

# Artificial Intelligence in E-Healthcare and M-Healthcare

Lead Guest Editor: Xingwang Li

Guest Editors: Lingwei Xu, Thomas Aaron Gulliver, and Han Wang





---

# **Artificial Intelligence in E-Healthcare and M-Healthcare**

Journal of Healthcare Engineering

---

## **Artificial Intelligence in E-Healthcare and M-Healthcare**

Lead Guest Editor: Xingwang Li

Guest Editors: Lingwei Xu, Thomas Aaron Gulliver,  
and Han Wang



---

Copyright © 2021 Hindawi Limited. All rights reserved.

This is a special issue published in "Journal of Healthcare Engineering." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Chief Editor

Loredana Zollo, Italy

## Editorial Board

Saverio Affatato, Italy  
Francesca Apollonio, Italy  
Hasan Ayaz, USA  
Óscar Belmonte Fernández, Spain  
Patrick Boissy, Canada  
Niccolò Camarlinghi, Italy  
Xiao-Jun Chen, China  
Hongwei Chen, USA  
Gin-Shin Chen, Taiwan  
Daniel H.K. Chow, Hong Kong  
Gianluca Ciardelli, Italy  
Elena De Momi, Italy  
Costantino Del Gaudio, Italy  
Ayush Dogra, India  
Daniel Espino, United Kingdom  
Mian Muhammand Sadiq Fareed, China  
Mostafa Fatemi, USA  
Jesus Favela, Mexico  
Jesus Fontecha, Spain  
Jean-Luc Gennisson, France  
Luca Giancardo, USA  
Antonio Gloria, Italy  
Kheng-Lim Goh, Singapore  
Pedro Gomes, Portugal  
Carlos Gómez, Spain  
Philippe Gorce, France  
Vincenzo Guarino, Italy  
Valentina Hartwig, Italy  
David Hewson, United Kingdom  
Ernesto Iadanza, Italy  
Cosimo Ieracitano, Italy  
Giovanni Improta, Italy  
Norio Iriguchi, Japan  
Mihajlo Jakovljevic, Japan  
Jingfeng Jiang, USA  
Zhongwei Jiang, Japan  
Rashed Karim, United Kingdom  
Pasi A. Karjalainen, Finland  
Chandan Karmakar, Australia  
John S. Katsanis, Greece  
Terry K.K. Koo, USA  
Jui-Yang Lai, Taiwan  
Xiang Li, USA  
Lun-De Liao, Singapore

Feng-Huei Lin, Taiwan  
Yuan-Pin Lin, Taiwan  
Qiu-Hua Lin, China  
Maria Lindén, Sweden  
Aiping Liu, China  
Dongfei Liu, Finland  
Francisco Lopez-Valdes, Spain  
Zufu Lu, Australia  
Ilias Maglogiannis, Greece  
Saverio Maietta, Italy  
Alessandro Mauro, Italy  
Mehran Moazen, United Kingdom  
Rafael Morales, Spain  
Jialin Peng, China  
Angkoon Phinyomark, Canada  
Vincenzo Positano, Italy  
Saeed Mian Qaisar, Saudi Arabia  
Alessandro Ramalli, Italy  
Alessandro Reali, Italy  
Lei Ren, United Kingdom  
Vito Ricotta, Italy  
Jose Joaquin Rieta, Spain  
Emanuele Rizzuto, Italy  
Sébastien Roth, France  
Simo Saarakkala, Finland  
Hélder A. Santos, Finland  
Emiliano Schena, Italy  
Maurizio Schmid, Italy  
Jiann-Shing Shieh, Taiwan  
Tiago H. Silva, Portugal  
Sharan Srinivas, USA  
Fabrizio Taffoni, Italy  
Redha Tair, France  
Jinshan Tang, USA  
Ioannis G. Tollis, Greece  
Kazunori Uemura, Japan  
Cesare F. Valenti, Italy  
Uche Wejinya, USA  
Ying Yang, United Kingdom  
Hongbo Zhang, Finland  
Haihong Zhang, Singapore  
Ping Zhou, USA

# Contents

## **Guest Editorial: Special Issue on Artificial Intelligence in E-Healthcare and M-Healthcare**

Xingwang Li , Lingwei Xu , Thomas Aaron Gulliver , and Han Wang 

Editorial (3 pages), Article ID 9857089, Volume 2021 (2021)

## **A High-Sensitivity Method Based on Advanced Optical Waveguide Technology to Detect HBsAg**

Pingping Xiao 

Research Article (8 pages), Article ID 8719147, Volume 2021 (2021)

## **Zero-Watermarking Algorithm for Medical Image Based on VGG19 Deep Convolution Neural Network**

Baoru Han , Jinglong Du , Yuanyuan Jia , and Huazheng Zhu 

Research Article (12 pages), Article ID 5551520, Volume 2021 (2021)

