect the voice health problems caused by their different perspectives, and a comprehensive system of influencing factors was constructed from both internal and external dimensions, combined with literature research and in-depth interviews with research subjects. The overall analysis system is constructed comprehensively, and the four elements of life, body, training, and society are selected in the final competition. There are 16 influencing factors, such as dietary habits, physical fitness, scientific degree of vocal production, degree of speaking, work and rest habits, health degree of vocal organs, frequency of practice, daily workload, exercise habits, mental health degree, frequency of daily performance, intensity of social pressure, habit of using voice, degree of mental identification, level of vocal singing, and degree of social demand, which are shown in Figure 1.

4. Construction of the Influence System of Vocal Health Care in Professional Vocal Performance and Analysis of the Factors

After sorting out the four elements and sixteen influencing factors, the overall evaluation system is constructed, which is divided into three levels: target level, criterion level, and element level, as shown in Table 1.

4.1. Life Element. Dietary habits (A1): reasonable dietary habits play an important role in the protection of voice in professional vocal music performance. Li [10] mentioned the important value and significance of dietary habits for professional vocal music performance in his research, so dietary habits are one of the important reference indicators.

Work and rest habits (A2): reasonable work and rest habits will affect the fatigue and health of voice in professional vocal music performance. In the study of Xu [11], the key value of work and rest habits for voice health care is very important. It is also very important to take it as an important reference index in the internal influencing factors.

Exercise habits (A3): maintaining good exercise habits plays an important role in reasonable voice health care. Therefore, Yang [12] has a certain demonstration on the role of exercise habits in the prevention and treatment of voice diseases. On the whole, good exercise habits have a certain monitoring index value for promoting scientific voice health care.

Voice habit (A4): voice habit is mainly reflected in the daily use of voice, which is a very dynamic personal habit. This key role value is mentioned in many professional vocal performers' voice health care. Therefore, it is necessary to adjust the corresponding voice habit, which is an indispensable key indicator factor in the big data of intelligent medical treatment.

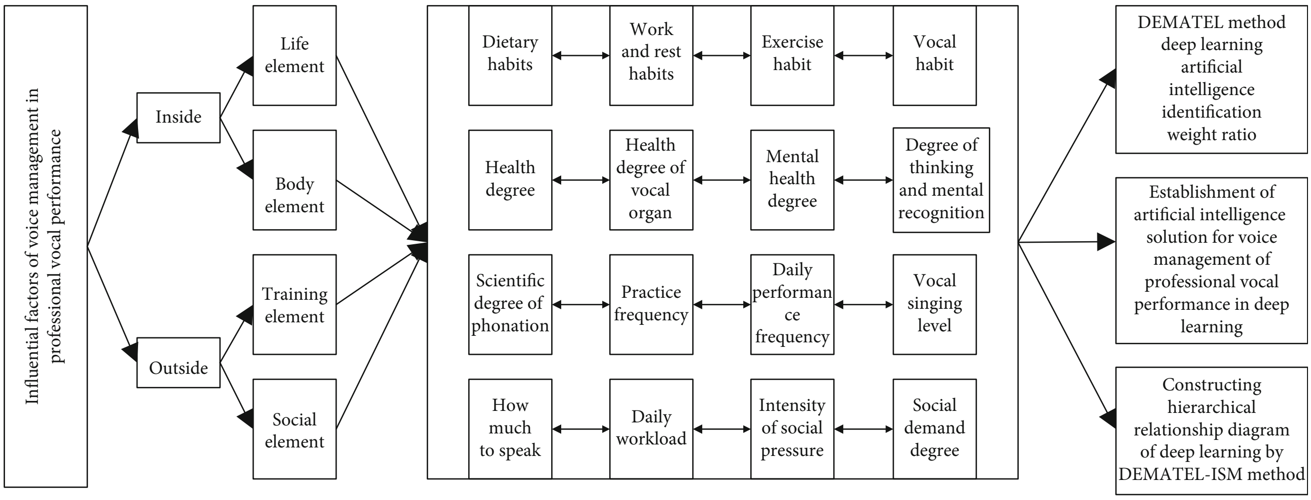

Influential factors of voice management in professional vocal performance
Inside
Outside
Life element
Body element
Training element
Social element
Dietary habits
Work and rest habits
Exercise habit
Vocal habit
Health degree
Health degree of vocal organ
Mental health degree
Degree of thinking and mental recognition
Scientific degree of phonation
Practice frequency
Daily performance frequency
Vocal singing level
How much to speak
Daily workload
Intensity of social pressure
Social demand degree
DEMATEL method deep learning artificial intelligence identification weight ratio
Establishment of artificial intelligence solution for voice management of professional vocal performance in deep learning
Constructing hierarchical relationship diagram of deep learning by DEMATEL-ISM method

Figure 1: Chart of influential factors of voice health in professional vocal music performance.

Table 1: Impact index system of voice health care in professional vocal performance.

Target layer	Criterion layer	Feature layer	
The influence index system of vocal health care in professional vocal performance	Life element U1	Dietary habits	A1
		Work and rest habits	A2
		Exercise habit	A3
		Vocal habit	A4
	Body element U2	Physical fitness and health level	A5
		Health degree of vocal organ	A6
		Mental health degree	A7
		Degree of thinking and mental recognition	A8
	Training element U3	Scientific degree of phonation	A9
		Practice frequency	A10
		Daily performance frequency	A11
		Vocal singing level	A12
	Social element U4	How much to speak?	A13
		Daily workload	A14
		Intensity of social pressure	A15
		Social demand degree	A16

4.2. Body Element. Physical fitness and health level (A5): the dynamic supervision of the health status of physical fitness has a very important essential value for the analysis of the state of voice health care.

Health degree of vocal organs (A6): for the voice health care of professional vocal performers, the overall pile body of the vocal organs of singing has important value significance. It is the key influencing factor of the overall body elements, so it is an important key monitoring factor index.

Mental health degree (A7): the degree of mental health for professional vocal music performers' voice health care also has important value and role, and it has an important impact on the vocal music performers to deal with the state of singing voice and voice fusion problems; Feng [13] showed in his research that the degree of mental health for stimulating voice health care has a certain correlation and is a very important observation index.

Thinking and mental identification degree (A8): Professional vocal music performers need to have certain thinking and mental identification ability. They should have full imagination ability for image description. The degree determines the understanding intention of voice health care. They can judge through daily reflection ability during observation, which has important observation index value.

4.3. Training Elements. Scientific degree of vocalization (A9): an important external indicator in professional vocal performers is the degree of mastery of the scientific nature of vocalization, which is well reflected in the vibration frequency of vocalization and the overall effect of singing. A

Table 2: Comprehensive impact matrix of voice health care in professional vocal performance.

Factor	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
A1	0.07	0.09	0.12	0.12	0.15	0.15	0.09	0.10	0.11	0.10	0.08	0.08	0.07	0.11	0.12	0.06
A2	0.18	0.13	0.19	0.19	0.22	0.24	0.19	0.19	0.19	0.17	0.17	0.17	0.14	0.19	0.20	0.17
A3	0.18	0.17	0.11	0.16	0.20	0.20	0.17	0.16	0.19	0.16	0.16	0.14	0.12	0.16	0.18	0.10
A4	0.13	0.14	0.12	0.16	0.17	0.23	0.14	0.15	0.22	0.22	0.22	0.21	0.19	0.15	0.15	0.15
A5	0.25	0.25	0.23	0.26	0.21	0.31	0.25	0.26	0.29	0.28	0.28	0.25	0.24	0.27	0.27	0.20
A6	0.18	0.18	0.14	0.25	0.21	0.19	0.20	0.21	0.25	0.24	0.24	0.23	0.21	0.19	0.18	0.17
A7	0.16	0.18	0.17	0.14	0.19	0.18	0.11	0.15	0.15	0.14	0.14	0.13	0.12	0.15	0.19	0.11
A8	0.18	0.19	0.16	0.22	0.17	0.20	0.19	0.14	0.23	0.19	0.19	0.20	0.19	0.21	0.22	0.16
A9	0.17	0.17	0.17	0.26	0.21	0.27	0.19	0.23	0.19	0.25	0.25	0.24	0.17	0.24	0.24	0.21
A10	0.20	0.20	0.17	0.26	0.21	0.27	0.19	0.23	0.26	0.18	0.25	0.24	0.22	0.23	0.19	0.16
A11	0.23	0.25	0.20	0.29	0.26	0.31	0.24	0.26	0.29	0.28	0.21	0.27	0.24	0.27	0.27	0.23
A12	0.19	0.19	0.15	0.27	0.23	0.28	0.19	0.24	0.26	0.25	0.25	0.18	0.19	0.21	0.25	0.21
A13	0.13	0.13	0.12	0.22	0.17	0.23	0.15	0.15	0.22	0.21	0.21	0.20	0.12	0.15	0.15	0.11
A14	0.23	0.23	0.19	0.26	0.25	0.28	0.23	0.22	0.23	0.23	0.23	0.19	0.17	0.17	0.25	0.21
A15	0.20	0.22	0.19	0.25	0.24	0.27	0.22	0.21	0.22	0.21	0.23	0.20	0.17	0.22	0.17	0.20
A16	0.16	0.16	0.14	0.25	0.20	0.23	0.18	0.17	0.25	0.24	0.24	0.23	0.21	0.23	0.23	0.13

Table 3: Influence degree, influenced degree, cause degree, and centrality of each factor.

Factor		Degree of influence	Influenced degree	Centrality	Degree of cause	Factor attribute
Dietary habits	A1	1.63	2.86	4.48	-1.23	Consequence factor
Work and rest habits	A2	2.92	2.88	5.79	0.04	Causal factor
Exercise habit	A3	2.56	2.56	5.12	0.00	Causal factor
Vocal habit	A4	2.75	3.56	6.31	-0.81	Consequence factor
Physical fitness and health level	A5	4.07	3.28	7.35	0.79	Causal factor
Health degree of vocal organ	A6	3.28	3.83	7.11	-0.55	Consequence factor
Mental health degree	A7	2.40	2.95	5.34	-0.55	Consequence factor
Degree of thinking and mental recognition	A8	3.04	3.07	6.11	-0.03	Consequence factor
Scientific degree of phonation	A9	3.49	3.54	7.03	-0.05	Consequence factor
Practice frequency	A10	3.46	3.35	6.81	0.11	Causal factor
Daily performance frequency	A11	4.11	3.36	7.47	0.75	Causal factor
Vocal singing level	A12	3.53	3.14	6.68	0.39	Causal factor
How much to speak?	A13	2.66	2.78	5.44	-0.11	Consequence factor
Daily workload	A14	3.56	3.17	6.72	0.39	Causal factor
Intensity of social pressure	A15	3.44	3.27	6.71	0.17	Causal factor
Social demand degree	A16	3.27	2.57	5.84	0.70	Causal factor

comprehensive analysis of the voice will help to fully control the scientific nature of the overall vocalization.

Exercise frequency (A10): voice management has a certain value of monitoring the frequency of voice use. When the frequency of voice management is increasing, it will put forward better requirements for voice management. But the endurance of each person is not the same. Therefore, comprehensive big data analysis and evaluation are needed to obtain a better reasonable range.

Daily performance frequency (A11): for professional vocal performers, the pressure faced by the performance and the impact on the voice are very obvious, and the problems brought by different frequencies are very different. Professional performers often bear strong pressure when dealing with the impact of large frequency of performance. But there will also be a certain limit, so for this artificial intelligence, tracking and analysis through large data can obtain considerable value and significance.

Vocal singing level (A12): for professional vocal performers, the singing level will determine the level of voice use frequency, and this is a positive correlation; when the singing level is higher, it has an important impact on voice health value and significance, and when the singing level is lower, the pressure on voice health is often less. Therefore, it is one of the important observation indicators.

4.4. Social Element. The degree of speaking (A13): the dynamic supervision of speaking at ordinary times has important observation value and significance for voice health care of professional vocal music performance. Speaking is a very noisy thing for singers, so the degree of speaking is an important observation index.

Daily workload (A14): the amount of daily workload and transactional work has an important impact on the voice health of professional vocal performers. Therefore, the evolution and change of daily workload are two of the factors that have a direct impact on voice health.

Intensity of social pressure (A15): the intensity of social pressure has important influence on the psychology of professional vocal performers and will also affect the health care of the voice. The social pressure often has an important critical value, and the results are not consistent at different stages, so close attention is needed.

Social demand degree (A16): the demand for professional vocal performers is large or small, which is an important manifestation of establishing the social value of professional vocal performers, and will also affect the formation of the overall voice health awareness, so it is one of the important external factors.

5. Influence Model and Analysis of Vocal Health Care in Professional Vocal Performance

5.1. Introduction to Methods and Models. Through the use of DEMATEL method to determine the impact of various factors directly affecting the relationship between vocal performers, calculate the impact of various factors and the extent of the impact of the relationship, to obtain the center, the reasons, and ultimately the cause and effect factors. Based on the comprehensive influence matrix calculated by DEMATEL, the ISM interpretation structure model is used to establish a multilevel hierarchical interpretation structure model. The integration of these two methods can reduce the influence of artificial factors in a traditional reachable matrix, construct statistical factors synthetically, and empower and scientifically analyze the related research.

5.2. Model Calculation Process. The DEMATEL research method was used to set up a comprehensive survey and interview, and a five-level evaluation system was established, with 0, 1, 2, 3, and 4 as the hierarchical relationship. 0 has no effect, 1 has a slight effect, 2 has a small effect, 3 has a moderate effect, and 4 has a great effect on professional vocal performers, professional vocal performance educators, and professional vocal performance learners. A total of 10 subjects were interviewed. The direct impact matrix X^d is obtained by scoring the relevance of each item.

Table 4: Weight ratio of comprehensive influence criterion layer of voice health care in professional vocal music performance.

Factor	Life element U1	Body element U2	Training element U3	Social element U4
Centrality	21.71	25.92	27.98	24.71
Normalization	21.64%	25.83%	27.90%	24.63%

The normalized direct influence matrix X is obtained by normalizing the direct influence matrix X^d by X^d as in

$$X^d = \begin{bmatrix} 0 & A_{1,2} & \cdots & A_{1,j} \\ A_{2,1} & 0 & \cdots & A_{2,j} \\ \vdots & \vdots & \vdots & \vdots \\ A_{i,1} & A_{i,2} & \cdots & A_{i,j} \end{bmatrix}. \tag{1}$$

Among $1 \le i \le n, 1 \le j \le n$, n is the total number of influencing elements.

A comprehensive influence matrix T is calculated. According to the formula, as shown in formula (2), the obtained comprehensive influence matrix is shown in Table 2.

$$= X(1 - X)^{-1} = (t_{ij}). \tag{2}$$

And calculate the influence degree (R), the influenced degree (D), the centrality ($R + D$), and the cause degree ($R - D$) in the comprehensive influence matrix. Centrality ($R + D$) is a direct manifestation of the influence of a factor. As shown in Table 3, cause degree > 0 indicates that the element has great influence on other elements, which is called cause element. The result element means that the cause degree < 0 indicates that the element is greatly influenced by other elements, which is called the result element, as shown in

$$R = \sum_{j=1}^{n} T_{ij}(1 \le i \le n, 1 \le j \le n), \tag{3}$$

$$D = \sum_{j=1}^{n} T_{ij}(1 \le i \le n, 1 \le j \le n). \tag{4}$$

Establishing the comprehensive influence weight of voice health care. Because the centrality is used to calculate the comprehensive impact value of each criterion layer element on the whole, the comprehensive proportion is shown by the numerical value. "Therefore, the weight of the first-

Table 5: Grading table of influencing factors of voice health care in professional vocal performance.

Ladder level	Element set	Level description
L1	Diet habit A1; work and rest habit A2; exercise habit A3; voice habit A4; health degree of vocal organs A6; mental health degree A7; thinking and mental recognition degree A8; speaking degree A13; social pressure intensity A15; social demand degree A16	Surface influencing factors
L2	Scientific degree of phonation A9; frequency of practice A10; level of vocal singing A12; daily workload A14	Middle-level influencing factors
L3	Physical fitness and health A5; daily performance frequency A11	Root cause influence factor

level evaluation index can be obtained by normalizing the centrality of each index," and the weight vector formula of the comprehensive influence matrix index is shown in

$$A = (a_n|\text{fund}, \tag{5}$$

$$\sum_{n=1}^{4} a_n = 1\boxtimes 0 \leq a_n \leq 1\boxtimes n = 1, \cdots, 4), \tag{6}$$

where a_n is the weight of the index U_n. The calculation results are shown in Table 4.

The calculation of the reachability matrix is based on the comprehensive influence matrix T, and the overall influence matrix H is obtained through the publication calculation, as shown in

$$H = I + T. \tag{7}$$

A reachability matrix K is calculated; the calculation publication of a given threshold value is shown as

$$K_{ij} = 1, \text{if } h_{ij} \geq \lambda \quad (i, j = 1, 2, \cdots, n), \tag{8}$$

$$K_{ij} = 0, \text{if } h_{ij} < \lambda \quad (i, j = 1, 2, \cdots, n). \tag{9}$$

The determination of the threshold λ will divide and stratify the structure of the influence reachability matrix. The specific value is obtained according to the analysis of multiple values to obtain a satisfactory result in line with the basic level. After multiple values, $\lambda = 0$ is finally selected. A multilevel hierarchical ISM model is established, and each level is divided according to the K matrix and the corresponding conditions (public notice 9).

$$R_i \cap S_i = R_i \quad (i = 1, 2, \cdots, n). \tag{10}$$

R_i is a reachable set and S_i is the set of preceding items. Construct the grading table of comprehensive influencing factors of voice health care in professional vocal music performance, as shown in Table 5.

5.3. Result Analysis of Influencing Factors of Voice Health Care in Professional Vocal Performance. Through the calculation of the above model method, it can be seen that various factors have great influence on voice health care in professional vocal music performance. No factor is independent, it is through mutual influence to finally affect the sound, only long-term attention to these issues affects the voice, to form a systematic voice care thinking and voice health effects. Some of them evolve gradually because of various temporary environments and endogenous changes. Through the relevant research and questionnaire survey, the relevant analysis model is established. Among various factors, professional vocal music performance training factors and body factors are the key factors affecting voice health care. Physical fitness, frequency of daily performance, and social needs are the three major causal factors, which are the key factors affecting voice health care. In order to apply AI technology to the construction of sound health care in professional vocal music performance, we need to consider the training elements and the establishment of body elements. At the same time, three important factors should be considered in the key indicators, namely, the physical quality of vocal performers, the frequency of daily participation, and the change of social needs. Through these three factors, a systematic comprehensive algorithm system is constructed to create a professional vocal performance voice health artificial intelligence technology. The scale of influencing factors constructed by ISM tells researchers that physical fitness and daily performance frequency are the fundamental influencing factors. Scientific vocal method, practice frequency, vocal level, and the reasonable arrangement of daily work are the median factors; other factors belong to the surface factors. In the vocal health care of professional vocal music performance, the physical quality and daily performance frequency of the performers are the key factors affecting their voice health. Therefore, in artificial intelligence settings, this one can be used as a key factor to judge whether the health of voice care is fully considered and judged, when one of the two settings is higher. At the same time as feedback, remind the performers concerned to make corresponding adjustments to maintain a healthy voice.

Compared with other disease management, vocal management in professional vocal music performance is more affected by many complicated factors, such as various living habits, using voice habits, physical quality, and frequency of use. Therefore, when constructing the assessment system of artificial intelligence, we need to think systematically, holistically, and integrally with various factors, and we cannot ignore the important value and significance of various factors to voice management. According to the expert guidance of this method, what is constructed at present is a way in accordance with a certain color of human evolution. With the deepening of research and the participation in the evolution of scholars and thinking, the value and significance are relatively recognized and revised. This research needs to be further optimized and promoted in the future.

6. Conclusion

Artificial intelligence technology in intelligent medicine is of great value to human health in the future. Artificial intelligence algorithms are used to scientifically plan and manage people's daily lives, studies, and work. Through the artificial intelligence technology, it carries on the order management to everybody's health and proposes the scientific guidance. Avoiding further deterioration of the singers and the environment is a promising technical direction. Combined with the professional vocal actor's voice health management environment, it can effectively distinguish the relationship between singers and environment from both internal and external aspects. AI settings can avoid deviations. This paper fully respects the relevant laws of professional development, from the professional vocal performance of the internal and external elements constituting a large number of evolution factors. Finally, the weight relation of the system is formed by identifying these subtle factors, and the root factors, intermediate factors, and surface factors are calculated by the ISM structural equation model. An artificial intelligence algorithm system is composed of different weights, which scientifically and reasonably combines professional theory and practice. We hope that the future of intelligent health care in vocal performance will provide a new perspective for vocal health management and vocal enthusiasts and professionals around the world will provide a scientific voice health management program.

Data Availability

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

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This work was financially supported by the Chongqing Natural Science Foundation (Postdoctoral Fund) Project: Development and Research of Interactive Accurate Evaluation Platform for Professional Solfeggio Training Based on Reinforcement Learning (No. cstc2021jcyj-bsh0202, host: Qin Wei).

References

[1] Y. Chen, *Basic Theory of Vocal Music*, vol. 6, Jiuzhou Publishing House, Beijing, 2018.

[2] D. R. Gunjawate, V. U. Aithal, U. Devadas, and V. Guddattu, "Evaluation of singing vocal health in Yakshagana singers," *Journal of Voice*, vol. 31, no. 2, pp. 253.e13–253.e16, 2017.

[3] U. Devadas, P. C. Kumar, and S. Maruthy, "Prevalence of and risk factors for self-reported voice problems among Carnatic singers," *Journal of Voice*, vol. 34, no. 2, pp. 303.e1–303.e15, 2020.

[4] B. J. Baird, T. E. Mokhtari, C. K. Sung, and E. Erickson-DiRenzo, "A preliminary study of vocal health among collegiate a cappella singers," *Journal of Voice*, vol. 34, no. 3, pp. 486.e1–486.e11, 2020.

[5] R. Arunachalam, P. Boominathan, and S. Mahalingam, "Clinical voice analysis of Carnatic singers," *Journal of Voice*, vol. 28, no. 1, pp. 128.e1–128.e9, 2014.