## **Service Migration Policy Optimization considering User Mobility for E-Healthcare Applications**

Xuhui Zhao , Jianghui Liu , Baofeng Ji , and Lin Wang 

Research Article (13 pages), Article ID 9922876, Volume 2021 (2021)

## **Clinical Effects of Form-Based Management of Forceps Delivery under Intelligent Medical Model**

Siming Xin , Zhizhong Wang , Hua Lai , Lingzhi Liu , Ting Shen , Fangping Xu , Xiaoming Zeng , and Jiusheng Zheng 

Research Article (8 pages), Article ID 9947255, Volume 2021 (2021)

## **ASC Performance Prediction for Medical IoT Communication Networks**

Fagen Yin , Pingping Xiao , and Zefeng Li 

Research Article (7 pages), Article ID 6265520, Volume 2021 (2021)

## **Efficient Algorithms for E-Healthcare to Solve Multiobject Fuse Detection Problem**

Ijaz Ahmad , Inam Ullah , Wali Ullah Khan, Ateeq Ur Rehman , Mohammed S. Adrees, Muhammad Qaiser Saleem, Omar Cheikhrouhou , Habib Hamam, and Muhammad Shafiq 

Research Article (16 pages), Article ID 9500304, Volume 2021 (2021)

## **Cross-Database Micro-Expression Recognition Exploiting Intradomain Structure**

Yanliang Zhang , Ying Liu , and Hao Wang 

Research Article (9 pages), Article ID 5511509, Volume 2021 (2021)

## **Predicting Mental Health Problems with Automatic Identification of Metaphors**

Nan Shi, Dongyu Zhang , Lulu Li, and Shengjun Xu

Research Article (11 pages), Article ID 5582714, Volume 2021 (2021)

## **Analysis on the Development Status of Coal Mine Dust Disaster Prevention Technology in China**

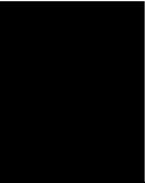
Hui Zhang, Weihai Han, Youlin Xu , and Zhenfeng Wang

Review Article (9 pages), Article ID 5574579, Volume 2021 (2021)

## **Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning**

Shujun Zhang , Shuhao Xu, Liwei Tan, Hongyan Wang , and Jianli Meng

Research Article (9 pages), Article ID 5524769, Volume 2021 (2021)



---

**An Improved Double Channel Long Short-Term Memory Model for Medical Text Classification**

Shengbin Liang, Xinan Chen , Jixin Ma , Wencai Du , and Huawei Ma 

Research Article (8 pages), Article ID 6664893, Volume 2021 (2021)

## Editorial

# Guest Editorial: Special Issue on Artificial Intelligence in E-Healthcare and M-Healthcare

Xingwang Li <sup>1</sup>, Lingwei Xu <sup>2</sup>, Thomas Aaron Gulliver <sup>3</sup> and Han Wang <sup>4,5</sup>

<sup>1</sup>School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

<sup>2</sup>Department of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China

<sup>3</sup>Department of Electrical and Computer Engineering, University of Victoria, Victoria BC V8W 2Y2, Canada

<sup>4</sup>Faculty of Data Science, City University of Macau, Macau 999078, China

<sup>5</sup>College of Physical Science & Engineering, Yichun University, Yichun 336000, China

Correspondence should be addressed to Xingwang Li; [lixingwang@hpu.edu.cn](mailto:lixingwang@hpu.edu.cn)

Received 17 November 2021; Accepted 17 November 2021; Published 2 December 2021

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

## 1. Background

With the development of society and the changes in people's lifestyles, more chronic illnesses have become the world's biggest killers. According to the World Health Organization (WHO), ischemic heart disease, stroke, chronic obstructive lung disease, and lower respiratory infections have remained the top killers during the past decade in both developed and developing countries. Artificial intelligence (AI) has emerged as a promising technology in healthcare with many applications, including diagnosis, treatment, prevention, and medical payment systems. Recently, a great number of new information technologies have emerged to assist the applications of AI in healthcare, such as big data and mobile internet networks. Therefore, AI-based electronic (E)-healthcare and medical (M)-healthcare solutions are aimed at gaining information, processing it, and giving a well-defined output to the end-user (the physician, the patient, or the caregiver).

AI-based E-healthcare and M-healthcare have shown significant potential to achieve the goals and demands of medical treatment. Owing to the data-driven approach, AI has brought a paradigm shift from information fields to other applications, such as intelligent diagnosis and treatment, intelligent payment, intelligent drug research and development, and intelligent health management and diagnosis related groups. Consequently, many researchers in the information and communication technology field are focusing on AI for healthcare applications.

The prime aim of this Special Issue is to bring together researchers to report recent research in AI in healthcare and exchange new ideas with innovative technologies and solutions towards AI in E-healthcare and M-healthcare applications. The proposed submissions and presentations should be original and unpublished works.

The Special Issue is composed of 11 outstanding contributions. The topics of these articles are mainly concerned with AI technology for E-healthcare and M-healthcare. We believe that these articles will play a role in inspiring our readers.

## 2. Summaries of Accepted Articles

The first article "An Improved Double Channel Long Short-Term Memory Model for Medical Text" authored by Shengbin Liang et al. presented an improved Double Channel (DC) mechanism as a significant enhancement to Long Short-Term Memory (LSTM) to solve the triage and precise treatment of patients. The experimental results have shown that the DC-LSTM model proposed by the authors has significantly superior accuracy and ROC compared with the basic CNN-LSTM model.

The second article is "Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning" authored by Shujun Zhang et al. In the article, three categories of deep learning object detection networks including Faster R-CNN, YOLOV3, and SSD were applied to implement automatic lesion detection with the best precision of 89.77%.

Meanwhile, statistical analysis of the locations, shapes of the lesions, and possible related diseases was conducted with valid conclusions. The research contributed to the intelligent assisted diagnosis and prevention and treatment of ischemic stroke.

The third article “Analysis on the Development Status of Coal Mine Dust Disaster Prevention Technology in China” authored by Hui Zhang et al. introduced the basic concepts, generation, distribution, and hazards of coal mine dust and analyzed the characteristics, applicable conditions, and use effects of various dust control measures such as ventilation dust removal and wet dust removal. The authors proposed specific prevention and control measures for related occupational diseases and discussed the development trend of dust prevention and control technology in the hope of providing guidance and reference for coal mine dust prevention and control.

The fourth article “Predicting Mental Health Problems with Automatic Identification of Metaphors” authored by Nan Shi et al. proposed a method for automatically detecting metaphors in texts to predict various mental health problems, specifically anxiety, depression, inferiority, sensitivity, social phobias, and obsession. They performed experiments on a composition dataset collected from second-language students and on the eRisk2017 dataset collected from Social Media. The experimental results showed that the proposed approach can help predict mental health problems in authors of written texts, and the proposed algorithm performs better than other state-of-the-art methods.

The fifth article “Cross-Database Micro-Expression Recognition Exploiting Intradomain Structure” authored by Yanliang Zhang et al. overcame this problem by exploiting the intradomain structure. Nonparametric transfer features were learned through intradomain alignment, while at the same time, a classifier was learned through intradomain programming. In order to evaluate the performance, a large number of cross-database experiments were conducted in CASMEII and SMIC databases. The comparison of results showed that this method can achieve a promising recognition accuracy and with high computational efficiency.

The sixth article “Efficient Algorithms for E-Healthcare to Solve Multiobject Fuse Detection Problem,” authored by Ijaz Ahmad et al., focused on the improved K-nearest neighbor (MK-NN) algorithm for electronic medical care to realize intelligent medical services and applications. The authors introduced modifications to improve the efficiency of MK-NN, and a comparative analysis was performed to determine the best fuse target detection algorithm based on robustness, accuracy, and computational time. Experimental results showed that the improved K-NN algorithm was the best model in terms of robustness, accuracy, and computational time.

The seventh article “ASC Performance Prediction for Medical IoT Communication Networks,” authored by Fagen Yin et al., investigated the secrecy performance of medical IoT communication networks. To improve the secrecy performance, the authors adopted a cooperative communication strategy. The authors also used the average secrecy

capacity (ASC) as a metric, and the expressions are first derived. Then, a secrecy performance intelligent prediction algorithm was proposed. The extensive simulations were used to verify the proposed method. Compared with other methods, the proposed algorithm realized a better prediction precision.

The eighth article is “Clinical Effects of Form-Based Management of Forceps Delivery under Intelligent Medical Model,” authored by Siming Xin et al. In the article, patients with forceps delivery in Maternal and Child Health Hospital Affiliated to Nanchang University were divided into two groups: form-based patients from January 1, 2019, to December 31, 2020, were selected as the study group, while traditional protocol patients from January 1, 2017, to December 31, 2018, were chosen as the control group. The conclusions were that form-based management could help assess the security of forceps delivery comprehensively, as it could not only improve the success rate of the one-time forceps traction scheme but also reduce the incidence of maternal and neonatal adverse outcomes effectively.