[6] E. S. Tepe, E. S. Deutsch, Q. Sampson, S. Lawless, J. S. Reilly, and R. T. Sataloff, "A pilot survey of vocal health in young singers," *Journal of Voice*, vol. 16, no. 2, pp. 244–250, 2002.

[7] Z. A. Flavia, T. Lloyd Adam, and G. Julia, "Associations of education and training with perceived inging voice function among professional singers," *Journal of Voice*, vol. 35, no. 3, pp. 500.e17–500.e24, 2021.

[8] M. G. García and J. M. López, "Voice habits and behaviors: voice care among flamenco singers," *Journal of Voice*, vol. 31, no. 2, pp. 246.e11–246.e19, 2017.

[9] B. Irene and P. H. Wilson, "Working 9-5: causal relationships between singers' day jobs and their performance work, with implications for vocal health," *Journal of Voice*, vol. 31, no. 2, pp. 243.e27–243.e34, 2017.

[10] L. Yunfei, "Voice health care in popular singing," *Voice of the Yellow River*, vol. 19, pp. 112-113, 2016.

[11] X. Kun and Z. Xiaojun, "A study on the present situation of teachers' artistic voice application," *Grand Stage*, vol. 12, pp. 190-191, 2012.

[12] Y. Huimeng, *The Importance of Prevention and Treatment of Voice Diseases in Vocal Music Training*, Liaoning Normal University, 2016.

[13] C. Feng, "On voice protection in vocal music singing and teaching," *Northern Music*, vol. 9, pp. 78–80, 2020.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2595143, 15 pages
https://doi.org/10.1155/2022/2595143

Research Article

A Task Assignment Method Based on User-Union Clustering and Individual Preferences in Mobile Crowdsensing

Zihao Shao, Huiqiang Wang, Yifan Zou, Zihan Gao, and Hongwu Lv

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Huiqiang Wang; wanghuiqiang@hrbeu.edu.cn

Received 12 January 2022; Revised 19 March 2022; Accepted 15 April 2022; Published 9 May 2022

Academic Editor: Barbara Guidi

Mobile crowdsensing (MCS) offers a novel paradigm for large-scale sensing with the proliferation of smartphones. Task assignment is a critical problem in mobile crowdsensing (MCS), where service providers attempt to recruit a group of brilliant users to complete the sensing task at a limited cost. However, selecting an appropriate set of users with high quality and low cost is challenging. Existing works of task assignment ignore the data redundancy of large-scale users and the individual preference of service providers, resulting in a significant workload on the sensing platform and inaccurate assignment results. To tackle this issue, we propose a task assignment method based on user-union clustering and individual preferences, which considers the influence of clustering data quality and preference-based sensing cost. Firstly, we design a user-union clustering algorithm (UCA) by defining user similarity and setting user scale, which aims to balance user distribution, reduce data redundancy, and improve the accuracy of high-quality user aggregation. Then, we consider individual preferences of service providers and construct a preference-based task assignment algorithm (PTA) to achieve the diversified sensing cost control needs. To evaluate the performance of the proposed solutions, extensive simulations are conducted. The results demonstrate that our proposed solutions outperform the baseline algorithm, which realizes the individual preference-based task assignment under the premise of ensuring high-quality data.

1. Introduction

The pervasive adoption of mobile smart devices and the rapid development of communication network technologies have accelerated the unprecedented expansion of mobile crowdsensing (MCS) in many aspects of our daily lives. MCS [1] is a compelling paradigm that allows a large group of individuals to collaboratively sense data and extract information about social events and national phenomena with common interest using mobile devices (e.g., smartphones, smart glasses).

Task assignment is a critical problem in MCS, where service providers attempt to recruit a group of brilliant users to complete the sensing task at a limited cost. Thus, the core goal of task assignment is to make a good balance between data quality and sensing cost [2]. Specifically, suppose the sensing platform always assigns inappropriate tasks to users and keeps users away from daily activities. In that case, users will refuse to perform tasks, leading to revenue loss for service providers and reducing the sensing utility. In addition, service providers may have individual preferences (i.e., maximizing benefits, minimizing costs) when selecting user data, further increasing the complexity and diversity of the task assignment.

By now, various methods to improve data quality have been proposed for task assignment. While service providers can enjoy the convenience provided by user data, large-scale user data will lead to the increase of redundant data.

Data redundancy is a potential threat to data quality, which will increase the workload of the platform and reduce the accuracy of the task assignment. Shahraki et al. [3] pointed out that cluster analysis can solve the problem of data redundancy. The most common data clustering mechanism is the k-means algorithm, despite the effectiveness, applying the k -means algorithm in MCS to realize task assignments still need to tackle complexity and balance challenges. Firstly, the k-means algorithm may require a high calculation time. Unfortunately, most task assignment methods are time-sensitive. That is, the acquisition of high-quality data should not be at the expense of time. Secondly, these algorithms are still a risk of poor data balance due to work neglecting the difference in the scale of clustered users. Thus, leveraging clustering analysis to form balanced, low-redundancy, and high-quality data in sensing regions with a limited time is a crucial problem in the MCS task assignment.

Sensing cost is another important issue to consider in task assignment, which aims to select suitable users to achieve task assignment. Up to now, the research works on sensing cost mainly focused user recruitment cost [4–6], user travel cost [7–9], and data transferring cost [10–13]. Most of the existing works for sensing cost consider a homogeneous preference model, which assumes all service providers have the same preference. Each service provider selects user data independently and randomly according to the same task preference. However, this model is at best an approximation, because different service providers indeed have various tastes and preferences. Such heterogeneity in preferences of service providers has been observed in [14].

The shortcomings of existing works drive us to explore a new task assignment method from data quality and sensing cost for realistic MCS applications. Our research efforts aim to achieve a practical task assignment in different individual preferences for real MCS with varying user data quality, while ensuring high-quality clustering data and preference-based sensing cost. More specifically, we first formulate the problem of task assignment. This formulation carefully considers the quality of clustering data and the individual preference for sensing cost. Afterward, a task assignment method is proposed based on user-union clustering and individual preferences. Different from prior works on task assignment, we first considered data redundancy caused by large-scale user data, leveraged the clustering method to reduce data redundancy and improve the accuracy of high-quality user aggregation. Then, based on this solution, we analyzed the impact of individual preferences and solved the diversified task assignment under the individual preference sensing cost.

In summary, this paper makes the following contributions:

(i) We formulate the task assignment problem from two perspectives. High-quality clustering data and the individual preference sensing cost are considered in our formulation

(ii) A UCA-based solution is proposed to balance user data scale, reduce data redundancy, and ultimately improve platform efficiency and data quality

(iii) A PTA-based solution is proposed to solve the task assignment under the individual preference sensing cost. To the best of our knowledge, this is the first work that validates from different perspectives of the task assignment the benefits of exploiting individual preferences and that gains insights through simulations based on real-world data

The rest of the paper is organized as follows: Section 2 discusses related work. Section 3 introduces our system model and problem formulation. Our UCA and PTA solutions are presented in Section 4. In Section 5, we evaluate our proposed method and present evaluation results. Finally, we conclude this paper in Section 6.

2. Related Work

Data quality and sensing cost become the main criterion to assign the tasks. Much work has been done to support the efficient task assignment in MCS. In the following, we shall introduce existing work in these two criteria.

2.1. Data Quality. Improving the accuracy of data quality is an essential design objective for most task assignments. Several factors have a significant impact on data quality, including data collection times, task duration, and data spatial-temporal coverage.

Data collection times refer to the number of times a target phenomenon is expected to be sensed. On the one hand, multiple measurements can reduce sensor reading errors and make sensing results approach the ground truth. Gong et al. [15] pointed out data quality keeps increasing as the collection times increase, characterized by a non-decreasing sub-modular function. On the other hand, there are tiny fluctuations of sensing data even in short durations and small areas. Xiong et al. [16] proposed that data quality will no longer increase when the collected data exceeds a certain threshold. However, multiple measurements are necessary to improve data quality in most cases.

Task duration is the period from the instant a task is published to the deadline. Wang et al. [17] proposed a two-level heterogeneous pricing mechanism based on the timeliness and location dependence of random arrival in MCS. The proposed greedy task selection algorithm can help users choose the appropriate task to maximize the total revenue and realize task assignment. Zeng et al. [18] took the execution time of workers as the optimization goal, and proposed an adaptive Top-k worker selection algorithm to select the most appropriate workers and achieve efficient task assignment. Huang et al. [19] investigated and formulated the time-dependent task allocation problem, and characterized the cost of performing a sensing task for each mobile user. They proposed an efficient task assignment algorithm called the optimized allocation scheme of time-dependent tasks (OPAT), which can maximize the sensing capacity of each mobile user.

Data spatial-temporal coverage is another important metric to evaluate data quality and has been extensively studied. To evaluate the time coverage provided by a group of users over a period of time, Alagha et al. [20] considered users' location and mobility mode. They designed a stable coverage recruitment parameter to realize task assignment. To reduce the system cost, Song et al. [21] migrated certain qualified users to less popular tasks to increase data coverage and optimize other performance factors. To satisfy both the service provider's coverage sensing preference and the user's revenue preference, Yucel et al. [22] proposed a coverage-aware stable task assignment method and proved that the user's revenue is proportional to the task coverage scale. Experimental results show that this method achieved accurate task assignment on the premise of ensuring user satisfaction and coverage quality.

The above works are good references for addressing the data quality of task assignment. However, most of these studies have not considered the importance of clustering analysis. Guo et al. [23] analyzed some common problems in task assignment and pointed out that cluster-based task assignment is necessary for future MCS task assignment. In the research of user data clustering, Du et al. [24] combined the data quality of users and proposed a Bayesian co-clustering truth discovery model to capture the fine-grained reliability of users on different clusters. This model enhances the usability of each user under the most appropriate task, which is conducive to observing aggregated tasks. Jin et al. [25] proposed a novel MCS system framework that integrates an incentive, a data aggregation, and a data perturbation mechanism. The data aggregation mechanism incorporated workers' reliability to generate highly accurate aggregated results. So far, the research of user clustering data is still in its infancy. It is crucial to consider clustering data quality evaluation to reduce redundancy and improve platform efficiency.

2.2. Sensing Cost. Sensing cost is the costs paid to perform tasks, including user recruitment cost, user travel cost, and data transferring cost. The first is paid by the sensing platform to recruited users for their involvement; the latter two are paid by users for their movements for data collection and data upload, respectively.

User recruitment cost includes per-user recruitment cost and per-data collection cost. To control the recruitment cost of users, Liu et al. [4] studied the user recruitment problem on both the user's and subarea's sides and proposed a three-step strategy, including user selection, subarea selection, and user-subarea-cross (US-cross) selection. Extensive experiments on two real-world data sets show that user recruitment algorithms can effectively enhance the data inference accuracy under a budget constraint. In practical application, Campioni et al. [5] improved recruitment algorithms for vehicular crowdsensing networks, which aims to select participants within a crowdsensing network such that the most sensing data is obtained for the lowest possible cost. Zhao et al. [6] classified the extrinsic utility into the task payoff shared with other participants and the resource cost incurred by participation. Based on this, they proposed a social-aware incentive mechanism by deep reinforcement learning (DRL-SIM) to control user recruitment cost and derive the optimal long-term sensing strategy for all vehicles.

User travel cost relies on the traveling paths of users, which could be fixed, predetermined, or predictable based on users' historical trajectories [7]. In fixed/predetermined-path-based MCS, each user can perform tasks alone or near their traveling path. In this case, the task assignment problem can often be transformed into a set cover problem or bipartite graph matching problem. Wei et al. [8] considered user moving cost and sensing level. They proposed a greedy task assignment algorithm, GP-BS, to select the most cost-effective participant iteratively. In predictable-path-based MCS, the traveling path of each worker is not predetermined. It is tough to accurately predict the specific locations of users in the future at a fine granularity. Wang et al. [9] proposed an approach that exploits the spatial-temporal causality among travel speeds of road sections by a time-lagged correlation coefficient function, which aims to overcome the uneven spatial-temporal distribution of vehicles and the variation of their data-offering intervals. For the sparse MCS scene, Wang et al. [10] propose a deep learning-enabled industrial sensing and prediction scheme, aiming to achieve high-precision prediction of future moments under the hypothesis of sparse historical data.

Data transferring cost is the cost generated for uploading sensing data. Wang et al. [11] considered that the users' main concern is the cost of data uploading, which affects their willingness to participate in a crowdsensing task. The proposed efficient prediction-based user recruitment for MCS can achieve a lower recruitment payment and the highest delivery efficiency. In [12], a data transfer solution for crowdsensing was proposed to minimize the number of users under the constraints of the quality of sensing data and coverage area of all cell towers. When multiple tasks share a pool of staff with bandwidth constraints, a multitask allocation strategy is proposed in [13] to ensure platform revenue.

Task assignment algorithms for MCS were designed following the different sensing costs. However, the algorithms proposed in the existing works are usually designed based on a fixed choice. That is, they all neglect the individual preferences for sensing cost. In our previous work [26], we have pointed out the influence of individual preferences on selection. Therefore, it is necessary to consider the individual preferences to ensure the practicality of task assignment.

In summary, despite the variety of the literature on data quality and sensing cost in MCS task assignments, the goal is defined chiefly from the overall system's point of view without considering the individual preferences and the importance of clustering data. Hence, they may not necessarily achieve high accuracy and rationality in the task assignment.

3. System Model and Problem Formulation

In this section, we first give the system model for task assignment in MCS. Then, we formulate the task assignment problem.

3.1. System Model

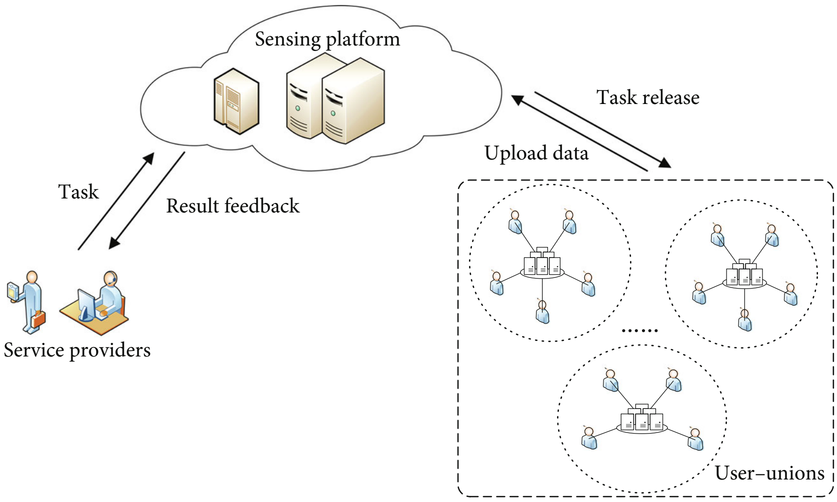


Figure 1: Framework for the mobile crowdsensing.

3.1.1. Model Construction. We consider a typical MCS architecture, including a trusted sensing platform, a set of m sensing users, and a set of k service providers, as shown in Figure 1.

For the task assignment, service providers can publish different sensing tasks and task centers to the sensing platform, denoted by $T = \{t_1, t_2, \cdots, t_k\}$, $t_{j\text{–center}}$, respectively. The sensing platform assigns tasks to sensing users, denoted by $U = \{u_1, u_2, \cdots, u_m\}$. To reduce data redundancy and the burden on the sensing platform, users form the user-union before uploading data. Service providers select appropriate users to realize task assignment according to individual preferences.

In this paper, we make the following assumptions.

(i) The initial locations of users are uniformly distributed in a specific region

(ii) The sensing platform is only responsible for the data calculation between users and service providers

(iii) Service providers have different individual preferences and decide the final choice. Service providers can only select one sensing user to achieve the task assignment

Such assumptions are practical in enterprise or agreement-based cooperation scenarios [27].

3.2. Problem Formulation

3.2.1. Data Quality Problem

(1) Data Quality Problem Formulation. In data quality research, considering a large number of sensing users, each user uploading data in an independent way will lead to a decrease in sensing utility. Therefore, we leverage the clustering method to reduce data redundancy and improve the accuracy of high-quality user aggregation. We evaluate the data quality and transform the user clustering problem into the maximum similarity matching problem, which can be expressed as follows:

$$\text{maximize sim} = \sum_{U,T} \text{sim}_{u_i,t_{j\text{–center}}} = \sum_{U,T} \sum_{u_i \in U, t_{j\text{–center}} \in T} f\left(u_i, t_{j\text{–center}}\right) \tag{1}$$

The goal of Equation (1) is to form a union with the highest user similarity from large-scale participating users, so as to reduce data redundancy and improve the efficiency of the sensing platform. $f(u_i, t_{j\text{–center}})$ represents the similarity between u_i and $t_{j\text{–center}}$, which can be expressed as follows:

$$\text{sim}_{u_i,t_{j\text{–center}}} = f\left(u_i, t_{j\text{–center}}\right) = \frac{1}{1 + \sqrt{\sum_{a=1}^{n} \omega_a * \left(c_{a,u_m} - c_{a,t_{j\text{–center}}} / c_{a,t_{j\text{–center}}}\right)^2}}, \tag{2}$$

where c_{a,u_m} represents the value of u_m under evaluation index a and $c_{a,t_{j\text{–center}}}$ represents the value of $t_{j\text{–center}}$ under evaluation index a.

(2) Method Construction. Step 1: Define the user similarity as a two-tuple.

Definition 1. Define the user similarity as $\text{sim}_{u_i,t_{j\text{–center}}} = f(u_i, t_{j\text{–center}})$.

$f(u_i, t_{j\text{–center}})$ is the participants of both clustering data, where u_i denotes the user, and $t_{j\text{–center}}$ denotes the center of task t_j.

Step 2: Calculate the similarity between u_i and $t_{j\text{-center}}$, sort the calculation results, and construct the user-union clustering.

Step 3: Set a maximum user limit τ in each user-union to ensure the balance of the union, i.e., $\|\mathrm{sim}_{u_i, t_{j\text{-center}}}\| \le \tau$.

3.2.2. Sensing Cost Problem

(1) Sensing Cost Problem Formulation. In Section 3.2.1, we use user data to build user-unions, which realize user clustering, reduce data scale and ensure data quality. Based on this solution, we consider the diversity and individual preferences for service providers, and solve the diversified task assignment under the individual preference sensing cost.

For each task assignment problem, each user-union has n sets of user schemes and m sets of data sensing cost evaluation indexes, denoted by $Y = \{Y_1, Y_2, \cdots, Y_n\}$ and $G = \{G_1, G_2, \cdots, G_m\}$. Each user scheme represents a sensing cost requirement, which can be evaluated by the sensing cost indexes, denoted by $\{h_{11}, h_{12}, \cdots, h_{1m}\}$. According to the decision selection sample matrix, service providers select the appropriate user to realize task assignment. The decision selection sample matrix is expressed as follows:

$$H = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nm} \end{bmatrix}, \tag{3}$$

Based on the above conditions, we normalize the decision information matrix, and use the prospect theory [28] to obtain the positive and negative prospect value matrix. Finally, the acceptability advantage solution is used to sort the schemes and select the most suitable users. Therefore, we transform the preference-based sensing cost problem into the maximum comprehensive prospect value, which can be expressed as follows:

$$\begin{array}{ll} \max V & \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}^{+}\psi^{+}(\omega_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} v_{ij}^{-}\psi^{-}(\omega_j) \\ s.t. & \omega_j \in [0,1] \\ & \sum_{j=1}^{m} \omega_j = 1 \end{array} . \tag{4}$$

The goal of Equation (4) is to solve the maximum comprehensive prospect value, so as to achieve the preference-based task assignment. The objective function in the first line is to solve the optimal evaluation index weight. The second and third lines define the range of each index, respectively.

(2) Method Construction. Step 1: Normalize the decision matrix of user scheme. We define the user sensing costs as the cost index and the benefit index, denoted by h_{ij}^{b} and h_{ij}^{c}, $h_{ij}^{b}, h_{ij}^{c} \in H$, $h_{ij}^{b} \cup h_{ij}^{c} = H$, and $h_{ij}^{b} \cap h_{ij}^{c} = \varnothing$, which can be expressed as:

$$z_j = \frac{1}{n}\sum_{i=1}^{n} h_{ij}, \tag{5}$$

$$h_{ij}^{b} = \frac{h_{ij} - z_j}{\max\left\{\max_j\left(h_{ij}\right) - z_j, z_j - \min_j\left(h_{ij}\right)\right\}}, \tag{6}$$

$$h_{ij}^{c} = \frac{z_j - h_{ij}}{\max\left\{\max_j\left(h_{ij}\right) - z_j, z_j - \min_j\left(h_{ij}\right)\right\}}. \tag{7}$$

Step 2: Determine the positive and the negative prospect value matrix. The normalized decision matrix is recorded as $\bar{O} = \left(\overline{h_{ij}}\right)_{n\times m}$. We construct the positive and the negative prospect value matrix, which can be expressed as:

$$\begin{cases} h_j^{+} = \max\left\{h_{ij} \mid 1 \le i \le n\right\} \\ h_j^{-} = \min\left\{h_{ij} \mid 1 \le i \le n\right\} , \\ j = 1, 2, \cdots, m \end{cases} \tag{8}$$

where $Y^{+} = \{h_1^{+}, h_2^{+}, \cdots, h_m^{+}\}$ and $Y^{-} = \{h_1^{-}, h_2^{-}, \cdots, h_m^{-}\}$ represent the positive and the negative ideal scheme, respectively.

Step 3: Calculate the correlation coefficient. A proper task assignment usually needs a reference node to measure the prospect value of the scheme, rather than the actual value of the decision result. Therefore, we use the values of positive and negative ideal schemes as reference points, which can be expressed as:

$$\begin{cases} \varsigma_{ij}^{+} = \dfrac{\min_{i,j}\left|h_{ij} - h_j^{+}\right| + \varphi \max_{i,j}\left|h_{ij} - h_j^{+}\right|}{\left|h_{ij} - h_j^{+}\right| + \varphi \max_{i,j}\left|h_{ij} - h_j^{+}\right|} \\ \varsigma_{ij}^{-} = \dfrac{\min_{i,j}\left|h_{ij} - h_j^{-}\right| + \varphi \max_{i,j}\left|h_{ij} - h_j^{-}\right|}{\left|h_{ij} - h_j^{-}\right| + \varphi \max_{i,j}\left|h_{ij} - h_j^{+}\right|} \end{cases}, \tag{9}$$

where $\varsigma_{ij}^{+}, \varsigma_{ij}^{-}$ represent the positive and the negative correlation coefficients, respectively, $0 \le \varsigma_{ij}^{+} \le 1$ ⊠ $-1 \le \varsigma_{ij}^{-} < 0$, φ represents the resolution coefficient, define$\varphi = 0.5$.

Step 4: Construct prospect decision matrix. We construct a prospect value function to represent the subjective feelings of service providers about the user scheme selection, which can be expressed as:

$$v(h_i) = \begin{cases} \left(1 - \varsigma_{ij}^{-}\right)^{\alpha}, & \varsigma_{ij}^{-} \text{ is a reference point} \\ -\lambda\left[-\left(\varsigma_{ij}^{+} - 1\right)\right]^{\beta}, & \varsigma_{ij}^{+} \text{ is a reference point} \end{cases}, \tag{10}$$

where α and β represent the concave and the convex degree of the benefit and the cost value functions at the reference point, respectively, $0 < \alpha < 1$, $0 < \beta < 1$. λ represents the degree of loss aversion of the service provider.

According to Equation (10), we achieve the positive and the negative values of Y_i, which expressed as:

$$\begin{cases} v^{+}\left(h_{ij}\right) = \left(1 - \varsigma_{ij}^{-}\right)^{1} \\ v^{-}\left(h_{ij}\right) = -\lambda\left[-\left(\varsigma_{ij}^{+} - 1\right)\right]^{\beta} \end{cases}. \tag{11}$$

Probability weight is the subjective judgment made by the service provider according to the probability ω of the result of the task assignment, which can be expressed as:

$$\begin{cases} \psi(\omega)^{+} = \dfrac{\omega^{\eta}}{\left[\omega^{\eta} + (1 - \omega^{\eta}\right]^{1/\eta}}, \varsigma_{ij}^{-} \text{ is a reference point} \\ \psi(\omega)^{-} = \dfrac{\omega^{\gamma}}{\left[\omega^{\gamma} + (1 - \omega)^{\gamma}\right]^{1/\gamma}}, \varsigma_{ij}^{+} \text{ is a reference point} \end{cases}, \tag{12}$$

where $\alpha = \beta = 0.88$, $\lambda = 2.25$, $\eta = 0.61$, $\gamma = 0.69$ [28], η and γ represent the fitting parameters of the probability weight function on the left and right of the reference point, respectively.

We calculate the comprehensive prospect value of each user scheme, which can be expressed as:

$$V_i = \sum_{j=1}^{m} v_{ij}^{+}\psi^{+}\left(\omega_j\right) + \sum_{j=1}^{m} v_{ij}^{-}\psi^{-}\left(\omega_j\right). \tag{13}$$

Step 5: Weight optimization. The weight of the user scheme should be reasonably assigned, aiming to obtain the maximum comprehensive prospect value, which can be expressed as:

$$V_i^{*} = \sum_{j=1}^{m} v_{ij}^{+}\psi^{+}\left(\omega_j^{*}\right) + \sum_{j=1}^{m} v_{ij}^{-}\psi^{-}\left(\omega_j^{*}\right). \tag{14}$$

The multi-attribute hesitant fuzzy evaluation matrix is transformed into the multi-attribute comprehensive prospect matrix.

Step 6: Sort user schemes to determine the preference-based task assignment.

According to the comprehensive prospect matrix, we calculate the positive (i.e., f^{+}) and the negative (i.e.,f^{-}) ideal solutions of each index, which can be expressed as:

$$\begin{cases} f^{+} = \left\{\max_i v(h_{i1}), \max_i v(h_{i2}), \cdots, \max_i v(h_{im})\right\} \\ f^{-} = \left\{\min_i v(h_{i1}), \min_i v(h_{i2}), \cdots, \min_i v(h_{im})\right\} \end{cases}. \tag{15}$$

We also need to calculate the group benefit value (i.e., B_i), individual regret value (i.e., R_i), and comprehensive index value (i.e., BR_i).

$$\begin{cases} B_i = \sum_{j=1}^{m} \omega_j\left(f^{+} - v\left(h_{ij}\right)\right)/\left(f^{+} - f^{-}\right) \\ R_i = \max_j\left\{\omega_j\left(f^{+} - v\left(h_{ij}\right)\right)/\left(f^{+} - f^{-}\right)\right\} \end{cases}, \tag{16}$$

$$BR_i = \kappa\frac{B_i - B_i^{-}}{B_i^{+} - B_i^{-}} + (1 - \kappa)\frac{R_i - R_i^{-}}{R_i^{+} - R_i^{-}}, \tag{17}$$

where B_i^{+}, B_i^{-} represent the maximum and minimum group benefit value, R_i^{+}, R_i^{-} represent the maximum and minimum individual regret value, and κ represents the decision preference. When $\kappa > 0.5$, it means that the service provider adopts the maximum group benefit to formulate the task assignment scheme. When $\kappa < 0.5$, it means that the service provider adopts the minimum individual regret to formulate the task assignment scheme. When $\kappa = 0.5$, it means that the service provider adopts the balance principle to formulate the task assignment scheme.

According to the judgment criteria of the VIKOR method [29], the value of B_i, R_i, and BR_i are arranged in descending order. We use BR_i to determine the first (i.e., Y_1) and second (i.e., Y_2) user schemes and realize the preference-based task assignment.

Condition 1 (Acceptability advantage). $BR(Y_2) - BR(Y_1) \geq 1/(n-1)$, where m is the number of options.

Condition 2 (Acceptability stable). Y_1 has the best B_i or R_i.

When Condition 1 and Condition 2 are both satisfied, Y_1 is the optimal user scheme to realize task assignment. When only Condition 1 is satisfied, Y_1 and Y_2 are compromise solutions. When only Condition 2 is satisfied, Y_1, Y_2, $\cdots$, Y_N are approximate ideal schemes.

4. Proposed Task Assignment Solutions

4.1. User-Union Clustering Algorithm. Traditional clustering algorithms are deficient in the efficiency and balance of clustering results. To solve this issue, we propose the user-union clustering algorithm (UCA), as shown in Algorithm 1.

Algorithm 1 realizes the generation of user-union. UCA provides a guarantee for the balance of clustering data by setting an upper limit. The function of ProperCluster (x_i) is to assign u_i to a suitable user-union. CS_j is a two-tuple, which represents the storage of existing user data and the similarity value of the center task in the jth user-union. From 1 to 4, the algorithm is used to calculate the similarity between u_i and $t_{j\text{-center}}$, which aims to quantify the behavioral characteristics of each user. From 5 to 18, the algorithm is used to control the scale of users, which can balance the number of users in the user-union.

Computational complexity. The k-means algorithm is a simple and efficient clustering algorithm, and the computational complexity of the algorithm is $O_2(tkmn)$, where t

```
Input: User data $U = \{u_1, u_2, \cdots, u_m\}$, task set $T = \{t_1, t_2, \cdots, t_k\}$, maximum user limit $\tau$
Output: Set of $K$ task clusters $tc = \{tc_1, tc_2, \cdots, tc_k\}$
ProperCluster($x_i$)
Determine the center of the initial task sets and user data evaluation indexes ($C_{u_i} = (c_{1,u_i}, c_{2,u_i}, \cdots, c_{n,u_i})$
Calculate the user similarity by Equation (1), and sort data in descending order $\mathrm{STC} = \{\mathrm{sim}_{u_i,t_{j\text{-center}}} \mid j = 1, 2, \cdots, K\}$
    for $j \longleftarrow 1$ to $K$ DO
        if $\|\mathrm{sim}_{u_i,t_{j\text{-center}}}\| < \tau$
            $u_i$ enter $tc_j$
            $u_i$ and $sim_{u_i,t_{j\text{-}center}}$ are saved in $CS_j = \{\{u_i, \mathrm{sim}_{u_i,t_{j\text{-center}}}\} \cdots\}$
            break
        else
            if $\mathrm{sim}_{u_i,t_{j\text{-center}}}$ is less than the minimum similarity value in $CS_j$
                continue
            else
                $u_i$ joins the $j$th union and deletes edge user ($u_e$)
                ProperCluster($u_e$)
repeat
    for $i \longleftarrow 1$ to $N$ DO
        ProperCluster($x_i$)
until saturate task requirements or reach the maximum number of iterations
End
```

Algorithm 1: User-union clustering algorithm (UCA).

represents the number of iterations, m represents the user scale, n represents the type of user data evaluation index, and k represents the number of clustering tasks. UCA is an improvement of the k-means algorithm, which uses user similarity to realize user clustering and improves the balance of user scale. First, each user needs to calculate the similarity with k, and the complexity is kmn. Next, the value of user similarity is compared with the edge point, when the number of users in the user-union reaches saturation, and the complexity is 1. In the worst case, UCA spends k times for comparison. Therefore, the computational complexity of UCA is $O_1(k^2mn)$. In practical application scenarios, to ensure the clustering accuracy, the number of algorithm iterations (i.e., t) is usually greater than clustering tasks (i.e., k); therefore, $O_2 - O_1 = (t - k)kmn > 0$.

Space complexity. The k-means algorithm needs to store user data and the clustering tasks data, and the space complexity of is $(k + m)n$. Like the k-means algorithm, UCA also needs to store user data and clustering task data, and the space complexity is $(k + m)n$.

4.2. Preference-Based Task Assignment Algorithm. Algorithm 1 provides high-quality data. Then, we propose the preference-based task assignment algorithm (PTA) to solve the diversified task assignment under the individual preference sensing cost, as shown in Algorithm 2.

Algorithm 2 realizes the reasonable and diverse task assignment by calculating the value of group benefit, individual regret, and a comprehensive index. This is a mode of task assignment selection from individual preference, which guides the decision of service providers. From 1 to 2, the algorithm is used to normalize the decision matrix. From 3 to 4, the algorithm is used to determine the positive and negative ideal solutions and calculates the correlation coefficient. From 5 to 6, the algorithm mainly constructs the prospect decision matrix through optimized weights. From 7 to 19, the algorithm is used to sort user schemes, and achieve preference-based task assignment based on the VIKOR method.

5. Performance Evaluation and Discussion

5.1. Basic Simulation Setup. In our experiments, the data we used came from the real Dartmouth College Wi-Fi campus trace data set [30], which was an experiment on the open-source middleware NSense. This data takes sound collection as an example, including timestamps, the distance between test points and sensing nodes, data collection methods, and data collection environments. We define data collection methods and environments as benefit indexes. Other metrics are defined as cost indexes. We consider two user distribution spaces [31] (i.e., sparse and dense regions) and employ different metrics to measure the performance in UCA and PTA.

In the research of task assignments, high-quality user data can improve the accuracy of assignments. User data clustering can reduce data redundancy and improve the overall quality of user data. Therefore, we first verify the performance of UCA. We compare the performance with three common clustering algorithms [14] (i.e., K-means, K-means improve, and fuzzy C-means clustering algorithm) by calculating the accuracy (ACC), normalized mutual information (NMI), and running time. The K-means improve algorithm limits the number of users of the K-means, aiming to control the balance of user distribution scale.

ACC is used to measure the accuracy of the users' classification after clustering, and compared to the actual

```
Input: Decision sample matrix $H$
Output: Optimal task assignment
Initialization
Normalize the sample matrix by Equations (5)–(7)
Determine $Y^+ = \{h_1^+, h_2^+, \cdots, h_m^+\}$ and $Y^- = \{h_1^-, h_2^-, \cdots, h_m^-\}$ by Equation (8)
Calculate the correlation coefficient by Equation (9)
Build a prospective decision matrix and calculate the prospective value
Optimize the index weights to obtain the best comprehensive prospect value by Equations (10)–(14)
Calculate $B_i$, $R_i$, and $BR_i$ by Equations (15)–(17), confirm the first and second value of $BR_i$ (i.e., $Y_1$ and $Y_2$)
for $Y_1$ and $Y_2$ do
    if only meet Condition 1 then
        $Y_1$ and $Y_2$ are compromise solutions
    end if
    else if only meet Condition 2 then
        Calculate the largest $N$ by $BR(Y_2) - BR(Y_1) < 1/(N-1)$, and $Y_1, Y_2, \cdots, Y_N$ are approximate ideal schemes
    end if
    if both meet Conditions 1 and 2 then
        $Y_1$ is the optimal solution
    end if
end for
```

Algorithm 2: Preference-based task assignment algorithm (PTA).

classification in the prior knowledge, which can be expressed as:

$$\mathrm{ACC} = \frac{\sum_{i=1}^{N} v(s_i, \mathrm{map}(r_i))}{N}, \tag{18}$$

where N is the number of users, map is a mapping function that maps the classification of the clustering results to the original data set, s_i is the original classification of user data in prior knowledge. When $s_i = \mathrm{map}(r_i)$, the value of v is 1. Otherwise, the value of v is 0.

NMI is used to evaluate the similarity between the clustering results and the distribution of the original dataset, which can be expressed as:

$$\mathrm{NMI}(X, Y) = \frac{I(X, Y)}{\sqrt{H(X)H(Y)}}, \tag{19}$$

where $I(X, Y)$ represents mutual information between X and Y. $H(X)$ and $H(Y)$ represent the information entropy of distributions X and Y, respectively.

Next, we verify the performance of PTA on the premise of obtaining high-quality user data, which aims to realize preference-based task assignment under the individual preference sensing cost. We compare the performance with two methods (i.e., VIKOR [29] and TOPSIS [32]) by calculating the compatibility degree and execution time. The VIKOR method determines the optimal task assignment scheme without the prospect value. The TOPSIS method is a common method to solve the ideal point.

Compatibility degree [33] is used to verify the rationality of task assignment, which can be expressed as:

$$\begin{cases} \mathrm{compd}_{\mathrm{met}i} = \dfrac{1}{l-1} \displaystyle\sum_{\mathrm{met}j=2}^{l} p_{\mathrm{met}i,\mathrm{met}j} \\ p_{\mathrm{met}i,\mathrm{met}j} = 1 - \dfrac{6}{m(m^2-1)} \displaystyle\sum_{d=1}^{m} f_d^2 \end{cases}, \tag{20}$$

where $\mathrm{compd}_{\mathrm{met}i}$ represents the compatibility of the ith method, $p_{\mathrm{met}i,\mathrm{met}j}$ represents the degree of correlation between i and j, m represents the number of schemes, and f_d represents the sorting difference of the dth scheme in i and j.

5.2. Experiment Results of UCA. It is meaningless to use UCA to reduce data redundancy for the small scale of users in remote regions. Therefore, for the experiment of UCA, we analyze the clustering effect of large-scale users, when the number of users varies from 100 to 1000, respectively. Figures 2–4 show the performance in terms of ACC, NMI, and running time, achieved by the four algorithms.

Apparently, UCA outperforms the three baselines (i.e., higher ACC, higher NMI, and lower running time), no matter how the number of users varies. In Figure 2, the augmentation of user data decreases the accuracy of all clustering algorithms. The reason is that the increase of user scale leads to the rise in low-similarity users, which reduces the clustering accuracy. The accuracy of UCA is better than these three algorithms, and the ACC is basically above 0.82. Compared with the best performance K-means algorithm, the accuracy is improved by about 10%. The reason is that UCA calculates user similarity and sets boundary user replacement

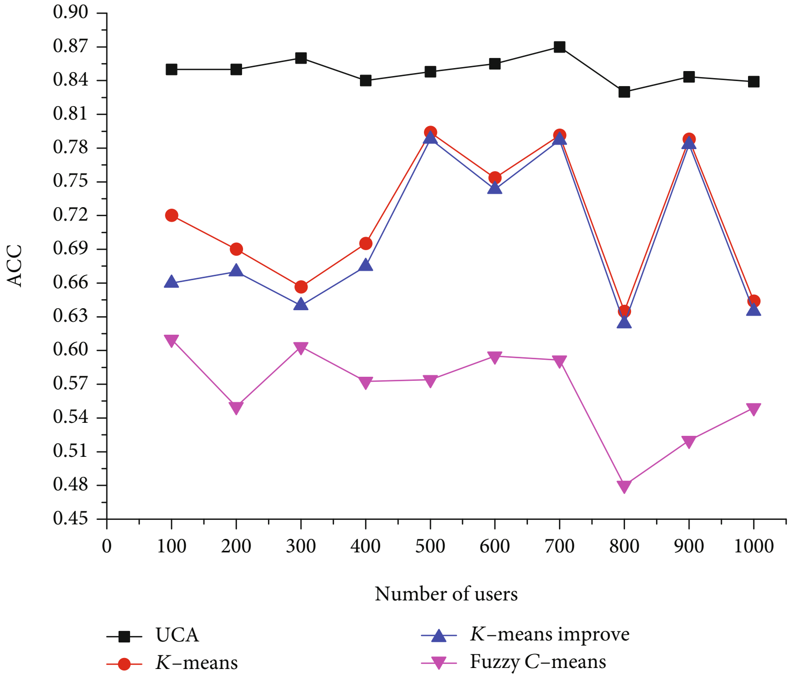


Figure 2: ACC vs. number of users.

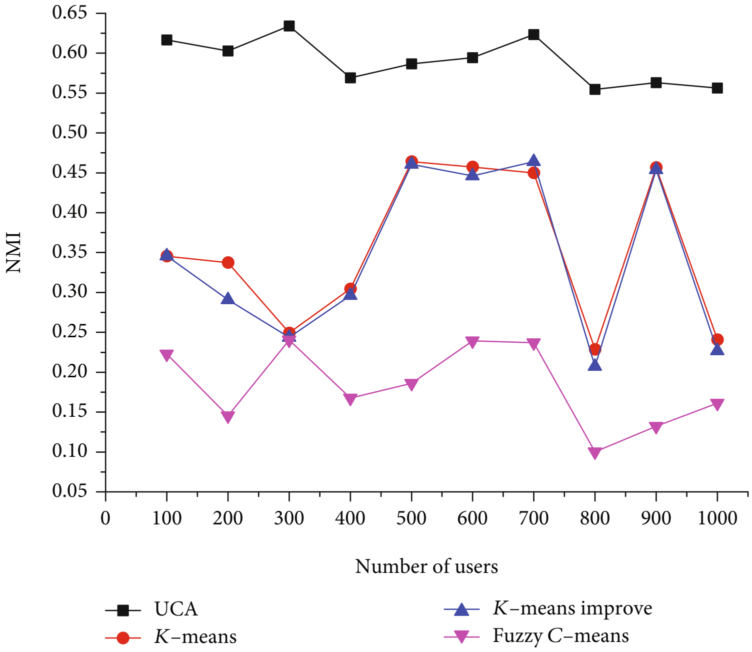


Figure 3: NMI vs. number of users.

rules, which can balance the number of users in different unions and ensure that users with high similarity are clustered together as much as possible. In addition, the accuracy of the K-means improve is lower than the K-means algorithm. This means that a single restriction on the size of users is not conducive to the formation of high-quality user clustering.

We also perform extensive simulations to validate the reduction of running time achieved by UCA under various user scales, as shown in Figure 5. As seen, the results of the four algorithms show an upward trend, and UCA has the lowest running time. The reasons are as follows: Firstly, the K-means algorithm uses random clustering centers to achieve user clustering through multiple iterations. The

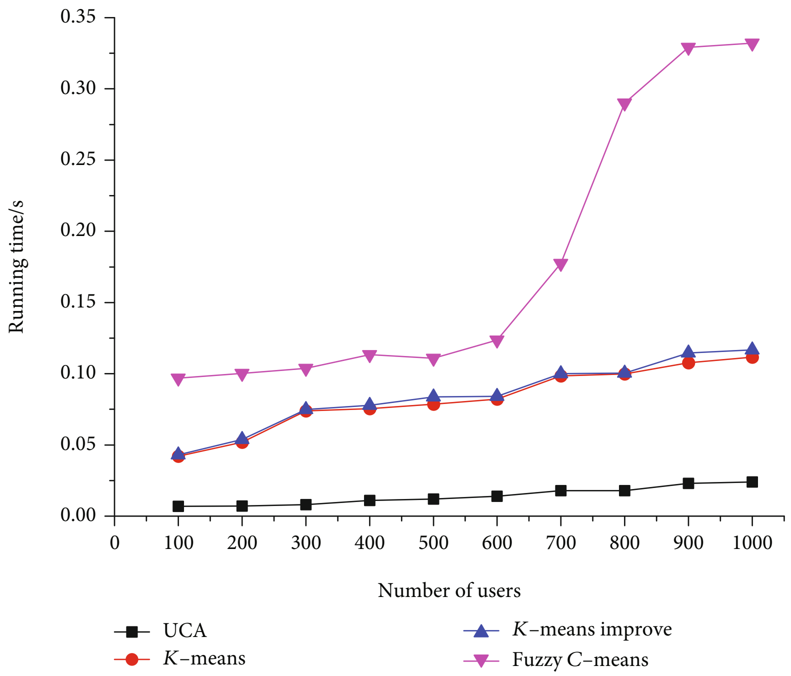


Figure 4: Running time vs. number of users.