The ninth article “Service Migration Policy Optimization considering User Mobility for E-Healthcare Applications,” authored by Xuhui Zhao et al., proposed a service migration solution based on migration zone and formulated service migration cost with a comprehensive model that captures the key challenges. The authors also formulated a service migration problem into Markov decision process to obtain optimal service migration policies that decide where to migrate in a limited area. The authors proposed three algorithms to resolve the optimization problem given by the formulated model. The results showed that the proposed service migration approach reduces the total cost by up to 3 times compared to no migration and outperforms the general solution in terms of the total expected reward.

The tenth article “Zero-Watermarking Algorithm for Medical Image Based on VGG19 Deep Convolution Neural Network,” authored by Baoru Han et al., proposed a robust zero-watermarking algorithm for medical images’ security based on VGG19. The proposed algorithm utilized Hermite chaotic neural network to scramble the watermarking image for secondary protection, which enhanced the security of the algorithm. Compared with the existing related works, the proposed algorithm was simple to implement and could effectively resist local nonlinear geometric attacks, with good robustness, security, and invisibility.

The eleventh article is “A High-Sensitivity Method Based on Advanced Optical Waveguide Technology to Detect HBsAg,” authored by Pingping Xiao and Xiaoxiong Hu. A novel method for the detection of the hepatitis B surface antigen (HBsAg) at low concentrations, using the ultrahigh-order guided mode acting as the probe excited by a symmetrical metal-cladding waveguide, was proposed in the article. The measurement results indicated that this new method can precisely detect HBsAg at the concentrations in the lower region of the clinical gray area (i.e., below 20 ng/mL), the lower region of the current clinical gray area of HBsAg (below 20 ng/ml) can be measured, and the resolution can be reached (2 ng/mL).

### **Conflicts of Interest**

The Guest Editors declare that they have no conflicts of interest.

### **Acknowledgments**

The Guest Editors would like to thank all the authors who submitted their valuable contributions and anonymous reviewers who carefully scrutinized and evaluated the articles submitted. They would also like to extend their sincere thanks to the Editor-in-Chief of the Journal of Healthcare Engineering, Prof. Zollo, for providing this opportunity and lots of guidance throughout the process and the editorial staff of Hindawi for their continuous support and professional instruction during the whole publication of the special section.

## Research Article

# A High-Sensitivity Method Based on Advanced Optical Waveguide Technology to Detect HBsAg

Pingping Xiao <sup>1</sup> and Xiaoxiong Hu<sup>2</sup>

<sup>1</sup>Department of Electronic Information Engineering, College of Physics and Engineering Technology, Yichun University, Yichun, Jiangxi 33600, China

<sup>2</sup>Department of Infectious Diseases, The People's Hospital of Yichun City, Yichun, Jiangxi 33600, China

Correspondence should be addressed to Pingping Xiao; [xpp7967@163.com](mailto:xpp7967@163.com)

Received 25 May 2021; Revised 25 June 2021; Accepted 19 July 2021; Published 27 July 2021

Academic Editor: Xingwang Li

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

A novel method for the detection of the hepatitis B surface antigen (HBsAg) at low concentrations, using the ultrahigh-order guided mode acting as the probe excited by a symmetrical metal-cladding waveguide, is proposed. The method using the fact of the minimum value of the absorption peaks is proportional to the concentration of the sample to be detected to realize the detection of the hepatitis B virus at extremely low concentrations. It is realized that the low concentration of the HBsAg measurement relied on the principle of the minimum value of the absorption peak and the concentration having a good linear relationship. The measurement results indicate that this new method can precisely detect HBsAg at the concentrations in the lower region of the clinical gray area (i.e., below 20 ng/mL), the lower region of the current clinical gray area of HBsAg (below 20 ng/ml) can be measured, and the resolution can be reached (2 ng/mL).

## 1. Introduction

It is well known that the infection of hepatitis B virus (HBV) can cause acute and chronic hepatitis B and even can lead to death of the person who got infected [1]. Patients with HBV for a long time may progress to liver fibrosis, cirrhosis, liver cancer, and a series of complications such as portal hypertension syndrome and hepatic encephalopathy, which seriously threaten their lives [2]. According to the incomplete statistics, all around the world, there are about two billion people infected with HBV in their lives [3, 4], and about 12% of them have developed chronic hepatitis B virus infections. And there are about 650,000 people who lost their lives because of liver failure, cirrhosis, and liver cancer caused by HBV infection [5–7]. The rate of the cirrhosis and hepatocellular carcinoma (HCC) caused by HBV infection is 30% and 45%, respectively [8], and therefore, it is very significant to develop methods that are able to precisely detect the concentrations of HBV in human blood, for both clinical medicine and our society.

Qualitative detection of HBsAg is the main basis for hepatitis B patients in clinics. The common detection methods of HBsAg include colloidal gold immunochromatography assay (GICA) [9], enzyme-linked immunosorbent assay (ELISA), chemiluminescence method (CLIA) [10], and electrochemiluminescence immunoassay (ECLIA) [11]. There are advantages and disadvantages for the four main detection methods. At present, the most widely used technology is the ELISA because of its high sensitivity in the world. It is very simple, low cost, and sensitive for the GICA. CLIA can be used to detect lower concentration or microconcentration samples in the reaction, but the detection is limited. ECLIA is a new label immunoassay technology after radioimmunoassay, enzyme immunoassay, and chemiluminescence immunoassay. The detection conditions are monopolized, the detection equipment is expensive, and the clinical accessibility is poor.

To improve the resolution of detection, an optical detection technology, surface plasmon resonance (SPR)

technology, is proposed by some researchers to detect the concentrations of the matter contained in liquid samples. The SPR technology uses the resonance absorbance peak as the probe excited at the interface of metal and nonmetal by the interaction between laser and free electrons in metal and has high detection sensitivities. The SPR technology also has the following outstanding advantages: no need of bulky equipment, real-time detection, label-free measurement, and little sample amount requirement [12]. Due to the outstanding advantages of real-time detection and little sample amount requirement, the SPR technology has been widely used in the studies of interactions between biomolecules and the combination dynamics between antibodies and antigens [13]. Biacore is the most typical commercial instrument that uses the SPR technology.

In our preliminary studies, we found that although the SPR technology already has wide applications, its operation principle has some issues or concerns. Firstly, the surface plasma wave can only be transmitted at the interface between the metal and medium, and then the full width at half maximum (FWHM) of the attenuation of the total reflection absorption peak is large because metals absorb light in the range of visible and near-infrared wavelengths. Secondly, since the coupling angle of the prism coupling system of the surface plasma wave is large, the prism coupling system has a large effective refractive index. These two issues severely limited the further improvement on the detection sensitivity of the SPR sensor. Additionally, since the SPR technology uses the evanescent field as the probe, the probing range is only within a few hundred nanometers, which limits the applications of the SPR technology only to the detection of macromolecular samples. Although several new technologies, such as leaky optical waveguide (LW) [14], reverse symmetry waveguide (RSW) [15], and long-range surface plasma wave (LSPR) [16], have been developed, these newly developed technologies made no breakthroughs in the operation principle of the SPR technology, and like in the existing SPR technology, in these newly developed technologies, the sample is still placed in the evanescent field where the light field is weak. With the consideration of these issues in the operation principle of the existing SPR technology, this paper proposes a new high-sensitivity detection technology to detect HBsAg based on the symmetrical metal-cladding waveguide technology.

## 2. Operation Principle

**2.1. Detection Principle.** A schematic diagram for the structure of the sensor chip named symmetrical metal-cladding waveguide (SMCW) is illustrated in Figure 1. From bottom to top, the sensor chip consists of the substrate glass (0.5 mm), the thick metal (gold or silver) film (200 nm) sputtered on one side, the gasket glass (0.5 mm) with a circular hole in the center, the upper glass (0.5 mm), and the thin metal (gold or silver) film (30 nm) sputtered on one side.

The attenuated total reflection spectrum shown in Figure 2 can be obtained when the incident angle of the incident beam scans from 0 to 90° in the three-layer symmetrical metal-cladding waveguide structure. Figure 2 shows one of the absorption peaks of the attenuated total reflection spectrum at a small incident angle.

According to Cao [17], the minimum of the absorption peak of the attenuation total reflection in the structure of a three-layer metal-cladding waveguide can be written as

$$R_{\min} \propto \left( 1 - \frac{4\text{Im}(\Delta\beta^L)}{\left[ \sqrt{\text{Im}(\beta^0)} + \text{Im}(\Delta\beta^L/\sqrt{\text{Im}(\beta^0)}) \right]^2} \right), \quad (1)$$

where  $\text{Im}(\beta^0)$  is the intrinsic loss determined by the absorption of the waveguide material and  $\text{Im}(\Delta\beta^L)$  is the radiation loss determined by the thickness of the upper metal film.

From equation (1), obviously, when  $\text{Im}(\Delta\beta^0) = \text{Im}(\Delta\beta^L)$ , i.e., when the intrinsic loss is equal to the radiation loss, the minimum value of the attenuated total reflection peak is equal to zero. When the sensor chip for the experiments is prepared, the structural parameters of the sensor chip are determined, and then the radiation loss of the sensor chip is also determined. If there is a change in the concentration of the ingredient to be detected in the sample, the intrinsic loss will vary, and then the minimum of the absorption peak of the attenuation total reflection will also change. According to this principle, the detection of sample concentration by measuring the minimum of the absorption peak of the attenuation total reflection may be realized.