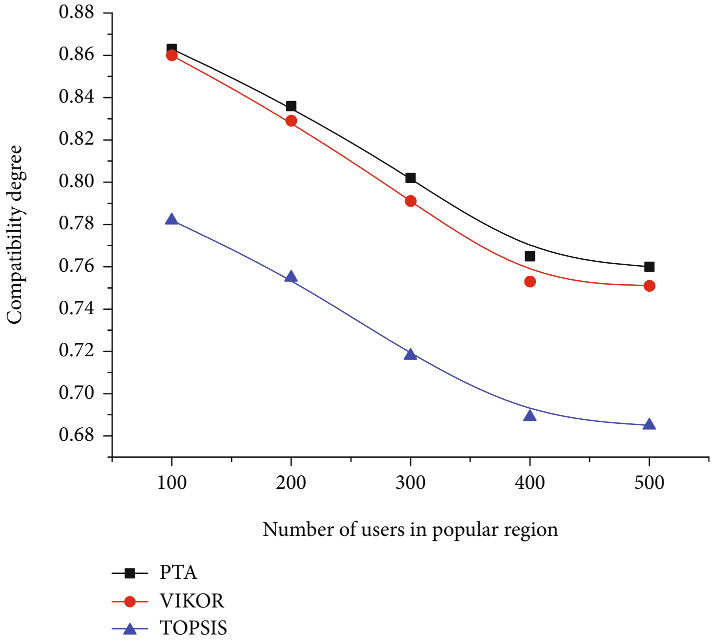


Figure 5: Compatibility degree vs. popular region.

growth of the user scale leads to more iterations and time overhead. In addition, setting user boundaries in this algorithm may cause more time costs. Unlike the *K*-means algorithm, UCA only needs to calculate the similarity between users and task centers, and compare boundary users to realize the user-union, which can reduce running time. Secondly, the Fuzzy *C*-means algorithm provides more flexible clustering results, but is more sensitive to boundary users. With the growth of user scale, the existence of enormous boundary users will require a longer time overhead for this algorithm.

In general, for the large-scale user clustering scenario, the proposed user-union clustering algorithm has the characteristics of high classification accuracy and fast calculation speed. It can provide high-quality user data for the preference-based task assignment.

Table 1: Decision matrix of task assignment scheme.

Scheme	G_1	G_2	G_3	G_4
Y_1	35	1521	0.69	0.68
Y_2	29	1487	0.78	0.67
Y_3	30	1495	0.77	0.72
Y_4	25	1503	0.72	0.74
Y_5	32	1508	0.74	0.69

Table 2: The value of B_i, R_i, and BR_i.

Scheme	B_i	R_i	BR_i
Y_1	0.984	0.3	1
Y_2	0.261	0.161	0
Y_3	0.413	0.183	0.184
Y_4	0.393	0.216	0.289
Y_5	0.673	0.222	0.504

5.3. Experiment Results of PTA. On the premise of ensuring high-quality clustering users, we perform the performance of PTA to realize the diversified task assignment under the individual preference sensing cost. In addition, different user scales may have various user characteristics, which affect the performance of execution time and compatibility. As a result, we first use an example to demonstrate the feasibility of small-scale data. Then, we consider PTA performance in two scenarios by calculating execution time and compatibility degree.

5.3.1. Example. According to UCA, we achieve five alternative task assignment schemes (i.e., the sensing cost for five users), as shown in Table 1.

Step 1. Normalize the sample matrix by Eqs. (5)–(7), G_1 and G_2 are the cost index, G_3 and G_4 are the benefit index. The decision selection sample matrix is

$$H = \begin{bmatrix} -1 & -1 & -1 & -0.5 \\ 0.2 & 0.619 & 0.5 & -0.75 \\ 0 & 0.238 & 0.333 & 0.5 \\ 1 & -0.143 & -0.5 & 1 \\ -0.4 & -0.381 & -0.167 & -0.25 \end{bmatrix}. \tag{21}$$

Then, we achieve the positive ideal assignment scheme (i.e., $Y^+ = \{1, 0.619, 0.5, 1\}$), and the negative ideal assignment scheme (i.e., $Y^- = \{-1,-1,-1,-0.75\}$).

Step 2. According to Equation (8), the correlation coefficient of positive and negative ideal scheme is

$$\begin{aligned} \varsigma^+ &= \begin{bmatrix} 0.333 & 0.333 & 0.333 & 0.368 \\ 0.556 & 1 & 1 & 0.333 \\ 0.500 & 0.680 & 0.818 & 0.636 \\ 1 & 0.515 & 0.429 & 1 \\ 0.417 & 0.447 & 0.529 & 0.412 \end{bmatrix}, \\ \varsigma^- &= \begin{bmatrix} 1 & 1 & 1 & 0.778 \\ 0.455 & 0.333 & 0.333 & 1 \\ 0.500 & 0.395 & 0.360 & 0.412 \\ 0.333 & 0.486 & 0.600 & 0.333 \\ 0.625 & 0.567 & 0.474 & 0.636 \end{bmatrix}, \end{aligned} \tag{22}$$

respectively.

Step 3. According to Equation (11), the positive and negative prospect value matrix of each scheme is

$$\begin{aligned} v^+ &= \begin{bmatrix} 0 & 0 & 0 & 0.266 \\ 0.586 & 0.700 & 0.700 & 0 \\ 0.543 & 0.643 & 0.675 & 0.627 \\ 0.700 & 0.557 & 0.446 & 0.700 \\ 0.422 & 0.479 & 0.568 & 0.411 \end{bmatrix}, \\ v^- &= \begin{bmatrix} -1.575 & -1.575 & -1.575 & -1.503 \\ -1.101 & 0 & 0 & -1.575 \\ -1.223 & -0.826 & -0.502 & -0.925 \\ 0 & -1.190 & -1.374 & 0 \\ -1.399 & -1.336 & -1.159 & -1.410 \end{bmatrix}, \end{aligned} \tag{23}$$

respectively.

Step 4. Optimize the index weights to obtain the best comprehensive prospect value by Equation (4), where $\omega_1, \omega_2, \omega_3, \omega_4 \in [0.1, 0.3]$. We achieve the optimal solution (i.e., $\omega^* = \{0.3, 0.3, 0.3, 0.1\}$) and the comprehensive prospect matrix is

$$v^* = \begin{bmatrix} -0.516 & -0.516 & -0.516 & -0.206 \\ -0.174 & 0.223 & 0.223 & -0.268 \\ -0.228 & -0.066 & 0.049 & -0.041 \\ 0.223 & -0.212 & -0.308 & 0.130 \\ -0.324 & -0.285 & -0.199 & -0.163 \end{bmatrix}. \tag{24}$$

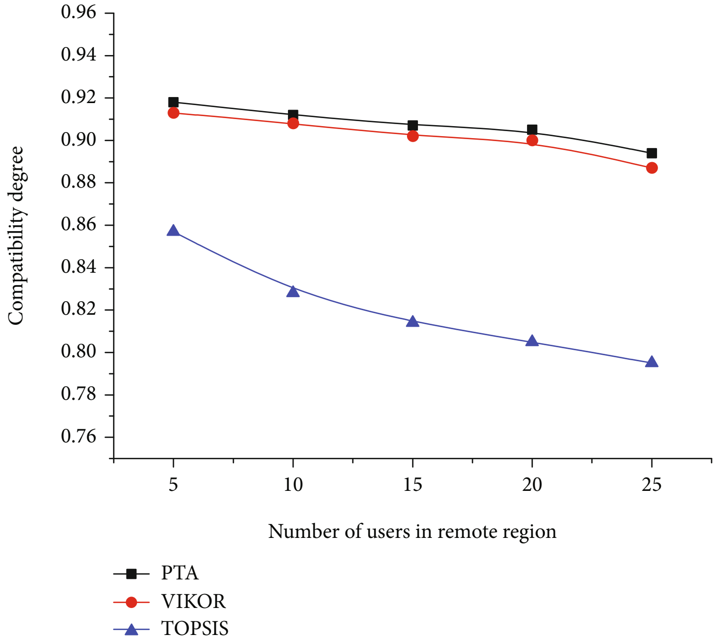


Figure 6: Compatibility degree vs. remote region.

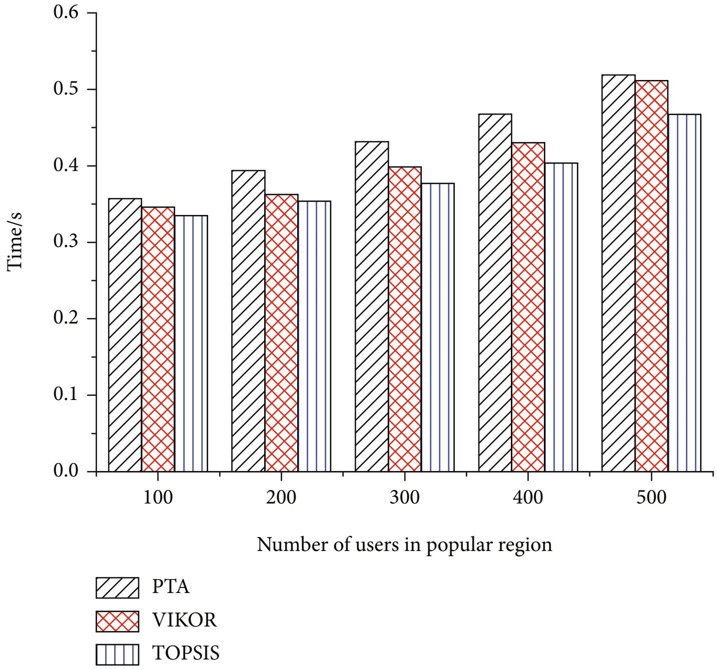


Figure 7: Execution time vs. popular region.

Step 5. Use the VIKOR method to sort the schemes. Calculate the value of B_i, R_i, and BR_i, as shown in Table 2.

Table 2 presents the value of B_i, R_i, and BR_i for the five users. As seen, Y_2 has the optimal value of BR_i and B_i, which satisfices Condition 2. Y_3 has the sub-optimal value of BR_i, and $BR_3 - BR_2 = 0.184 < 1/(5-1)$, which does not satisfy Condition 1. Obviously, Y_2 and Y_3 are both acceptable and ideal solutions. The service provider can choose Y_2 or Y_3 according to individual preference, and PTA implements the preference-based task assignment. In addition, we also found an interesting phenomenon that PTA usually chooses low-cost and high-quality schemes. The reason is as follows. Firstly, PTA solves the diversified task assignment under the individual preference sensing cost. That is, service providers play a decisive role in the task assignment. Considering the profit orientation of service providers, low-cost and high-quality schemes are more competitive in selection. Secondly, sensing users are competitive and work hard. Users try to

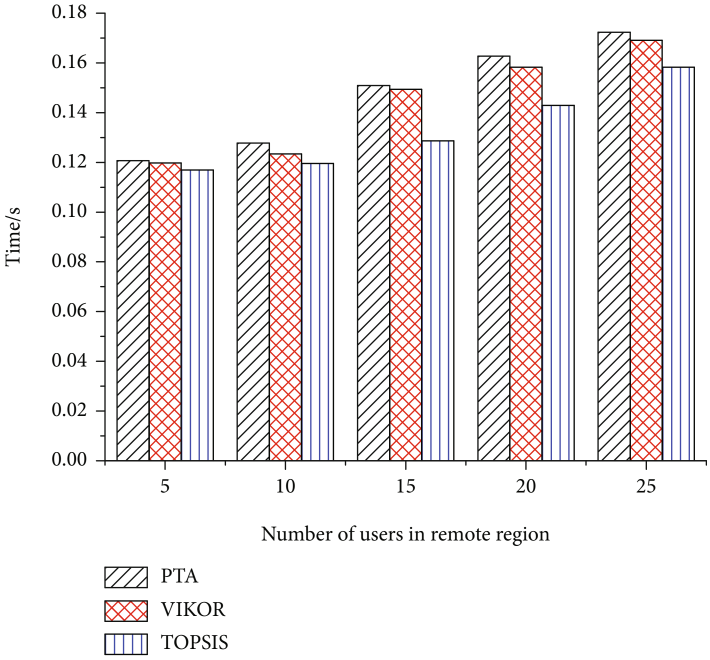


Figure 8: Execution time vs. remote region.

improve the quality of uploaded data to win in a task as much as possible. Thirdly, in the calculation results of the best comprehensive prospect value, the weight of the cost index is much greater than the benefit index, which further promotes PTA to choose low-cost and high-quality user solutions.

5.3.2. Performance Comparison. Next, we provide simulation results by three methods in various scenarios.

(1) Compatibility Degree. According to our definition of location regions [31], we conduct simulations to observe the effect of compatibility degree on different solutions when users are in different regions (i.e., popular region and remote region), as shown in Figures 4 and 6.

Generally, high compatibility degree means that the user data is representative and reliable, which means the higher accuracy of task assignment. Figures 4 and 6 show the performance of compatibility degrees under different numbers of users and regions. It is seen that as the number of users increases, the compatibility degree of these methods decreases. The compatibility degree of PTA is better than these two methods, and the compatibility degree in the remote region is better than in the popular region. The reason is as follows. First, as the number of users increases, more similar users participate in sensing tasks, especially in a popular region, which reduces the differences between users. As a result, the sensing platform is challenging to select suitable users, which leads to a decrease in the compatibility of these solutions. Second, compared with the VIKOR method, PTA adds prospect theory to reflect that decision-makers are more sensitive to losses than revenues. Poor indexes are more difficult to compensate by superior indexes, and the selected user data is more balanced to ensure the accuracy of task assignment. Third, compared with the TOPSIS method, PTA does not need to satisfy both the optimal positive ideal solution and the worst negative ideal solution. The final selection meets the individual preferences of the service provider. Furthermore, we also found that the performance of the three methods in remote regions is better than in popular regions. The reason is that large-scale users in popular regions lead to the high similarity between data, which makes it difficult to assign tasks accurately. On the contrary, the small scale of users in remote regions is conducive to accurate task assignment.

(2) Execution Time. We perform extensive simulations to validate the execution efficiency of PTA under various regions, compared with two solutions, as shown in Figures 7 and 8.

Figures 7 and 8 both show the execution time of PTA under various regions. We find that the execution time of the three solutions increases stably when the number of users enlarges. More alternative users in the sensing platform lead to more computational overhead. In addition, the execution time in the popular region is generally higher than that in the remote region. The reason is that more similar users are contained in the popular region, and more calculations are needed to find suitable candidate users. PTA is slightly worse than VIKOR and TOPSIS methods in execution time, because PTA makes user selection from multiple perspectives, which increases the execution time.

In general, the performance of PTA is acceptable in the preference-based task assignment. The reason is as follows. Firstly, PTA has a more significant advantage in the accuracy of user selection (i.e., the highest compatibility degree), which can ensure the accuracy of task assignment. Besides, as another part of the task assignment method, UCA has the characteristics of high classification accuracy and fast calculation speed, which can make up for the lack of execution time of PTA.

6. Conclusions

In this paper, we addressed a task assignment problem in MCS. We proposed a task assignment method based on user-union clustering and individual preferences. Specifically, we analyzed and formulated the task assignment problem from two perspectives, respectively. We first define the user similarity and propose a user-union clustering algorithm (UCA) to reduce data redundancy and achieve high-quality clustering data. Based on this solution, we further consider individual preferences of service providers and propose a preference-based task assignment algorithm (PTA) to meet the needs of diversified sensing cost and achieve the task assignment with individual preference. To evaluate the performance of the proposed solutions, we conducted extensive simulations. The results show that our method realizes the individual preference-based task assignment under the premise of ensuring high-quality clustering data. However, our method usually chooses low-cost and high-quality user data, which may suppress the revenues of users. At the same time, for the user-union, using exact values to evaluate data may reduce the accuracy of evaluation. In future works, we will balance the revenues between users and service providers, improve the accuracy of clustering data quality evaluation, and develop a task assignment method with lower complexities.

Data Availability

The authors declare that all the data and materials in this manuscript are available.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This research was funded by the Nature Science Foundation of China, grant number 61872104, and Fundamental Research Fund for the Central Universities in China, grant number 3072020CF0603.

References

[1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[2] W. Gong, B. Zhang, and C. Li, "Task assignment in mobile crowdsensing: present and future directions," *IEEE Network*, vol. 32, no. 4, pp. 100–107, 2018.

[3] A. Shahraki, A. Taherkordi, Ø. Haugen, and F. Eliassen, "Clustering objectives in wireless sensor networks: a survey and research direction analysis," *Computer Networks*, vol. 180, article 107376, 2020.

[4] W. Liu, Y. Yang, E. Wang, and J. Wu, "User recruitment for enhancing data inference accuracy in sparse mobile crowdsensing," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1802–1814, 2020.

[5] F. Campioni, S. Choudhury, K. Salomaa, and S. G. Akl, "Improved recruitment algorithms for vehicular crowdsensing networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1198–1207, 2019.

[6] Y. Zhao and C. H. Liu, "Social-aware incentive mechanism for vehicular crowdsensing by deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2314–2325, 2021.

[7] W. Gong, B. Zhang, C. Li, and Z. Yao, "Task allocation in semi-opportunistic mobile crowdsensing: paradigm and algorithms," *Mobile Networks and Applications*, vol. 25, no. 2, pp. 772–782, 2020.

[8] X. H. Wei, Z. J. Li, Y. Y. Liu, S. Gao, and H. S. Yue, "SDLSC-TA: subarea division learning based task allocation in sparse mobile crowdsensing," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1344–1358, 2021.

[9] C. Wang, Z. Xie, L. Shao, Z. Zhang, and M. Zhou, "Estimating travel speed of a road section through sparse crowdsensing data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3486–3495, 2019.

[10] E. Wang, M. Zhang, X. Cheng et al., "Deep learning-enabled sparse industrial crowdsensing and prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6170–6181, 2021.

[11] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16–28, 2018.

[12] H. Xiong, D. Zhang, L. Wang, J. P. Gibson, and J. Zhu, "EEMC," *ACM transactions on intelligent systems and technology*, vol. 6, no. 3, pp. 1–26, 2015.

[13] J. T. Wang, Y. S. Wang, D. Q. Zhang, F. Wang, Y. D. He, and L. T. Ma, "PSAllocator: Multi-Task Allocation for Participatory Sensing with Sensing Capability Constraints," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 1139–1151, Portland, Oregon, USA, 2017.

[14] A. Pugazhenthi and L. S. Kumar, "Selection of Optimal Number of Clusters and Centroids for K-means and Fuzzy C-means Clustering: A Review," in *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, pp. 1–4, Patna, India, 2020.

[15] W. Gong, B. Zhang, and C. Li, "Location-based online task scheduling in mobile crowdsensing," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, Singapore, 2017.

[16] H. Xiong, D. Zhang, G. Chen, L. Wang, V. Gauthier, and L. E. Barnes, "*iCrowd*: near-optimal task allocation for piggyback crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 2010–2022, 2016.

[17] Z. B. Wang, R. Tan, J. H. Hu et al., "Heterogeneous incentive mechanism for time-sensitive and location-dependent crowdsensing networks with random arrivals," *Computer Networks*, vol. 131, pp. 96–109, 2018.

[18] B. Zeng, X. Yan, X. Zhang, and B. Zhao, “BRAKE: bilateral privacy-preserving and accurate task assignment in fog-assisted mobile crowdsensing,” *IEEE Systems Journal*, vol. 15, no. 3, pp. 4480–4491, 2021.

[19] Y. Huang, H. Chen, G. Ma et al., “OPAT: optimized allocation of time-dependent tasks for mobile crowdsensing,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2476–2485, 2022.

[20] A. Alagha, R. Mizouni, S. Singh, H. Otrok, and A. Ouali, “SDRS: a stable data-based recruitment system in IoT crowdsensing for localization tasks,” *Journal of Network and Computer Applications*, vol. 177, article 102968, 2021.

[21] S. Song, Z. Liu, Z. Li, T. Xing, and D. Fang, “Coverage-oriented task assignment for mobile crowdsensing,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7407–7418, 2020.

[22] F. Yucel, M. Yuksel, and E. Bulut, “Coverage-aware stable task assignment in opportunistic mobile crowdsensing,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3831–3845, 2021.

[23] W. Z. Guo, W. P. Zhu, Z. Y. Yu, J. T. Wang, and B. Guo, “A survey of task allocation: contrastive perspectives from wireless sensor networks and mobile crowdsensing,” *IEEE Access*, vol. 7, pp. 78406–78420, 2019.

[24] Y. Du, Y. E. Sun, H. Huang et al., “Bayesian co-clustering truth discovery for mobile crowd sensing systems,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1045–1057, 2020.

[25] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, “Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems,” *IEEE/ACM Transactions on Networking*, vol. 26, no. 5, pp. 2019–2032, 2018.

[26] Z. H. Shao, H. Q. Wang, and G. S. Feng, “PUEGM: a method of user revenue selection based on a publisher-user evolutionary game model for mobile crowdsensing,” *Sensors*, vol. 19, no. 13, p. 2927, 2019.

[27] J. Peng, Y. Zhu, Q. Zhao et al., “Fair energy-efficient sensing task allocation in participatory sensing with smartphones,” *The Computer Journal*, vol. 60, no. 6, pp. 850–865, 2017.

[28] A. Tversky and D. Kahneman, “Advances in prospect theory: cumulative representation of uncertainty,” *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[29] M. Shumaiza, A. N. A.-k. Akram, and J. C. R. Alcantud, “Group decision-making based on the VIKOR method with trapezoidal bipolar fuzzy information,” *Symmetry*, vol. 11, no. 10, p. 1313, 2019.

[30] S. Firdose, W. Lopes, R. Moreira, and P. Sofia, “CRAWDAD dataset copelabs/usense (v. 2017-01-27),” http://crawdad.org/copelabs/usense/20170127.

[31] Z. H. Shao, H. Q. Wang, Y. F. Zou, Z. H. Gao, and H. W. Lv, “From centralized protection to distributed edge collaboration: a location difference-based privacy-preserving framework for mobile crowdsensing,” *Security and Communication Networks*, vol. 2021, Article ID 5855745, 8 pages, 2021.

[32] K. Palczewski and W. Salabun, “The fuzzy TOPSIS applications in the last decade,” *Procedia Computer Science*, vol. 159, pp. 2294–2303, 2019.