**2.2. Sensitivity Analysis.** The basic structure of a three-layer planar optical waveguide is shown in Figure 3.

The TE mode distribution of the electric field in a three-layer planar optical waveguide can be obtained as

$$E_y(x) = \cos \begin{cases} A \exp(\alpha_0 x), & -\infty < x < 0, \\ \frac{A \cos(\kappa_1 x - \phi_{13})}{\cos \phi_{13}}, & 0 < x < h, \\ \frac{A \cos(\kappa_1 h - \phi_{13}) \exp[-\alpha_2(x-h)]}{\cos \phi_{13}}, & h < x < +\infty, \end{cases} \quad (2)$$

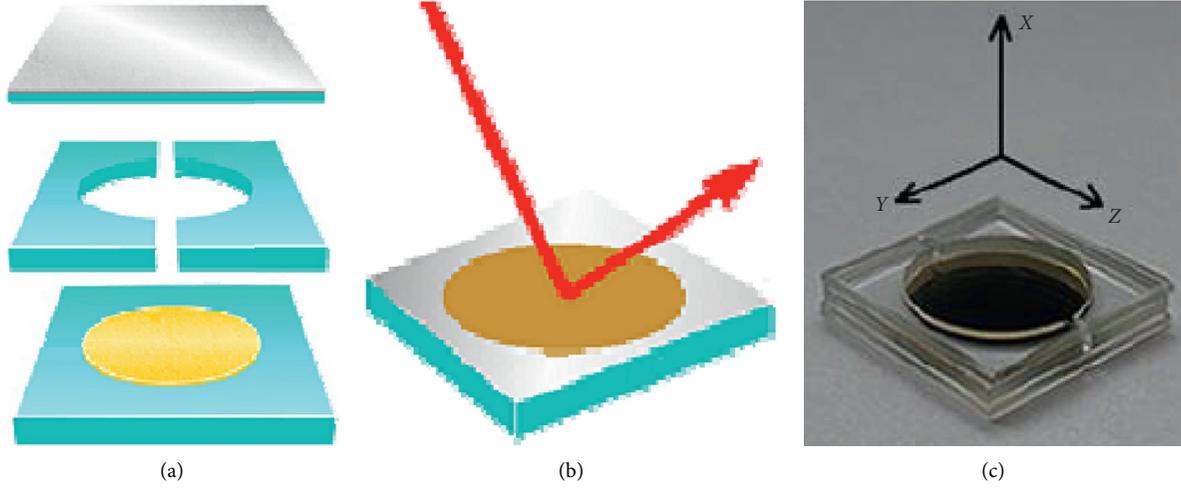


FIGURE 1: Schematic diagram to illustrate the structure of the symmetrical metal-cladding waveguide.

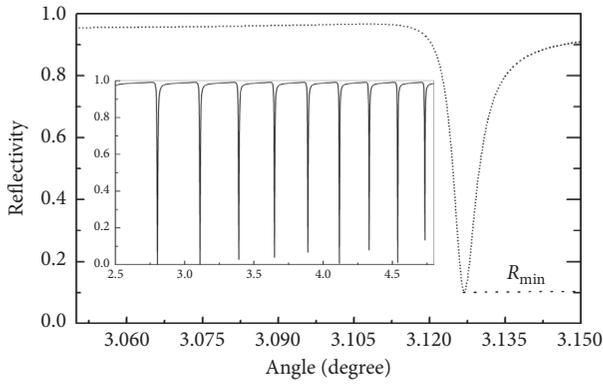


FIGURE 2: Absorption peaks of the attenuated total reflection.

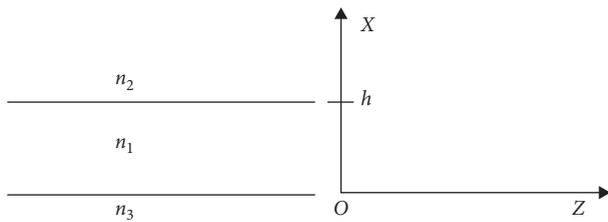


FIGURE 3: Schematic diagram of the three-layer waveguide.

where  $\phi_{12} = \tan^{-1}(\alpha_2/\kappa_1)$ ,  $\phi_{13} = \tan^{-1}(\alpha_3/\kappa_1)$ ,  
 $\alpha_3 = (k_0/\sqrt{N_{\text{eff}}^2 - n_3^2})$ ,  $\alpha_2 = (k_0/\sqrt{N_{\text{eff}}^2 - n_2^2})$ , and  
 $\kappa_1 = (k_0/\sqrt{n_1^2 - N_{\text{eff}}^2})$ .