[33] C. Ceballos and V. Pilaud, “Denominator vectors and compatibility degrees in cluster algebras of finite type,” *Transactions of the American Mathematical Society*, vol. 367, no. 2, pp. 1421–1439, 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3574588, 11 pages
https://doi.org/10.1155/2022/3574588

Research Article

A Combined Detection Algorithm for Personal Protective Equipment Based on Lightweight YOLOv4 Model

Li Ma, Xinxin Li, Xinguan Dai, Zhibin Guan, and Yuanmeng Lu

College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an, 710600 Shaanxi, China

Correspondence should be addressed to Xinxin Li; 1505011424@qq.com

Received 27 February 2022; Accepted 13 April 2022; Published 4 May 2022

Academic Editor: Maode Ma

Improper wearing of personal protective equipment may lead to safety incidents; this paper proposes a combined detection algorithm for personal protective equipment based on the lightweight YOLOv4 model for mobile terminals. To ensure high detection accuracy, a channel and layer pruning method (CLSlim) to lightweight algorithm is used to reduce computing power consumption and improve the detection speed on the basis of the YOLOv4 network. This method applies L1 regularization and gradient sparse training on the scaling factor of the BN layer in the convolutional module: global pruning threshold and local safety threshold are used to eliminate redundant channels, the layer pruning threshold is used to prune the structure of the shortcuts in the Cross Stage Partial (CSP) module for inference speed improvement, and finally, a lightweight network model is obtained. The experiment improves the YOLOv4 and YOLOv4-Tiny models for CLSlim lightweight separately in GTX2080ti environment. Results show that (1) CLSlim-YOLOv4 compresses the YOLOv4 model parameters by 98.2% and increases the inference speed by 1.8 times with mAP loss of only 2.1% and (2) CLSlim-YOLOv4-Tiny compresses the original model parameters by 74.3% and increases the inference speed by 1.1 times with mAP increase of 0.8%, which certificates that this improved lightweight algorithm serves better for the real-time ability and accuracy of combined detection on PPE with mobile terminals.

1. Introduction

Personal protective equipment (PPE) is equipment for workers avoiding or lightening accident injury at work. Common PPEs include safety hard hats, reflective clothing, and protective clothing in construction scenes [1]. OSHA (occupational safety and health administration) stipulates that workers must wear safety hard hats when entering the construction site, and special types of work shall wear appropriate personal protective equipment. Workers working at heights must wear safety hard hats and safety belts [2]. Outdoor workers shall wear safety hard hats and reflective clothes, etc. [3]. The traditional image-based PPE detection algorithm needs to extract the key region features first and then use the edge information or classification algorithm to recognize the PPE. Reference [4] uses the template matching method to judge personnel wear safety belts. Reference [5] uses edge contour information to identify hard hats. In Reference [6], it demonstrates the application of Artificial Intelligence (AI) and machine vision for the identification of personal protective equipment (PPE), particularly safety glasses in zones of the learning factory, where safety risks exist. Traditional PPE detection methods have the disadvantages of low precision and slow speed. However, with the rapid development of convolutional neural networks in the field of machine vision, many scholars use end-to-end target detection algorithms to detect PPE and achieve good results. Reference [7] uses the SSD target detection algorithm to detect the hard hat in real time and recognize its color information. In Reference [8], a convolutional neural network is used to identify workers and hard hats, and the normalization of wearing hard hats according to the overlap value of workers' heads and hard hats is verified.

Due to the complexity of the construction environment, workers wearing a single PPE could not fully protect their own safety, while the combined detection algorithm of multiple types of PPEs needs to verify the standardization of use at the same time. The verification method proposed in

Reference [8] will increase exponentially with the increase of PPE components, which will affect the recognition speed. Based on the YOLOv3 algorithm, Reference [9] detects multiple types of PPEs and verifies the standardization of wearing, which has high real-time performance but poor recognition accuracy for small targets or low resolution picture. At present, there are many ideas worthy of reference in academic circles to improve the detection accuracy of the algorithm, such as data enhancement methods that only increase the training cost without affecting the inference speed or inserting attention mechanism modules that only increase a small amount of reasoning cost in the training process [10, 11]. In 2020, Bochkovskiy et al. proposed their YOLOv4 algorithm [12]. Combining the popular convolutional network optimization techniques and using more complex network structures, this algorithm was able to proceed with fast and accurate training and detection on lower configuration servers and was identified as an excellent target detection algorithm. But the huge model and parameter calculation volumes make it not suitable for mobile ends in industrial scenarios with limited resources. Therefore, under the premise of ensuring high detection accuracy, reducing floating point operations, improving the inference speed, and making it deployable to mobile terminals with limited resources are an urgent problem that needs to be solved. By using the channel pruning method, the parameter quantity of YOLOv3 is compressed by 92%. Reference [13] maintains the detection accuracy of the original model and improves the inference speed by twice. By channel pruning of the improved YOLOv4 model, the algorithm in Reference [14] lost 2.43% of the detection accuracy, increased the prediction speed by 2.9 times, and compressed the model by 96%. By designing a lightweight convolutional neural network (CNN) which is named as Shuffle CNN, a Shuffle CNN-based AMC (Shuffle AMC) method is proposed for the ubiquitous IoT cyberphysical systems with orthogonal frequency division multiplexing (OFDM) in Reference [15]. For facial landmark detection, Reference [16] presents a novel loss function to train a lightweight student network (e.g., MobileNetV2).

To solve the problems of combined detection on the workers' multiple PPEs and improve the real-time performance and detection accuracy of terminals with limited resources in complex networks, this paper proposes a high accuracy PPE real-time detection algorithm with a smaller volume. Based on the popular YOLOv4 and YOLOv4-Tiny networks for model lightweight, it compresses the model efficiently by combining channel and layer pruning methods and gets a combined detection algorithm on PPEs with small volume and fast detection speed, which is suitable for mobile ends in industrial scenarios.

2. Improved YOLOv4 for PPE Real-Time Detection Algorithm

As the improved version on v3, YOLOv4 integrates the idea of the convolutional neural network algorithm based on the original YOLO frame and uses many strategies on the backbone network of feature extraction, neck network of feature fusion and the detection head of classification, and regression for the improvement of the v3 algorithm.

The YOLOv4 network structure is shown in Figure 1, and CSPDarknet53 is used as the backbone network. In the structure, the CSP structure can be lightweight and simultaneously improve the learning ability of CNN, reduce computing bottlenecks, and reduce memory costs [17]. CBM and CBL are joined up with batch normalization (BN) operation after regular convolution (Conv), with commonly used activation functions of Leaky ReLu, Mish, etc. Before feature fusion, the SPP module is introduced, which can effectively increase the network receptive field and obviously separate the contextual features compared to max pooling operation. The Path Aggregation Network (PANet) is the enhanced feature pyramid network [18], which effectively improves the problem of losing shallow feature information in the deep network by combining the methods of bottom-up and top-down paths [19].

To improve real-time performance, a lightweight algorithm YOLOv4-Tiny is proposed on the basis of YOLOv4, which is showed in Figure 2. In this YOLOv4-Tiny lightweight network model, three residual modules are used in the CSPDarknet53 backbone network, the Leaky ReLu function is used as the activation function, the FPN network is used in the multiscale feature fusion module, and two detection heads are used in classification and regression of the prediction module.

2.1. Activation Functions for the Modification of Class Probability. In regular target detection tasks, one object may belong to multiple categories as Figure 3(a). When there are many overlapping categories in the data set, a single detection box can be used to detect multiple classes simultaneously (e.g., person and male and dog and pug). Therefore, the single-label classification method has limitations in real scenes; the original YOLOv4 algorithm supposes that all classes are nonmutually exclusive. And the activation function sigmoid is used for the calculation of class probability as shown in

$$\sigma_{\text{sigmoid}}(z_i) = \frac{e^{z_i}}{e^{z_i} + 1}. \quad (1)$$

The function processes each class i independently and normalizes the prediction probability z_i of each class between [0, 1]. If $\sigma_{\text{sigmoid}}(z_i)$ is bigger than a certain threshold, like 0.5, there is a class in the grid cell; that is to say, an object can be predicted as multiple classes. And this paper focuses on the combined detection of workers wearing different classes of PPEs as shown in Figure 3(b). A worker can only belong to one category. For instance, the semantic definition of the classes is given as W, WH, WV, and WHV. If one worker is detected as a certain class (i.e., WHV), the other classes (i.e., W, WH, and WV) of the same target will be replaced. Therefore, the activation function SoftMax is used to calculate the class probabilities as shown in

$$\sigma_{\text{SoftMax}}(z)_i = \frac{e^{z_i}}{\sum_{i=1} e^{z_j}}. \quad (2)$$

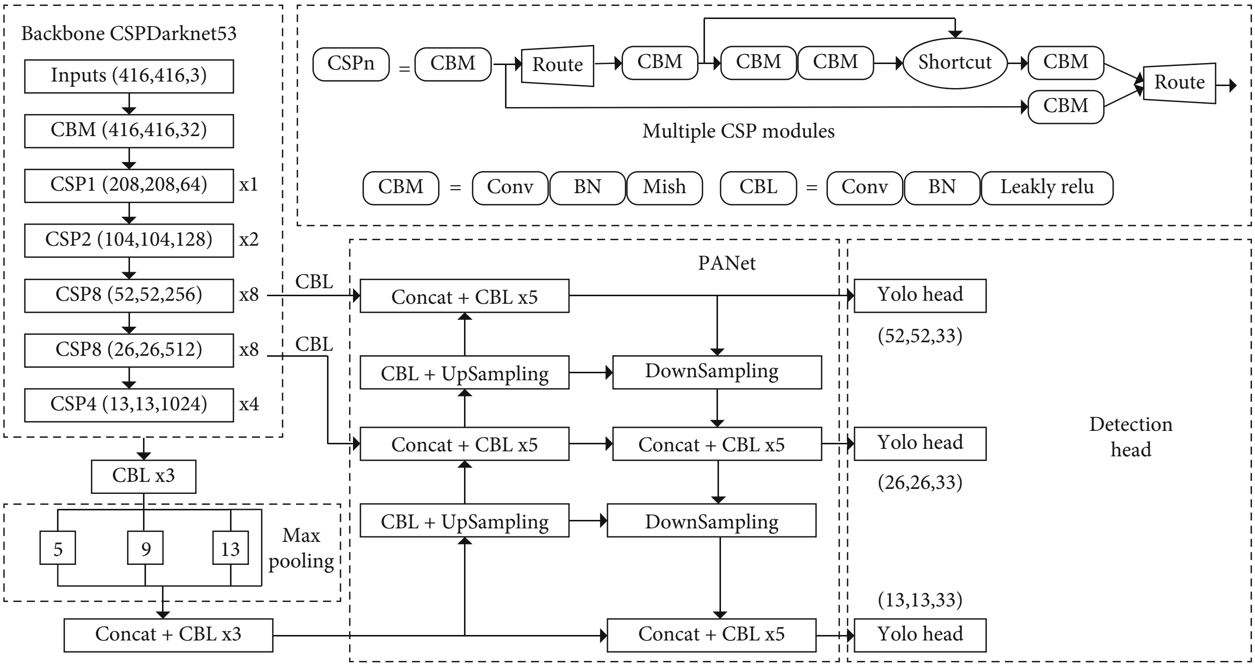


Figure 1: YOLOv4 algorithm structure.

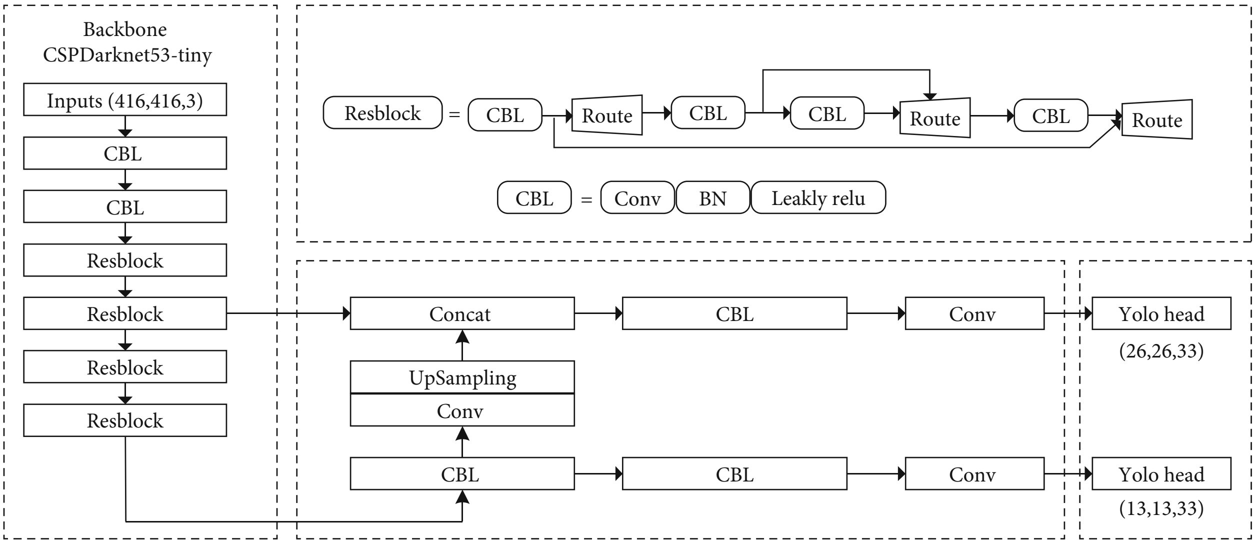


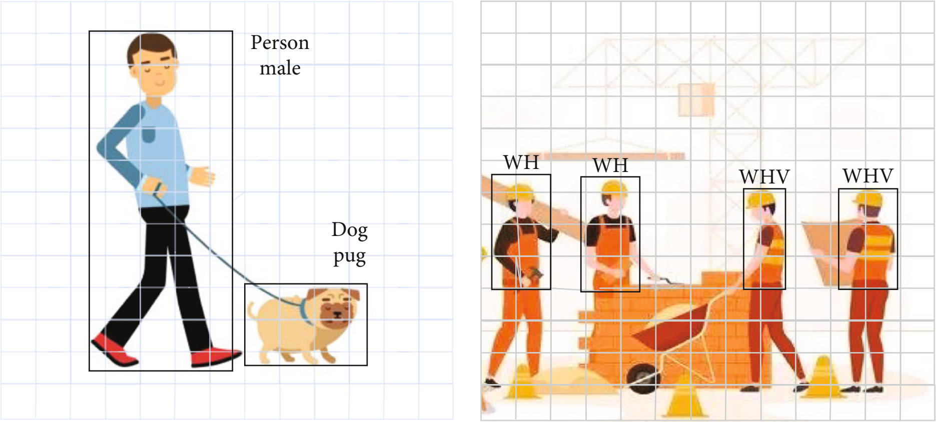
Figure 2: YOLOv4-Tiny algorithm structure.

The function supposes that all classes i are exclusive, and it normalizes the prediction probability z_i of each class and makes the sum of them 1.

2.2. Detection Box Modification and Duplication Strategy. In the prediction stage, the original YOLOv4 algorithm proceeds Nonmaximum Suppression (NMS) for one class each time because the combined detection method in this paper marks only the worker's upper part of the body; the similarity of classes is high. If a prediction box belongs to Class A and Class B at the same time, it is redundancy. So, the regular NMS algorithm is used for the prediction box of a certain class and afterwards all classes to eliminate the duplication of the same worker detected as multiple classes as shown in Figure 4.

3. Model Lightweight Based on CLSlim-YOLOv4

3.1. BN Layer and Scaling Factor. BN [20] was a data normalization method proposed, and it has been applied in most CNNs. Traditional standardization methods distribute the input of CNN between $[0, 1]$, but most of the activation functions in CNN, such as sigmoid and tanh, are linearly distributed in the interval $[0, 1]$, and standardization methods can reduce the nonlinear capability of the network.



(a)

(b)

FIGURE 3: Classes in regular object detection vs. multiple PPE detection.

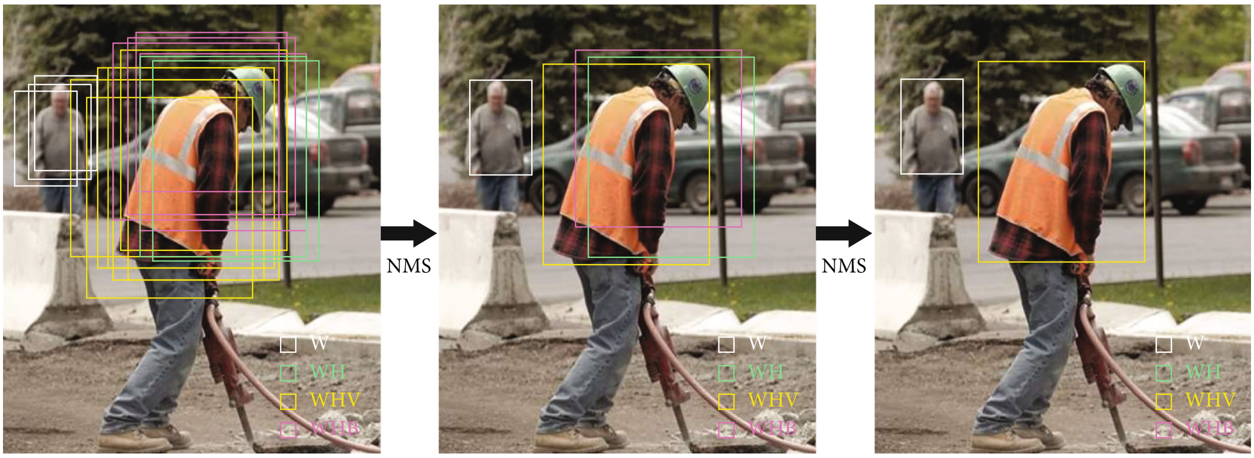


FIGURE 4: Two-stage method of NMS.

In order to reduce the effects of standardization on activation functions, two parameters under learning: scaling factor γ and shifting factor β are imported on the basis of standardization, and the values after standardization are zoomed and panned, which can regain the nonlinear expression ability of the convolution network to a certain extent. The flow of the BN algorithm is followed, in which $X_{\text{minibatch}}$ and y_i are the input and output of the BN layer; μ_X and σ_X^2 are the mean and variance values of the input of the BN layer; $\widehat{x}_i$ is the result after standardization.

```
Input: $X_{\text{minibatch}} = \{x_1, x_2, \cdots, x_m\}$ ;
        Parameter: $\gamma, \beta$;
Output: $\{y_i = BN_{\gamma,\beta}(x_i)\}$ ;
$\mu_x \longleftarrow 1/m\sum_{i=1}^{m} x_i$                    //Mini-batch min
$\sigma_X^2 \longleftarrow 1/m\sum_i^m (x_i - \mu_X)$          //Mini-batch variance
$\widehat{x}_i \longleftarrow x_i - \mu_X / \sqrt{\mu_X^2 + \varepsilon}$             //Standardization
$y_i \longleftarrow \gamma\widehat{x}_i + \beta \equiv BN_{\gamma,\beta}(x_i)$           //Scale and shift
```

ALGORITHM 1

In the YOLOv4 network, most convolution structures are composed of the convolution layer, BN layer, and activation function as shown in the CBM and CBL modules in Figure 1. If the scaling factor of the BN layer is very small, the value input into the activation function is very small, which represents the contribution of the corresponding channel to the network is also very low. Therefore, γ of the BN layer can be used as the scaling factor of channel pruning to evaluate the importance of the channel to the network without additional costs.

3.2. Sparse Training. In the process of network sparse training, γ in the CBM and CBL modules of the BN layer is considered the scaling factor of channel pruning and multiplied by the corresponding channel. Then, a sparse model is contained after combined training on network weights and scaling factors. Most of the scaling factors of channels tend to 0, and the corresponding loss function is shown in

$$\text{Loss} = \sum_{(x,y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \tau} g(\gamma), \tag{3}$$

where (x, y) are the training input sample and the corresponding label; Loss is the loss function of CNN normal

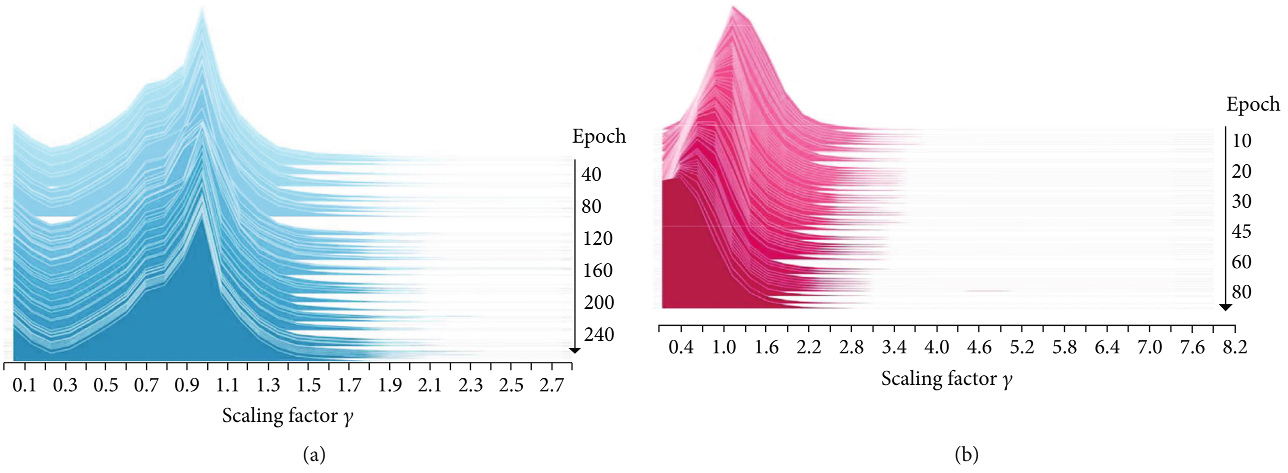


Figure 5: (a) Distribution of scaling factor without sparse process; (b) distribution of scaling factor with sparse process.

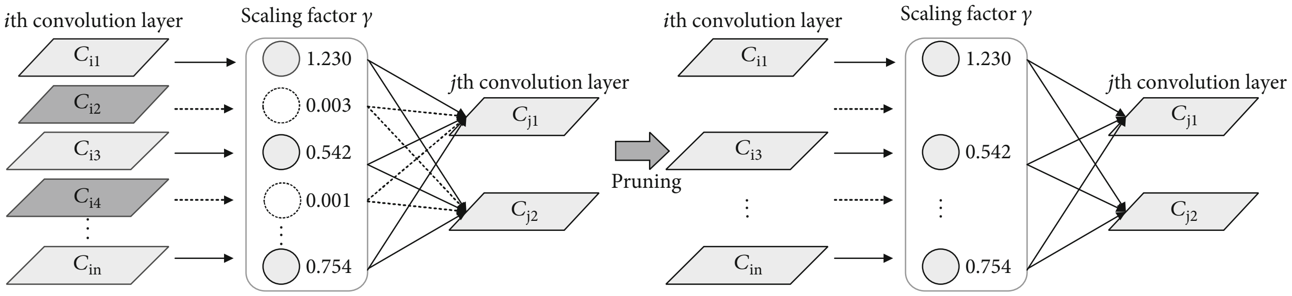


Figure 6: Model sparse training and channel pruning.

training; W is the weight of the network to be trained; $g(\gamma)$ is the penalty function on scaling factor, adopting $g = |s|$, i. e., L1 regularization; and λ is the penalty coefficient balancing two weights.