The dispersion equation of the TE guided mode can be easily derived from the boundary conditions as follows [18]:

$$\kappa_1 h = m\pi + \arctan\left(\frac{\alpha_3}{\kappa_1}\right) + \arctan\left(\frac{\alpha_2}{\kappa_1}\right), \quad m = 0, 1, 2, 3, \dots \quad (3)$$

By differentiating both sides of equation (3), we can have

$$\frac{\Delta N_{\text{eff}}}{\Delta n_1} = \frac{(n_1/N_{\text{eff}}) \cdot [h + (\alpha_3/(\kappa_1^2 + \alpha_3^2)) + (\alpha_2/(\kappa_1^2 + \alpha_2^2))]}{h + (1/\alpha_3) + (1/\alpha_2)} \quad (4)$$

The power  $P_{ab}$  that the TE mode carries (from  $a$  to  $b$ ) in the waveguide is obtained from

$$P_{ab} = \frac{\beta}{2\omega\mu_0} \cdot \int_a^b [E_y(x)]^2 dx, \quad (5)$$

where  $\mu_0$  is the vacuum permeability. The region from  $a$  to  $b$  can be on either the substrate or the core layer of the waveguide:

$$\frac{P_1}{P_{\text{Total}}} = \frac{[h + (\alpha_3/(\kappa_1^2 + \alpha_3^2)) + (\alpha_2/(\kappa_1^2 + \alpha_2^2))]}{h + (1/\alpha_3) + (1/\alpha_2)}, \quad (6)$$

where  $P_1$  is the optical power coupled into the sample and  $P_{\text{Total}}$  is the total power input by the laser beam. The derivative of the effective refractive index of the guided mode can be obtained as

$$\frac{dN_{\text{eff}}}{dn_1} = \frac{n_1}{N_{\text{eff}}} \cdot \frac{P_1}{P_{\text{Total}}}, \quad (7)$$

where  $n_1$  is the refractive index of the sample to be detected and  $N_{\text{eff}}$  is the effective refractive index of the resonant mode.

Similarly, the derivative of the effective refractive index of the TM mode can be obtained as

$$\frac{dN_{\text{eff}}}{dn_1} = \frac{n_1}{N_{\text{eff}}} \cdot \frac{P_1}{P_{\text{Total}}} \cdot \frac{1 - \delta}{h_{\text{eg}}}, \quad (8)$$

where

$$h_{\text{eg}} = h + (n_3^2 n_1^2 \alpha_3 / (n_0^4 \kappa_1^2 + n_1^4 \alpha_3^2)) + (n_2^2 n_1^2 \alpha_2 / (n_2^4 \kappa_1^2 + n_1^4 \alpha_2^2)).$$

Then, the uniform sensitivity formula  $S$  of an optical waveguide sensor with the resonant mode can be written as

$$S = \frac{dN_{\text{eff}}}{dn_1} = \eta \cdot \frac{n_1}{N_{\text{eff}}} \cdot \frac{P_1}{P_{\text{Total}}}, \quad (9)$$

where  $\eta$  is a physical quantity related to the polarization of the laser and structure of the sensor, i.e.,  $\eta = 1$  for the TE mode and  $\eta = ((1 - \delta)/h_{eg})$  for the TM mode.

Since  $(n_i/N_{\text{eff}}) \ll 1$  and  $(P_i/P_{\text{Total}}) \ll 1$  for the surface plasmon resonance structure, it can be seen from equation (9) that  $S \ll 1$ . However, for a sensor based on the three-layer symmetrical metal-cladding waveguide structure, since the analyte is located in the region of the oscillating field where the light energy is very high, the incident laser light is almost all coupled into the waveguide, i.e.,  $P_i \approx P_{\text{Total}}$ . Since the incident angle  $\theta_{\text{air}}$  is very small, the effective refractive index  $N_{\text{eff}} = \kappa_0 n_{\text{air}} \sin \theta_{\text{air}}$  approaches to zero. According to equation (9), in principle, the sensitivity  $S$  approaches to infinity, and therefore, the sensitivity of this detection method based on the three-layer symmetrical metal-cladding waveguide structure is very high.

Simulations for the sensitivities of the SPR sensing technology using the prism coupled technology, the LSPR sensing technology using the prism coupled technology, and the RSW sensing technology using prism coupled technology and SMCW sensing technology with free-space coupling technology were conducted. The parameters used in the simulations and the results of these simulations for the four sensing technologies are listed in Table 1 and shown in Figure 4, respectively. The simulation results indicate that the sensitivity of the sensing technology proposed here using the variation of the minimum value of the absorption peak ( $R_{\text{min}}$ ) with the extinction coefficient of the sample and based on the three-layer symmetrical metal-cladding waveguide structure is at least two to three orders of magnitude higher than those of the other three traditional optical sensing technologies.