The essence of model channel pruning is cutting out the connection between the input and output related to a channel. Due to the combined optimization of loss function under normal training and scaling factor γ, a valuable channel can be chosen based on the scaling factor during the sparse training process. In the training process, the scaling factor γ of the BN layer shows approximately a normal distribution that expectation is 1 in the nonsparse YOLOv4 network as shown in Figure 5(a). When the penalty coefficient in equation (3) is set as 0.0005, after sparse training on the model, most of the scaling factor γ all go close to 0 as shown in Figure 5(b).

3.3. Channel and Layer Pruning. After sparse training, the scaling factor γ introduced in Section 3.1 is taken as the basis evaluating the importance of the channel. This paper defines a global threshold to control the pruning ratio and introduces a local safety threshold to prevent overpruning on the number of convolution layer channels and maintain the integrity of network connectivity. Figure 6 shows the model sparse training and channel pruning: each channel of the kth convolution layer is given a scaling factor; after sparse training, the scaling factor approaches 0; the absolute value of the scaling factor smaller than the global threshold is removed; if the scaling factors of the channels in the whole layer are small than the global threshold, the channels with scaling factors bigger than the local safety threshold are reserved to prevent the whole layer pruned.

In the YOLOv4 network pruning process, some structures need to be handled, such as the CSPn module in the backbone network CSPDarknet53; and the max pool and unsample layers independent of the number of channels can be ignored directly. In the channel pruning process, the pruning ratio is settled firstly. And then, $|\gamma|$ of the BN layer to be pruned is ascending sort. And the global threshold $\tilde{\gamma}$ and local safety threshold π are determined based on the pruning ratio and channel ratio to be reserved for each layer. Channels to be deleted in each layer are set to 0, and others 1. Then, the pruning mask is obtained. The CFG structure of CSP1 in the Darknet framework is shown in Figure 7. As to the route layer, the characteristic chart of the corresponding index is output when there is only one parameter; concatenate operation is proceeded when there are two parameters. Therefore, the pruning masks of the corresponding input layers are connected to and used as their own pruning masks. The structure of the shortcut layer is similar to the residual module of ResNet. Therefore, all layers for shortcut connection need the same number of channels. The final pruning masks are generated after traversing the pruning masks of these layers and making logic or operation.

Channel pruning can greatly reduce the model and parameter calculation volumes but has little impact on

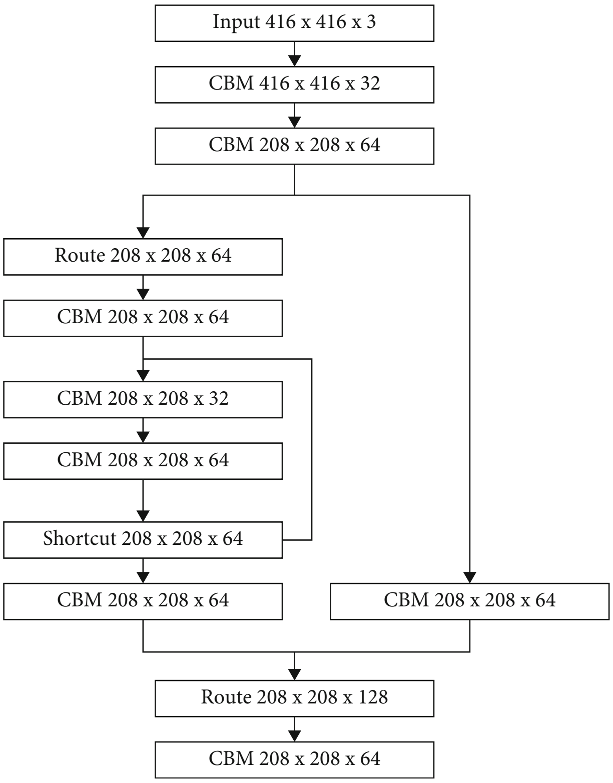


Figure 7: CFG structure of the CSP1 module.

improving the inference speed of the model. This paper proposes a layer pruning method on the basis of channel pruning, which works by sorting the scaling factor's mean value γ_{mean} for each convolution layer, pruning the layer with the smallest γ_{mean}, and introducing layer pruning coefficient S which is the number of shortcut structures to be pruned. There are 23 shortcut structures in the backbone network of YOLOv4. For the integrity of the network structure, each shortcut with the corresponding two convolution layers in its upper layer is pruned at the same time. If the layer pruning coefficient S is 8, there will be 24 layers pruned. Layer pruning can improve the inference speed of the model. Channel pruning and layer pruning are used to compress the width and depth of the model, respectively. The pruned model has a great improvement in parameter calculation, memory ratio, and inference speed.

3.4. Multiple Iterative Pruning and Fine-Tuning Model. Due to the reduction of the number of model channels and layers, there will be accuracy loss inevitably. Therefore, using the original data set to fine-tune the pruned model to regain accuracy of the model is necessary. In order to protect the model's complete network structure and high detection performance after pruning, the threshold with less precision loss can be used for pruning each time, and pruning and fine-tuning can be proceeded many times until the best pruning performance is achieved. The multiple pruning process is shown in Figure 8.

The CLSlim method proposed in the paper combines channel pruning and layer pruning to compress the model. After the CLSlim method used in YOLOv4, the number of the channel and layer can be reduced largely and the structure of the original model can be kept meanwhile, which will provide the possibility for application on embedded devices. Setting different layer pruning coefficients will cut off different numbers of shortcut structures, so the structure diagram of the pruning model is not fixed. When the layer pruning coefficient is 20, 20 shortcuts of backbone network CSPDarknet53 in the YOLOv4 will be pruned. Then, the pruning model network structure can be obtained as Figure 9, where Slim-CSPn is the pruned CSPn structure. Compared with the model before pruning, the CSPDarknet53 backbone network has been reduced by 60 layers.

The YOLOv4-Tiny network can also be lightweight improved using the CLSlim method. Since YOLOv4-Tiny's backbone CSPDarknet-Tiny network does not contain the shortcut structure, it only needs to make channel pruning in YOLOv4-Tiny to achieve the model compression effect.

4. Experiment Analysis

4.1. Experiment Environment. The experiment environment in this paper is under the Windows 10 professional operating system. Pytorch and Darknet Deep Learning Frameworks are applied for the realization of the combined detection algorithm on PPE. The configuration of the server is the graphics card of NVIDIA GTX2080ti and processor of Intel Core i7; RK3399pro is adopted as the performance verification embedded platform.

4.2. Experiment Data Set. Construction scenes generally separate into aerial work and ground work. As the most common and basic PPE for workers, a safety hard hat needs to be worn regularly at any time entering the construction site. Ground workers are mostly exposed to and in the shade of the sun. Due to the heavy dust on the construction site, it is difficult to accurately identify the positions of workers in this kind of bad environment. Therefore, ground workers should wear reflective clothing throughout the whole working period to avoid collision accidents; aerial workers should wear safety belts to avoid falling accidents. It takes the two kinds of works as an example, and the construction data set for PPE combined detection is shown in Table 1.

Experiment data mainly comes from the surveillance video of a building construction site in Xi'an. The camera takes pictures of workers standing, squatting, walking, and in other positions from multiple angles. But there is not enough data against rules. With these unbalanced sample categories, it might cause inaccurate experimental results. Therefore, to expand the data set, pictures of some certain categories against rules are taken independently at the construction site. The labeling boxes and sample pictures of each category are listed in Table 2. Label image is used to label the pictures, and data set in VOC format is set up, and samples are divided into the training set and test set by 8 : 2 for model training and performance verification, respectively.

4.3. Model Training and Pruning. The multiple iterative pruning and fine-tuning process in Section 3.4 is applied

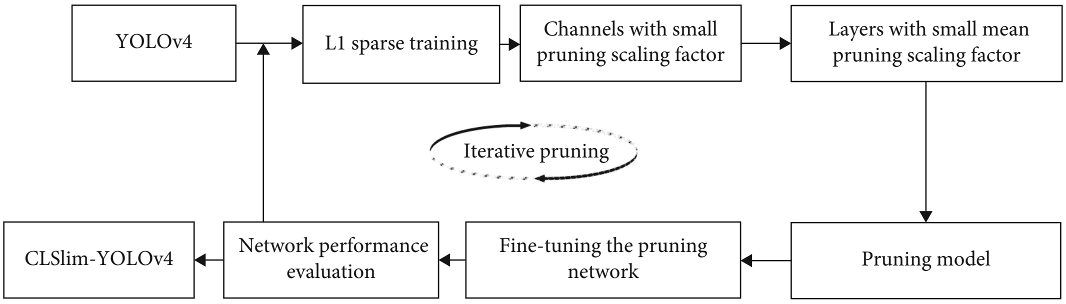

YOLOv4
L1 sparse training
Channels with small pruning scaling factor
Layers with small mean pruning scaling factor
Iterative pruning
Pruning model
Fine-tuning the pruning network
Network performance evaluation
CLSlim-YOLOv4

Figure 8: Multiple pruning process.

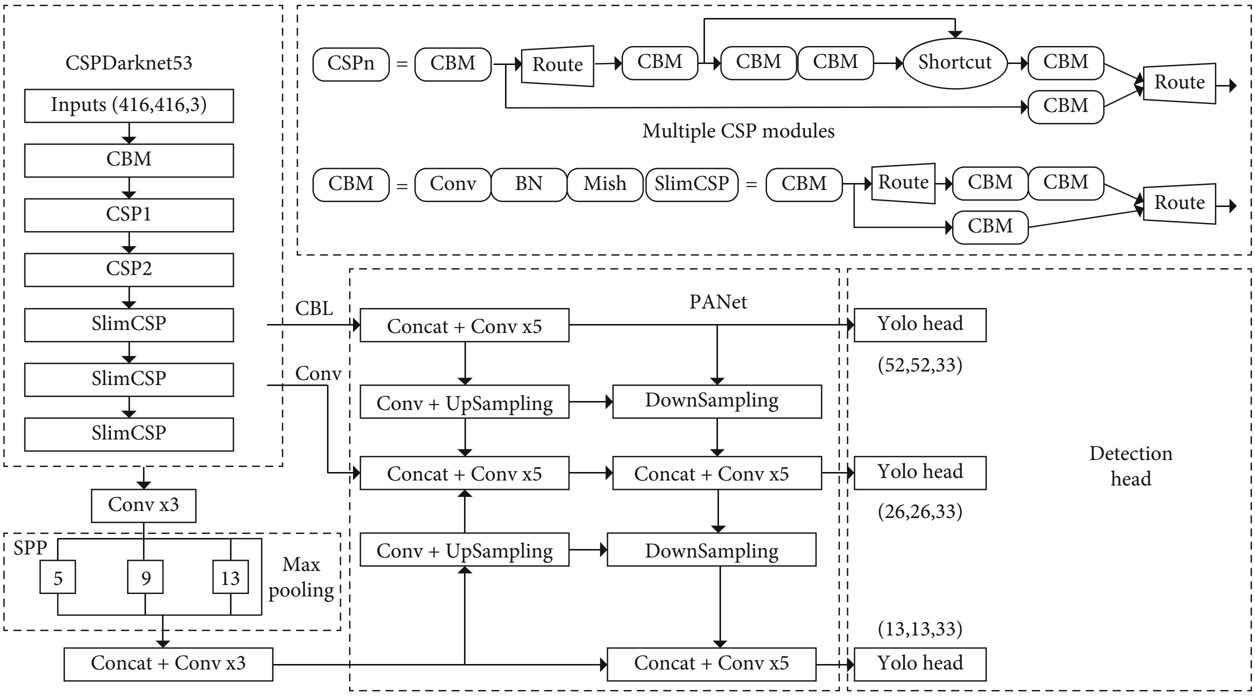


Figure 9: CLSlim-YOLOv4 structure ($S = 20$).

Table 1: The data set of combined detection on personal protective equipment.

Scene	Unsafe behavior	Label	Description
Ground work	Without safety hard hat Without reflective clothing	W	Worker
		WH	Worker with safety hard hat
		WV	Worker with reflective clothing
		WHV	Worker with safety hard hat and reflective clothing
Aerial work	Without safety hard hat Without safety belt	W	Worker
		WH	Worker with safety hard hat
		WB	Worker with safety rope
		WHB	Worker with safety hard hat and safety rope

for channel pruning and layer pruning on the YOLOv4 network, and then, the CLSlim-YOLOv4 model is obtained. The performance of the model is verified by using the self-built PPE real-time detection data set. The model training needs to be trained on computers with high configuration hardware and high-performance GPU graphics cards, and the experimental environment should be set up, and various dependency libraries should be installed on the server. For the comparison of the performances of the pruning model and nonpruning model, the self-built PPE real-time detection data set is imported to train the YOLOv4 model firstly; then, the trained model is used to prune. The model pruning experiment is also based on the self-built data set. And the pruning process is as follows:

Table 2: Numbers of categories and sample pictures.

(a)

Label	W	WH	WV
Number of labeling boxes	1756	5327	2227
Sample picture			

(b)

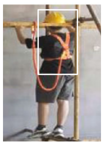
Label	WHV	WB	WHB
Number of labeling boxes	7284	3797	5442
Sample picture			

Table 3: Experimental parameters.

Parameter	Value
Learning rate (learning rate)	0.002324
Number of iterations (epoch)	400
Batch size (batch_size)	8
Momentum (momentum)	0.97
Weight decay (weight_decay)	0.0004569
Learning rate decay factor (Ir_factor)	0.1
Penalty coefficient (λ)	0.0005

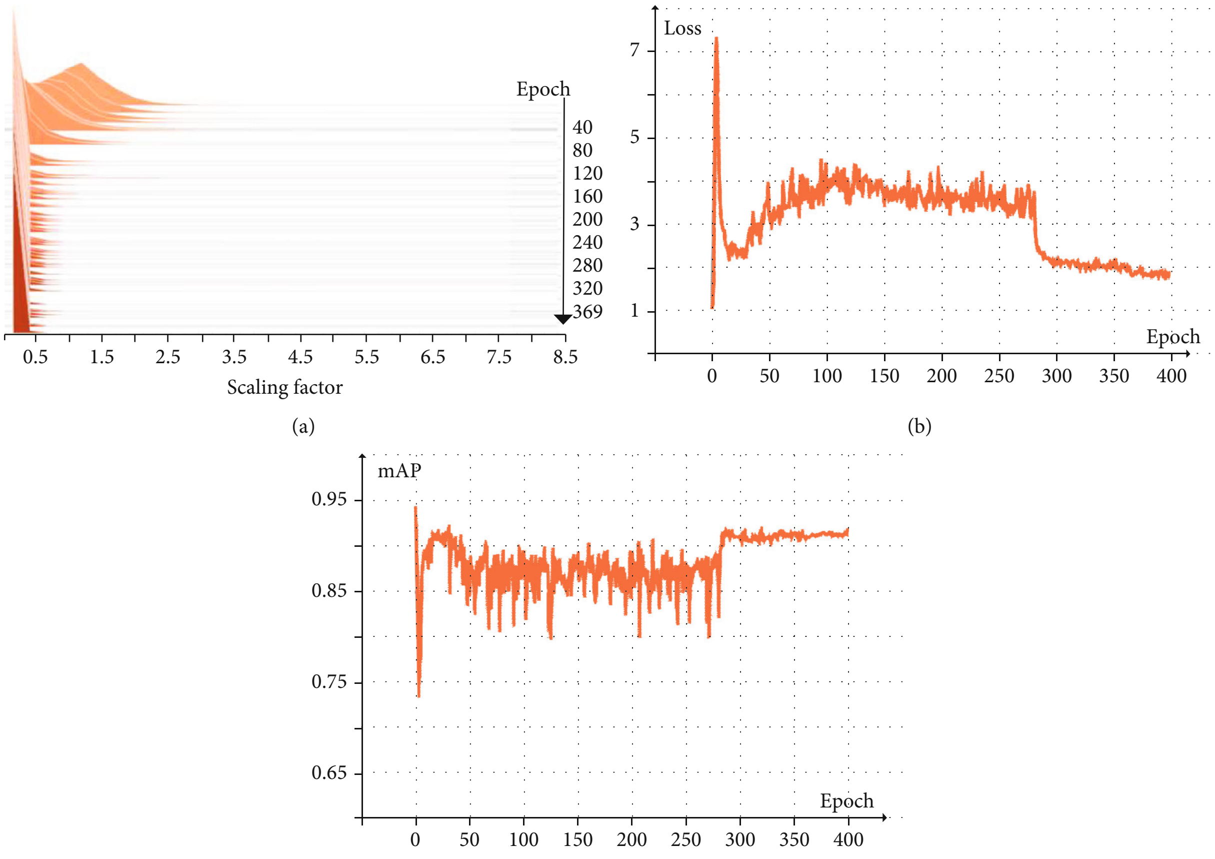
(1) *Sparse Training*. It is a game process between accuracy and sparsity of the model referring to the loss function during the model sparse training process in Section 3.2. If the penalty coefficient of the scaling factor is too high, the network is with fast sparsity but the accuracy drops fast too; else, the model's accuracy loss is low but with very slow sparsity. Therefore, choosing a suitable penalty coefficient is of great importance. Taking the experiment cases in previous studies, about 100 rounds of sparse training can reach maximum performance. The parameters applied in the sparse training experiments in this paper are shown in Table 3. A total of 400 rounds of training is proceeded to ensure sufficient time left after sparse training for further adjusting the model.

The sparse training process is shown in Figure 10(a). Large compression of the model is completed in about the first 100 rounds, and the accuracy is fine-tuned and restored in the following 300 rounds. The loss and mAP curves of the model are shown in Figures 10(b) and 10(c), in which while the scaling factor of the BN layer is in the period of substantial compression (20-100 epochs), the loss value of the model increases continuously and then goes to the adjustment period. The loss value decreases rapidly in the 280 rounds, and the mAP of the model rebounds significantly; the learning rate is reduced in the last 120 rounds, and the model accuracy is repaired.

The performances of the model before and after sparse training are compared in Table 4. The model accuracy after sparse training is 2% less than the original YOLOv4 model.

(2) *Channel Pruning and Layer Pruning*. The pruning ratios of the experiments in this paper are set to 0.8, 0.9, and 0.95; the ratio of channels to be reserved on each layer is 0.01, corresponding to the global threshold $\tilde{\gamma}$ and local safety threshold π in Section 3.3. To get the best pruning parameters, three groups of comparative experiments are designed for channel pruning of the model: YOLOv4-0.8, YOLOv4-0.9, and YOLOV4-0.9, and the performance of the pruning model is evaluated based on five kinds of indices: model size (model_size), mean accuracy (mAP), floating point of operations (FLOPs), parameters (params), and inference speed (inference). The model detection performances with different pruning ratios are shown in Table 5, in which, with low precision loss, the YOLOv4-0.9 model compresses its volume smaller and reduces the volume of floating point operations. So, the YOLOv4-0.9 model is chosen as the benchmark model for the layer pruning experiment.

Channel pruning can reduce the volume of model parameter calculation and improves nothing on inference speed. Therefore, layer pruning on shortcut structures of the backbone network is required after channel pruning. In this paper, layer pruning on the best model YOLOv4-0.9 chosen from channel pruning experiments is proceeded.



(a)

(b)

(c)

Figure 10: (a) Distribution of scaling factor in sparse training process; (b) loss curve in sparse training process; (c) mAP curve in sparse training process.

Table 4: Comparison of model performances before and after sparse training.

Model	Model_size (MB)	mAP (%)
YOLOv4	244 MB	93.4%
YOLOv4 after sparse training	244 MB	91.7%

Table 5: Comparison of the channel pruning experiments.

Experiment	Model_size (MB)	mAP (%)	FLOPs (G)	Params (M)	Inference (ms)
YOLOv4	235	94.9	59.80	63.96	33.9
YOLOv4-0.8	14.00	91.64	18.12	3.67	33.6
YOLOv4-0.9	4.46	91.40	7.64	1.16	33.8
YOLOv4-0.95	1.73	1.41	3.48	0.45	33.5

Layer pruning coefficient S is set as 8, 16, 20, 22, and 24, separately, and the five kinds of indices for model performance evaluation in channel pruning experiments are also applied in channel pruning. The model performances with different layer pruning ratios are shown in Table 6.

After the layer pruning on the YOLOv4-0.9 model, the model accuracy always decreases to a low level, so the original data set needs to be used and fine-tuned to rebound the lost accuracy. YOLOv4-0.9-20 is finally chosen for fine-tuning the training. The mean detection accuracy rebounds from 84.2% to 92.8%. And after the fine-tuning, the final pruning model CLSlim-YOLOv4 is obtained.

4.4. Evaluation of Pruning Model. To verify the effectiveness of the pruning method, YOLOv4 and YOLOv4-Tiny are both improved with lightweight CLSlim in this paper. The corresponding results are shown in Table 7.

Table 7 compares the model size, mean accuracy, and other indices of YOLOv4, YOLOv4-Tiny, and pruning model CLSlim-YOLOv4 and CLSlim-YOLOv4-Tiny. The size of CLSlim-YOLOv4 is compressed to 4.15 M, 1.76% of YOLOv4's, with inference speed increasing 1.8 times and 1.9 times in GTX2080ti and RK3399pro, respectively, FLOPs decreasing 12.1% and model accuracy decreasing 2.1%. Applying the model pruning method in YOLOv4-Tiny, the model size is compressed to 25.6%, parameter volume compressed to 25.5%, FLOPs decreasing 57%, inference speed increasing 1.1 times of the two type devices, and model accuracy increasing 0.8%. Parts of the CLSlim-YOLOv4-Tiny detection results in RK3399pro are shown in Figure 11.

Table 6: Comparison of the layer pruning experiments.

Experiment	Model_size (MB)	mAP (%)	FLOPs (G)	Params (M)	Inference (ms)
YOLOv4-0.9	4.46	91.40	7.64	1.16	33.8
YOLOv4-0.9-8	4.44	91.4	7.64	1.16	29.0
YOLOv4-0.9-16	4.34	89.4	7.52	1.13	23.0
YOLOv4-0.9-20	4.15	84.2	7.26	1.08	20.4
YOLOv4-0.9-22	4.06	65.1	6.78	1.05	19.5
YOLOv4-0.9-24	4.04	32.2	6.20	1.05	18.7

Table 7: Comparison of the model pruning experiments.

Model	Model_size (MB)	mAP (%)	FLOPs (G)	Params (M)	Inference (ms)	
					GTX2080ti	RK3399pro
YOLOv4	235	94.9	59.80	63.96	33.9	620
YOLOv4-Tiny	23.1	93.8	6.92	6.07	7.1	39
CLSlim-YOLOv4	4.15	92.8	7.26	1.08	18.7	320
CLSlim-YOLOv4-Tiny	5.92	94.6	4.00	1.55	6.5	33

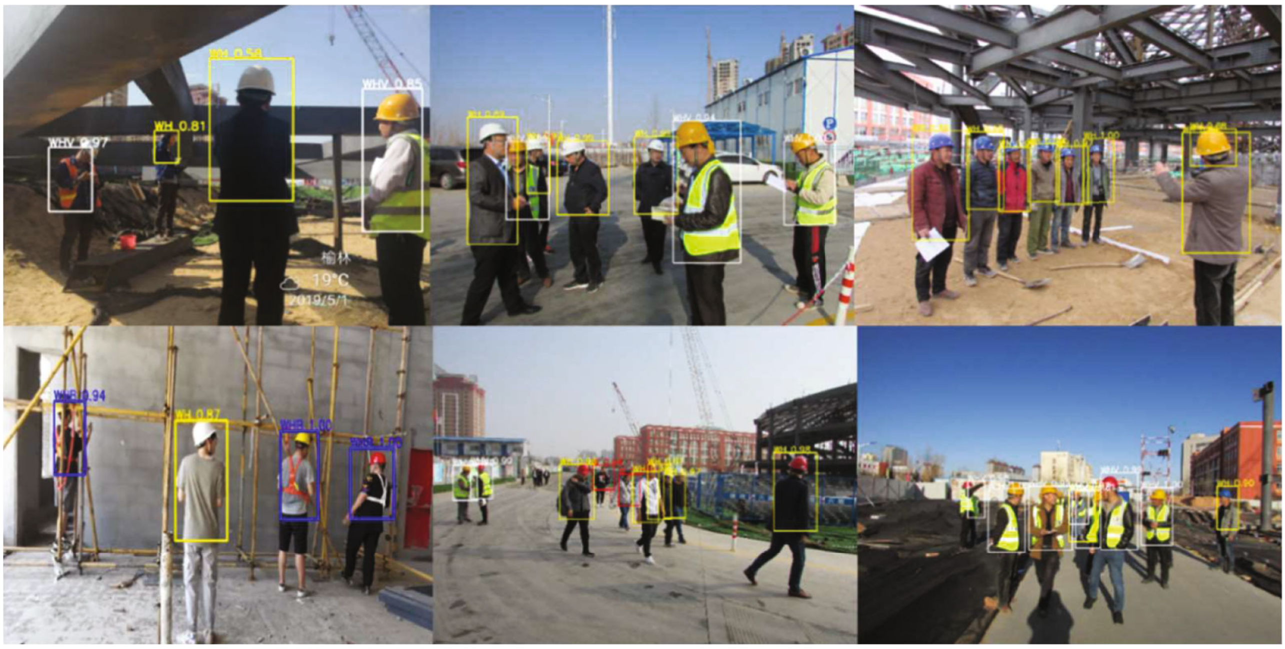


Figure 11: Recognition results of CLSlim-YOLOv4-Tiny in RK3399pro.

5. Conclusions

A combined detection algorithm on personal protective equipment for mobile terminals is proposed to check whether it is worn properly. The algorithm is realized based on the YOLOv4 network. Firstly, this algorithm applies L1 regularization and gradient sparse training on the scaling factor of the BN layer in the convolutional module of YOLOv4, and a global pruning threshold is settled to eliminate channel redundant parameters; at the same time, layer pruning thresholds are set to maintain the network structure integrity. After channel pruning, model size and parameter calculation volume decrease significantly. Then, the mean values of the scaling factors of each layer of the backbone network are sorted. Combining the layer pruning coefficient, several layers with small mean values of scaling factors are pruned, and the inference speed is improved. Afterwards, the pruning model CLSlim-YOLOv4 is obtained after 2-3 rounds of fine-tuning. To verify the effectiveness of the pruning method in this paper, with the same data set and test environment, the lightweight CLSlim method is imported into YOLOv4 and YOLOv4-Tiny. The test results show that with the premise of greatly reducing the parameter calculation volume and improving the inference speed, the accuracy losses of CLSlim-YOLOv4 are only 1%-2%; compared to YOLOv4-Tiny, CLSlim-YOLOv4-Tiny performs better in detection accuracy, parameter calculation, and inference speed. There might be false and missed detection in the real-life test of this research. Data sets can be expanded and enriched to improve the fitting ability of the model in afterwards research.

The combined detection algorithm on PPE in this paper can detect several kinds of PPEs at the same time, and ensuring the strong feature extraction ability of complex models, the model lightweight improvement is made to maintain high accuracy even with substantial parameter compression. This method satisfies the real-time ability and accuracy requirements of the combined detection of PPE in the real construction environment. The follow-up research will continue to combine the ones with other model lightweight strategies to improve the model inference speed and find model lightweight methods more suitable for source-limited mobile terminals.

Data Availability

The (PPE combined detection) data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

Acknowledgments

This research was funded by the National Natural Science Foundation of China (52104215) and the Key Industry Innovation Chain Project of Shaanxi Key Research and Development Plan (2021ZDLGY07-08).

References

[1] Osha, "Personal protective equipment," 2022, https://www.osha.gov/personal-protective-equipment.

[2] Osha, "Fall protection," 2022, https://www.osha.gov/fall-protection.

[3] Osha, "Occupational safety and health administration," 2022, https://www.osha.gov/laws-regs/standardinterpretations/2002-03-11.

[4] Y. Han, J. J. Zhang, H. Sun, J. Y. Yao, and S. D. You, "Design and implementation of intelligent safety inspection system for construction workers based on image recognition," *Journal of Safety Science and Technology*, vol. 12, no. 10, pp. 142–148, 2016.

[5] K. Shrestha, P. P. Shrestha, D. Bajracharya, and E. A. Yfantis, "Hard-hat detection for construction safety visualization," *Journal of Construction Engineering*, vol. 2015, 8 pages, 2015.

[6] B. Balakreshnan, G. Richards, G. Nanda, H. Mao, R. Athinarayanan, and J. Zaccaria, "PPE compliance detection using artificial intelligence in learning factories," *Procedia Manufacturing*, vol. 45, pp. 277–282, 2020.

[7] J. Wu, N. Cai, W. Chen, H. Wang, and G. Wang, "Automatic detection of hardhats worn by construction personnel: a deep learning approach and benchmark dataset," *Automation in Construction*, vol. 106, p. 102894, 2019.

[8] Z. Xie, H. Liu, Z. Li, and Y. He, "A convolutional neural network based approach towards real-time hard hat detection," in *2018 IEEE International Conference on Progress in Informatics and Computing (PIC)*, pp. 430–434, Suzhou, China, Dec 2018.

[9] N. D. Nath, A. H. Behzadan, and S. G. Paal, "Deep learning for site safety: real-time detection of personal protective equipment," *Automation in Construction*, vol. 112, p. 103085, 2020.

[10] G. Han, M. Zhu, X. Zhao, and H. Gao, "Method based on the cross-layer attention mechanism and multiscale perception for safety helmet-wearing detection," *Computers & Electrical Engineering*, vol. 95, article 107458, 2021.

[11] H. Wang, F. Lu, X. Tong, X. Gao, L. Wang, and Z. Liao, "A model for detecting safety hazards in key electrical sites based on hybrid attention mechanisms and lightweight Mobilenet," *Energy Reports*, vol. 7, pp. 716–724, 2021.

[12] A. Bochkovskiy, C. Y. Wang, and H. Liao, "*YOLOv4: optimal speed and accuracy of object detection*," 2020, http://arxiv.org/abs/2004.10934.

[13] P. Zhang, Y. Zhong, and X. Li, "Slim YOLOv3: narrower, faster and better for real-time UAV applications," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, Seoul, Korea (South), Oct. 2019.

[14] F. U. Huitong, W. A. N. G. Peng, L. I. Xiaoyan, L. U. Zhigang, and D. I. Ruohai, "Lightweight network model for moving object recognition," *Journal Of Xi'an Jiaotong University*, vol. 7, 2021.

[15] J. Yin, G. Liang, W. Jiang, S. Hong, and J. Yang, "ShuffleNet-inspired lightweight neural network design for automatic modulation classification methods in ubiquitous IoT cyber-physical systems," *Computer Communications*, vol. 176, pp. 249–257, 2021.

[16] A. P. Fard and M. H. Mahoor, "Facial landmark points detection using knowledge distillation-based neural networks," *Computer Vision and Image Understanding*, vol. 215, article 103316, pp. 1077–3142, 2022.

[17] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: a new backbone that can enhance learning capability of CNN," in *2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*, Seattle, WA, USA, June 2020.

[18] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA, July 2017.

[19] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, June 2018.

[20] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, SLille, France, July 2015.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 7896380, 8 pages
https://doi.org/10.1155/2022/7896380

Research Article

Influential Spreader Identification in Complex Networks Based on Network Connectivity and Efficiency

Rong Qiang [1] **and Jianshe Yang** [2]

[1] *School of Computer and Information Science, Southwest University, Chongqing 400715, China*
[2] *Basic Medical School, Gansu Medical College, Pingliang 744000, China*

Correspondence should be addressed to Jianshe Yang; yangjs@impcas.ac.cn

Received 15 February 2022; Accepted 28 March 2022; Published 13 April 2022

Academic Editor: Maode Ma

Influential spreader identification is a vital research area in complex network theory, which has important influence on application and popularization. Each of the existing methods has its own advantages and disadvantages, and there are still various methods proposed to solve this issue. In this paper, we come up with a new centrality of influential spreader identification based on network connectivity and efficiency (CEC). The consequences of spreader deletion can be generally divided into two parts, one is that the connectivity of network topology is destroyed, and the other is that network's performance is degraded, which makes the network unable to meet the functional requirement. Therefore, the relative changes of connectivity and efficiency of network before and after removing spreaders are used to present the influence of spreaders. We adopt susceptible-infected (SI) model, a well-known infectious disease model, to verify the effectiveness of CEC through the spreading ability simulation of spreaders in actual networks. And the simulation results demonstrate the superiority of CEC.

1. Introduction

At present, complex networks are closely associated with our real life, for example, networks [1, 2], traffic systems [3, 4], power grids [5, 6], and ecological networks [7, 8]. Influential spreader identification remains an open and vital research issue that has attracted increasing attention, which helps to understand the structure of networks and control the propagation process. Some hazards caused by load propagation and cascading effect, for example, North American blackout and WannaCry's spread, often begin with a small portion of spreaders but spread rapidly to the entire network [9, 10]; this small portion of spreaders has a great impact on network. Therefore, accurate quantification and identification of influential spreaders is very important. For instance, we can effectively suppress the spread of the virus and prevent its large-scale outbreak by vaccinating key individuals in infectious disease network [11]. In power grids, we can effectively prevent the cascading failure with taking prior precautions for circuits in vital areas [12]. In social network, such as MicroBlog and Twitter, we can control the dissemination of information to guide speech [13].

A variety of approaches have been proposed over the past few decades. Most of these approaches measure the influence of spreaders from the structural information of network.

There are a lot of methods proposed to search these key spreaders [14]. Degree centrality (DC) [15], one of the simplest and earliest methods, only counts the number of the directly connected spreaders and results in low complexity. Closeness centrality (CC) [16] measures spreader's capability to affect others through the network, while it will fail when applied to disconnected networks. Betweenness centrality (BC) [17] measures spreader's influence with the perspective from shortest path. Except these classical measures, some new methods have been proposed such as H-index centrality [18] and evidence theory [19]. Roberts et al. [20] suggested a centrality which considered the fourth-level neighbors as a trade-off measure. However, these centralities ignore the connections among spreaders, and then, the ClusterRank [21] was proposed by taking the effect of clustering coefficient into consideration. What is more, Kitsak et al. [22] measured the influence of spreaders from location perspective and put forward a new method named K-shell decomposition (Ks). The spreaders were moved layer by layer based on continuously updated DC value. The biggest problem of Ks is the poor distinguish capacity of centrality value, i.e., poor monotonicity. Then, some approaches were put forward to solve this issue. Zeng and Zhang [23] came up with a MDD approach by considering the degree of initial spreaders and removed spreaders. Bae and Kim [24] summed the Ks value of neighbors to measure the importance of spreaders. In addition, there are also some approaches based on iteration such as PageRank [25], LeaderRank [26], and Hits [27]. Different centralities reflect the influence of spreaders from limited parts; some researchers have proposed multiattribute ranking approaches which combines several centralities to comprehensively rank the influence of spreaders. Liu et al. [28] proposed an improved Ks and used TOPSIS to fuse DC, CC, and BC and improved K-shell decomposition. Yang et al. [29] combined DC, CC, and BC with VIKOR method and adopted entropy weighting method to reasonably obtain the weights of attributes. Wen and Deng proposed a local information dimensionality (LD) to rank key spreaders [30]. Wang et al. focused on the contribution of spreaders to network efficiency and proposed EffC method to identify influential spreaders [31].

In this paper, we consider the importance of spreaders from global information perspective, and then, a novel centrality called connectivity and efficiency centrality (CEC) is put forward. The consequences of network spreaders removal can be generally divided into two aspects [32, 33], one is that the connectivity of network topology is destroyed, and the other is that the performance of the network is degraded, which makes the network unable to meet the service requirement. Therefore, we consider the relative changes of connectivity and efficiency of network before and after removing spreaders, and the combination of them is taken as an indicator to determine the influence of spreaders. Note that the removal of spreaders will also delete the links connected to them at the same time. To assess the effectiveness of CEC, we adopt susceptible-infected (SI) model [34] to measure spreading ability of spreaders in actual databases, and we compare the performance between CEC and others to verify the superiority of CEC.

2. Centralities

Given a network $G(V, E)$, where V and E, respectively, represent the set of spreaders and the set of edges, they meet $m = |E|$ and $n = |V|$. $A = (a_{ij})_{n\times n}$ indicates the adjacent matrix; if spreader i and spreader j are connected by edge $e_{ij} \in E$, $a_{ij} = 1$; otherwise, $a_{ij} = 0$.

Degree centrality [15], one of the simplest and earliest local centrality, only counts the number of the directly connected spreaders and results in low complexity.

$$\mathrm{DC}(i) = \sum_{j}^{n} a_{ij}. \tag{1}$$

Degree centrality indicates spreaders' ability to communicate directly with others.

Closeness centrality [16] considers the influence of spreaders based on the distance between them. It measures spreader's capability to affect others through the network.

$$\mathrm{CC}(i) = \frac{n-1}{\sum_{j}^{n} d_{ij}}, \tag{2}$$

wherein d_{ij} represents the Euclidean distance between spreader i and spreader j. CC uses average transmission time of information to determine the influence of a spreader.

Betweenness centrality [17] measures spreader's influence with the perspective from shortest path. BC considers a spreader influential if it expressed as a "bridge."

$$\mathrm{BC}(i) = \frac{1}{(n-1)(n-2)}\left(\sum_{s,t\neq i} \frac{\mathrm{g}_{\mathrm{st}}(i)}{\mathrm{g}_{\mathrm{st}}}\right), \tag{3}$$

wherein g_{st} represents the number of the shortest paths between spreader s and spreader t and $\mathrm{g}_{\mathrm{st}}(i)$ indicates the number of shortest paths passing through spreader i. BC can reflect the degree of independence between spreaders.

K-shell decomposition [22] measures the influence of spreaders from location perspective, which has important milestone significance. The spreaders were moved layer by layer based on continuously updated DC value.

3. The Proposed Centrality

We consider the influence of spreaders from global information perspective. The influence of spreaders can be measured by the relative changes of some global characteristic parameters of network before and after removing corresponding spreaders. The consequences of network spreader deletion can be generally divided into two parts, one is that the connectivity of network topology is destroyed, and the other is that the network efficiency is degraded, which makes the network unable to meet the service requirement. Both the two aspects should be taken into consideration to give comprehensive identification results.

Definition 1. The network connectivity represents the average influence of network to maintain connectivity, which is indicated as the mean value of the ratio of number of connected spreader pairs to the total number of spreader pairs in network.

$$\mathrm{CON}(G) = \frac{1}{n(n-1)} \sum_{i\in G} \sum_{j>\mathrm{i}} l_{ij}, \tag{4}$$

wherein l_{ij} represents the connection parameter from

Table 1: The top 10 spreaders using different centralities: karate club.

Karate club						
Rank	DC	CC	BC	Ks	CEC	I(t)
1	34	1	1	1	1	1
2	1	3	34	34	34	34
3	33	34	33	33	3	3
4	3	32	3	31	33	33
5	2	9	32	3	32	2
6	4	14	9	14	2	9
7	32	33	2	2	14	14
8	9	20	14	4	9	32
9	14	2	20	8	4	4
10	24	4	6	9	20	20

spreader i to spreader j; if they have a connected path, including directly connected path and indirectly connected path, then $l_{ij} = 1$; otherwise, $l_{ij} = 0$.

Definition 2. The residual network is denoted as G_k after removing spreader k from G, and the relative changes of network connectivity can be defined as

$$\text{RE\_CON}(k) = \frac{|\text{CON}(G_k) - \text{CON}(G)|}{\text{CON}(G)}. \tag{5}$$

Definition 3. The network efficiency refers to the effectiveness of information transmission on the network. It is denoted as

$$\text{EFF}(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}}, \tag{6}$$

wherein d_{ij} refers to the shortest distance between spreader i and spreader j. Note that if spreader i and spreader j have no connected path, $d_{ij} = +\infty$ and $1/\mathrm{d}_{\mathrm{ij}} = 0$.

Definition 4. The residual network is denoted as G_k after removing spreader k from G, and the relative changes of network efficiency can be written as

$$\text{RE\_EFF}(k) = \frac{|\text{EFF}(G_k) - \text{EFF}(G)|}{\text{EFF}(G)}. \tag{7}$$

Definition 5. The proposed connectivity and efficiency centrality (CEC) can be defined as

$$\text{CEC}(k) = (1 + \text{RE\_CON}(k)) \times \text{RE\_EFF}(k). \tag{8}$$

The greater the value of $\text{CEC}(k)$, the more influential the spreader k.

Table 2: The top 10 spreaders using different centralities: Jazz musicians.

Jazz musicians						
Rank	DC	CC	BC	Ks	CEC	I(t)
1	136	136	136	60	136	136
2	60	60	153	168	149	149
3	132	168	60	108	60	96
4	168	70	149	122	5	70
5	70	83	168	33	153	60
6	99	132	167	58	185	153
7	108	122	189	66	189	167
8	83	194	115	100	168	5
9	158	174	96	132	34	83
10	7	158	83	179	83	168

Table 3: The top 10 spreaders using different centralities: USAir97.

USAir97						
Rank	DC	CC	BC	Ks	CEC	I(t)
1	118	118	118	67	8	118
2	261	261	8	94	261	67
3	255	67	261	109	118	261
4	152	255	201	112	13	8
5	182	201	47	118	201	182
6	230	182	182	131	152	313
7	166	47	255	146	182	201
8	67	166	152	147	313	152
9	112	248	313	150	67	248
10	201	112	13	152	258	258

Table 4: The top 10 spreaders using different centralities: email.

Email						
Rank	DC	CC	BC	Ks	CEC	I(t)
1	105	333	333	299	333	105
2	333	23	105	389	58	333
3	16	105	23	434	355	135
4	23	42	578	552	578	42
5	42	41	76	571	105	3
6	41	76	233	726	21	52
7	196	233	135	756	270	21
8	233	52	41	788	376	355
9	21	135	355	885	42	233
10	76	378	42	886	233	270

4. Simulation and Analysis

4.1. Datasets. We choose four actual networks to conduct experiments and simulations, which cover multiple fields and network scales. (i) Karate club [35]: it is a widely used dataset describing the relationship between karate club

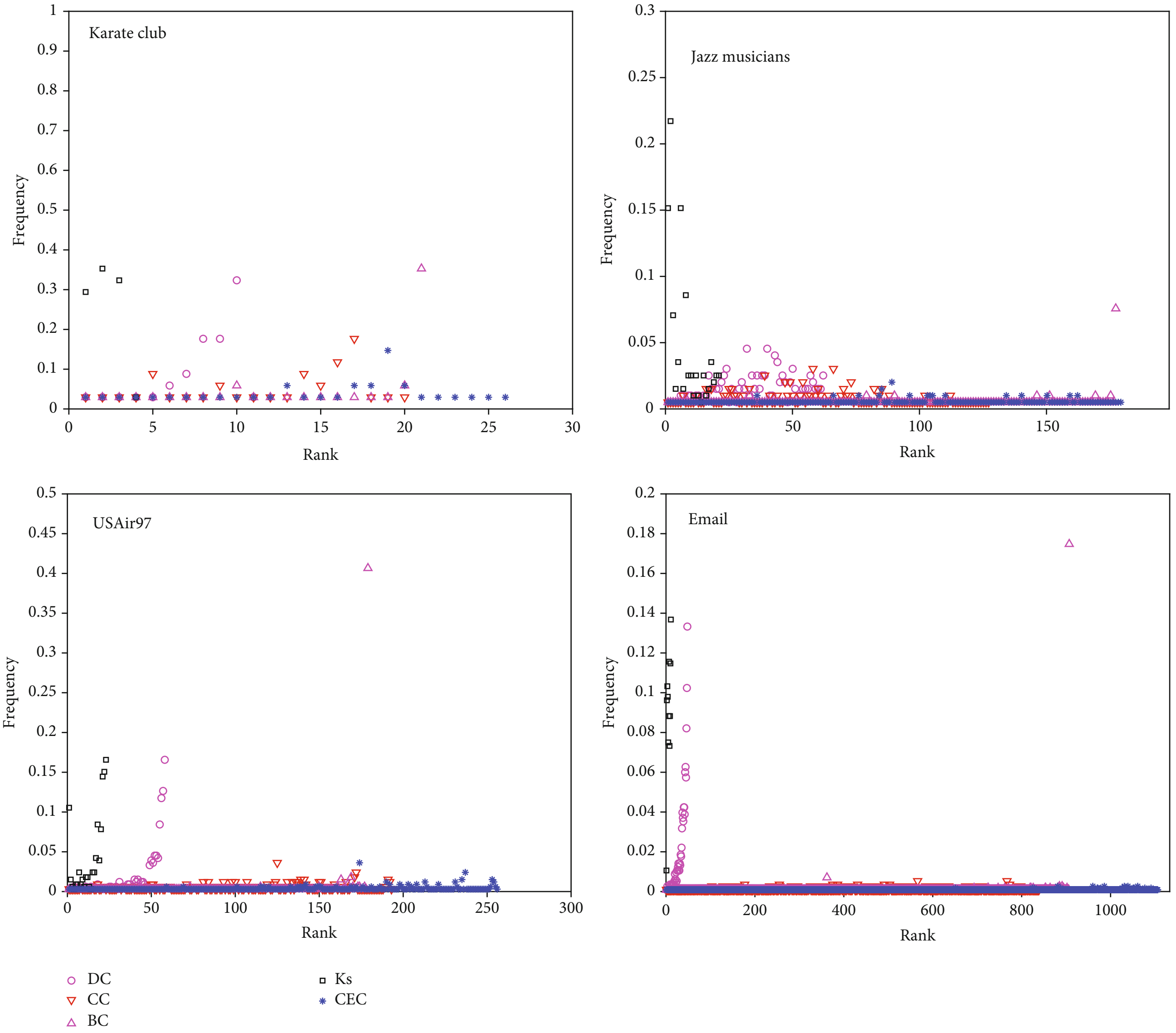


Figure 1: Example of a figure caption.

members. (ii) Jazz musicians [36]: it is a social dataset describing the cooperative relationship between jazz musicians. (iii) USAir97 [37]: it is a transportation dataset representing the airline relationship of American airports in 1997. (iv) Email: it describes the email exchange in a university.

4.2. Experiment and Analysis

4.2.1. Experiment 1: Comparison of Top 10 Spreaders Ranked by Different Centralities. The influence of each spreader in network is calculated using CEC and classical centralities. The actual spreading ability I(t) ($t = 25$) calculated by SI model is used as benchmark; the definition of I(t) will be introduced later. We pay attention to the top 10 spreaders sorted by several centralities. As shown in Table 1, in karate club network, the identification results of CEC and CC are the best due to their 10 same spreaders as I(t), and DC and BC have 9 same spreaders as I(t), while Ks owns 8 same spreaders. In Jazz musician network, shown in Table 2, there are 5 same spreaders with I(t) in top 10 lists using DC and CC, while it is only 2 using Ks. CEC owned 7 same spreaders as I(t) performs slightly worse than BC. Besides, the top 2 spreaders of CEC are the same with I(t). In USAir97 network (Table 3), DC and CC both own 6 same spreaders; the number of same spreaders of CEC, BC, and Ks is 9, 7, and 3, separately. In email network, depicted in Table 4, CEC, CC, and BC have 6 same spreaders as I(t), which is lightly greater than DC, while there is no any same spreader between Ks and I(t). In a word, CEC has the most similar performance with actual ranking results; that is, CEC can identify spreaders more accurately.

4.2.2. Experiment 2: Comparison of Capability of Different Centralities to Distinguish Spreaders' Spreading Ability. When ranking the influence of spreaders, we find that some spreaders have the same centrality value and it is impossible to distinguish them. This phenomenon will reduce the accuracy of centrality. We consider the frequency of spreaders

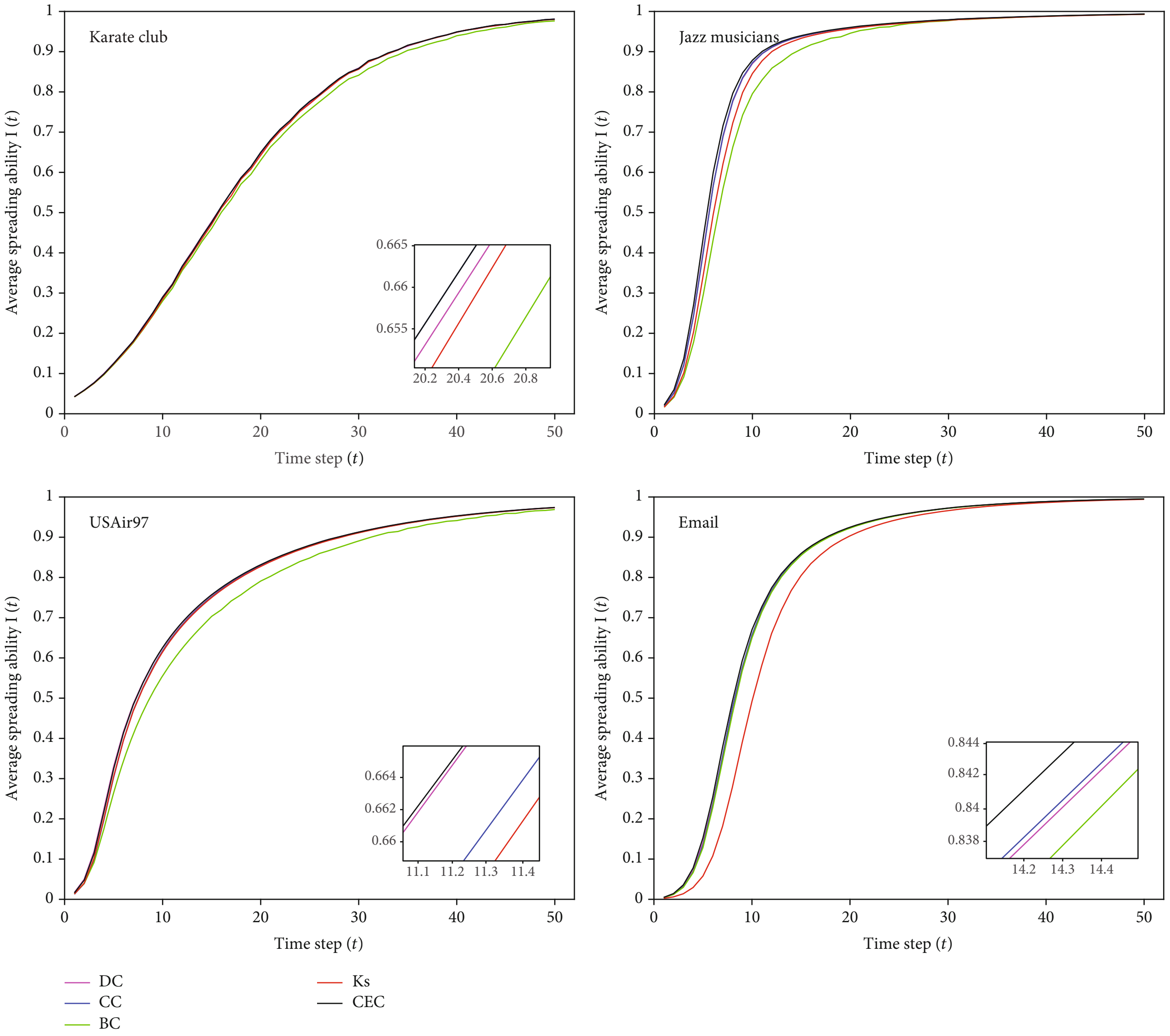


Figure 2: The average spreading ability of top 10 spreaders sorted with different centralities.

with same rank as an index to assess the distinguishing capability. The lower the frequency, the better the method. The experimental results of different centralities are shown in Figure 1. In the four networks, CEC has the lowest frequency; that is to say, CEC performs best in distinguishing spreaders' spreading ability. However, the frequency of DC and Ks is greater than other methods. The experimental result indicates the superiority of our method in distinguishing spreading ability.

4.2.3. Experiment 3: Comparison of the Average Spreading Ability of Top 10 Spreaders. We conduct transmission simulation with SI model [34] to examine the spreading ability of spreaders. We take spreader i as the source spreader and the spread process will start from the source spreader. The total number of infected spreaders will reach n_{it} after $t(t = 1, 2, \cdots)$ time step. Then, the spreading ability, denoted as $I_i(t) = n_{it}/n$, is expressed as ratio of infected spreaders to network size. And the average spreading ability of top 10 spreaders is represented as $I(t) = (\sum_{i=1}^{10} I_t(\mathrm{t}))/10$. We set $t_{\max} = 50$; the simulation results are presented in Figure 2.

From Figure 2, the average spreading ability of top 10 spreaders increases with t, and eventually almost the entire network is infected. In karate club network, we can see that the black curve and the blue curve overlap; that is to say, the spreading ability of CEC is the same with CC, because the top 10 spreaders of them are the same. It is clear that the spreading ability of CEC is superior to that of DC, BC, and Ks. In Jazz musician network, we can find that there are more infected spreaders of CEC than others, which demonstrates that the spreading ability of CEC is better than that of other methods. In USAir97 network, CEC is marginally better than CC, DC, and Ks, and BC is the poorest because the average number of infected spreaders of BC is much less than that of others. In email network, the number of infected spreaders at each step of CEC is marginally greater than DC, CC, and BC.

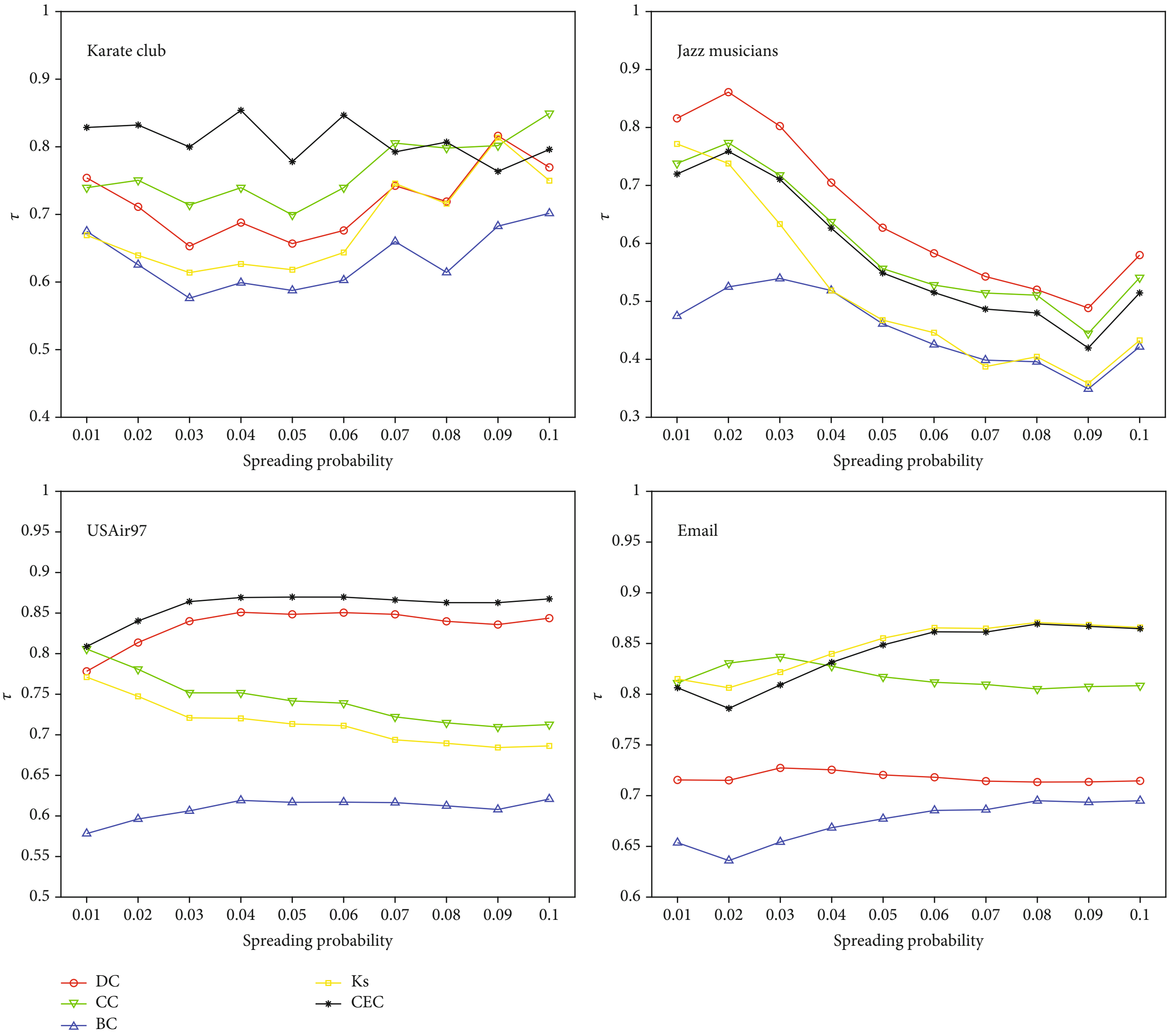


Figure 3: Comparison of the correlation between centralities and the actual ranking list.

4.2.4. Experiment 4: Comparison of the Correlation between Centralities and the Actual Ranking Result. We choose Kendall's tau coefficient (τ) [36] to be a linear correlation coefficient between the five methods and the actual ranking result. The value of τ ranges between [0, 1]; the larger the value of τ is, the more similar two sequences is. Give two sequences $X = \{x_1, x_2, \cdots, x_s\}$ and $Y = \{y_1, y_2, \cdots, y_s\}$. (x_i, y_i) is regarded as a positive sequence pair when $x_i > x_j$ and $\mathrm{y_i} > y_j$, or $x_i < x_j$ and $y_i < y_j$, or else it will be considered as a negative sequence pair. Then, Kendall's tau can be denoted as $\tau = (n_+ - n_- / n(n-1))$, where n_+ and n_- indicate the number of positive sequence pairs and negative sequence pairs, respectively, and $n = n_+ + n_-$.

We consider the ranking list at $t = 10$ obtained by SI model as the actual ranking result Ia; then, we calculate the correlation between Ia and centralities. As shown in Figure 3, CEC outweighs other centralities before spreading probability 0.07 in karate club network, and it is lower than CC after spreading probability 0.08. In Jazz musician network, DC has the greatest τ value, while it has very poor performance in email network, and the τ value of CEC is similar with CC. In USAir97 network, CEC outweighs other centralities across the spreading probability. In email network, the τ value of CEC is lower than CC and Ks before spreading probability 0.04, and it is similar with Ks after spreading probability 0.04. Overall speaking, CEC has the best correlation with actual ranking result in the four networks.

5. Conclusion

Identifying influential spreaders is essential for network invulnerability. In this paper, we pay attention to the approach of identifying influential spreaders based on global information, and the connectivity and efficiency centrality (CEC) are put forward to achieve this goal. Removing spreaders and the corresponding links will lead to two consequences: the destruction of network connectivity and the decline of network efficiency. Therefore, we consider both the two aspects to provide a novel centrality in identifying influential spreaders. The relative changes of network connectivity and efficiency before and after removing spreaders are taken as indicators to measure the influence of spreaders; we combine the relative changes of network connectivity and efficiency to give comprehensive identifying results. The greater the relative changes, the more influential the spreader. We conduct several experiments based on actual datasets, and the results show that CEC performs better than other methods.

Data Availability

All data are available in the manuscript references, which can be accessed at Pubmed, google scholar and other web resources.

Conflicts of Interest

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

References

[1] J. Kim and M. Hastak, “Social network analysis: characteristics of online social networks after a disaster,” *International Journal of Information Management*, vol. 38, no. 1, pp. 86–96, 2018.

[2] K. Q. Zhang, H. F. Du, and M. W. Feldman, “Maximizing influence in a social network: improved results using a genetic algorithm,” *Physica A: Statistical Mechanics and its Applications*, vol. 478, pp. 20–30, 2017.

[3] R. Ding, N. Ujang, H. Hamid et al., “Detecting the urban traffic network structure dynamics through the growth and analysis of multi-layer networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 503, pp. 800–817, 2018.

[4] M. Y. Zhang, B. Y. Liang, S. Wang, M. Perc, W. B. Du, and X. B. Cao, “Analysis of flight conflicts in the Chinese air route network,” *Chaos Solitons & Fractals*, vol. 112, pp. 97–102, 2018.

[5] Y. Yang, T. Nishikawa, and A. E. Motter, “Small vulnerable sets determine large network cascades in power grids,” *Science*, vol. 358, no. 6365, p. eaan3184, 2017.

[6] B. Schäfer, D. Witthaut, M. Timme, and V. Latora, “Dynamically induced cascading failures in power grids,” *Nature Communications*, vol. 9, no. 1, p. 1975, 2018.

[7] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi, “The multilayer nature of ecological networks,” *Nature Ecology & Evolution*, vol. 1, no. 4, 2017.

[8] G. Strona and K. D. Lafferty, “Environmental change makes robust ecological networks fragile,” *Nature Communications*, vol. 7, no. 1, 2016.

[9] M. Zhao, T. Zhou, B. H. Wang, and W. X. Wang, “Enhanced synchronizability by structural perturbations,” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 72, no. 5, article 057102, 2005.

[10] Y. Z. Yang, L. Yu, Z. L. Zhou, Y. Chen, and T. Kou, “Node importance ranking in complex networks based on multicriteria decision making,” *Mathematical Problems in Engineering*, vol. 2019, Article ID 9728742, 12 pages, 2019.

[11] M. J. Alvarez, Y. Shen, F. M. Giorgi et al., “Functional characterization of somatic mutations in cancer using network-based inference of protein activity,” *Nature Genetics*, vol. 48, no. 8, pp. 838–847, 2016.

[12] R. Albert, I. Albert, and G. L. Nakarado, “Structural vulnerability of the North American power grid,” *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 2, article 025103, 2004.

[13] A. Sheikhahmadi, M. A. Nematbakhsh, and A. Shokrollahi, “Improving detection of influential nodes in complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 436, pp. 833–845, 2015.

[14] L. Y. Lü, D. B. Chen, X. L. Ren, Q. M. Zhang, Y. C. Zhang, and T. Zhou, “Vital nodes identification in complex networks,” *Physics Reports*, vol. 650, pp. 1–63, 2016.

[15] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” *Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.

[16] L. C. Freeman, “Centrality in social networks conceptual clarification,” *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[17] M. E. J. Newman, “A measure of betweenness centrality based on random walks,” *Social Networks*, vol. 27, no. 1, pp. 39–54, 2005.

[18] A. Zareie and A. Sheikhahmadi, “EHC: extended H-index centrality measure for identification of users’ spreading influence in complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 514, pp. 141–155, 2019.

[19] P. Hu, W. L. Fan, and S. W. Mei, “Identifying node importance based on evidence theory in complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 429, pp. 169–176, 2015.

[20] P. Roberts, D. Gaffney, J. Lee-Thorp, and G. Summerhayes, “Identifying influential nodes in complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 4, pp. 1777–1787, 2012.

[21] D. B. Chen, H. Gao, L. Y. Lü, and T. Zhou, “Identifying influential nodes in large-scale directed networks: the role of clustering,” *PLoS One*, vol. 8, no. 10, article e77455, 2013.

[22] M. Kitsak, L. K. Gallos, S. Havlin et al., “Identification of influential spreaders in complex networks,” *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[23] A. Zeng and C. J. Zhang, “Ranking spreaders by decomposing complex networks,” *Physics Letters A*, vol. 377, no. 14, pp. 1031–1035, 2013.

[24] J. Bae and S. Kim, “Identifying and ranking influential spreaders in complex networks by neighborhood coreness,” *Physica A: Statistical Mechanics and its Applications*, vol. 395, pp. 549–559, 2014.

[25] A. Arasu, J. H. Cho, H. Garcia-Molina, A. Paepcke, and S. Raghavan, “Searching the web,” *ACM Transactions on Internet Technology (TOIT)*, vol. 1, no. 1, pp. 2–43, 2001.

[26] L. Lü, Y. C. Zhang, C. H. Yeung, and T. Zhou, “Leaders in social networks, the delicious case,” *PLoS One*, vol. 6, no. 6, 2011.

[27] J. M. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, vol. 98, pp. 668–677, 1998.

[28] Z. H. Liu, C. Jiang, J. Y. Wang, and H. Yu, “The node importance in actual complex networks based on a multi-attribute ranking method,” *Knowledge-Based Systems*, vol. 84, pp. 56–66, 2015.

[29] Y. Z. Yang, L. Yu, X. Wang, Z. L. Zhou, Y. Chen, and T. Kou, "A novel method to evaluate node importance in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 526, 2019.

[30] T. Wen and Y. Deng, "Identification of influencers in complex networks by local information dimensionality," *Information Sciences*, vol. 512, pp. 549–562, 2020.

[31] S. S. Wang, Y. X. Du, and Y. Deng, "A new measure of identifying influential nodes: efficiency centrality," *Communications in Nonlinear Science and Numerical Simulation*, vol. 47, pp. 151–163, 2017.

[32] T. Zhou, J. G. Liu, W. J. Bai, G. Chen, and B. H. Wang, "Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 5, article 056109, 2006.

[33] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[34] M. G. Pablo and L. E. Danon, "Community structure in jazz," *Advances in Complex Systems*, vol. 6, no. 4, pp. 565–573, 2003.

[35] "North American Transportation Atlas Data (NORTAD)," https://www.bts.gov/archive/publications/north_american_transportation_atlas_data/index.

[36] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, "Self-similar community structure in a network of human interactions," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 6, 2003.

[37] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1-2, pp. 81–93, 1938.