### 3. Sample Preparation

The original, national, and quantitative standard sample of HBsAg (NIBSC code: 00/880) was provided by the World Health Organization. The content of HBsAg in each bottle of freeze-dried serum is 16 ng, and the concentration is 160 ng/mL after adding 0.1 mL sterile ultrapure deionized water into each bottle. According to the dilution factor of the sample solution, a total of seven standard experimental samples with different concentrations besides pure distilled water were prepared.

Further sample preparation process is as follows:

- (1) Add 100  $\mu\text{L}$  of each of the seven standard experimental samples with different concentrations of HBsAg, respectively, to the seven same wells coated with the hepatitis B surface antibody, shake well, and then place the wells in an incubator of 37°C for about 1 hour.
- (2) Add 50  $\mu\text{L}$  of biotin-labeled capture antibody (secondary antibody) to each well, and then place the wells in an incubator of 37°C for about 30 minutes.
- (3) Use detergent to wash several times to wash away the unbound capture antibody, and then add 50  $\mu\text{L}$  of horseradish peroxidase-labeled streptavidin to each well, shake well, and place the wells in an incubator of 37°C for 30 minutes.

- (4) After washing, add 100  $\mu\text{L}$  of the reaction substrate o-phenylenediamine (OPD) to each well, shake well, and then place the wells in an incubator of 37°C for 30 minutes.
- (5) Terminate the reaction in each well by adding 100  $\mu\text{L}$  of acid, and then six orange-red liquids of different shades of the color after the oxidation by OPD are obtained. The used experimental samples of various specified concentrations and their corresponding labels are listed in Table 2.

## 4. Experimental Measurement

**4.1. Setup.** The schematic diagram of the experimental setup used for the measurements is shown in Figure 5. The model number of the laser source used was MW-SL-532 whose output laser has a center wavelength of 532 nm and has a maximum output optical power of about 30 mW. The two small holes (aperture 1 and aperture 2) were to ensure the collimation of the laser line. The polarization controller (polarizer) was used to select the TM polarization state of the laser. The high-reflection mirror was used to change the direction of the optical path to ensure that the incident light from the laser hits the center spot of the sensor chip. The sensor chip was placed on the goniometer which was connected to the two independently rotating parts controlled by a stepping motor, the inside part and the outside part, and the angular velocity of the outside part was twice of that of the inside part.

The photodetector (PD) used was a photoelectrical detection tube manufactured by ALPHALAS Inc. in Germany (model no. UPD-35-UVI), with a response wavelength range of 350–1700 nm. In the photoelectrical detection tube, the response signal is amplified by an analog signal amplifier, and then the amplified analog signal is converted into a digital signal by an analog digital signal conversion circuit. The final digital signal is input into the computer signal acquisition module, the acquired digital signals are further processed by self-programmed computer software, and then the measurement results will display on its monitor. In the experimental setup, the injection of the specimens into the sensor chip was realized by an electric peristaltic sampler with a low noise.

**4.2. Measurement.** At the beginning of the measurement, the inner rotating part that held the sensor chip was manually adjusted so that the incident point of the laser was at the center spot of the sensor chip, and the incident angle of the laser was about 2°. The incident angle of the laser was controlled by the computer program to scan in the range between 2° and 60°. The narrowest half-height full width and the lowest ATR absorption peak on the computer screen was selected to be the working peak. Then, the starting and ending laser incident angles of the selected working peak were set as the incident angle scanning range of the laser for the measurements, and then the laser was scanned back and forth within the set incident angle scanning range multiple times to make sure this scanning can repeat well. Since the

TABLE 1: Parameters used in the simulations for the four sensing technologies in the resonant mode.

Sensor	SPR	LSPR	RSW	SMCW
Prism ( $\epsilon_0$ )	3.24	3.24	3.24	1.0
Buffer layer ( $\epsilon_1, d_1$ )		1.78, 1060 nm		1.0
Upper metal film ( $\epsilon_2, d_2$ )	$-12 + i0.5$ , 50 nm	$-12 + i0.5$ , 15 nm	$-12 + i0.5$ , 50 nm	$-12 + i0.5$ , 39 nm
Upper glass ( $\epsilon_3, d_3$ )				2.25, 0.5 mm
Guide wave film ( $\epsilon_4, d_4$ )			2.25, 1500 nm	1.78, 0.5 mm (sample)
Bottom glass ( $\epsilon_5, d_5$ )				2.25, 0.5 mm
Bottom metal film ( $\epsilon_6, d_6$ )				$-12 + i0.5$ , 200 nm
Sample ( $\epsilon_7$ )	1.78	1.78	1.78	
$\Delta k$	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-6}$

Laser wavelength:  $\lambda = 488$  nm;  $\Delta k$ : the change of the imaginary part of the refractive index of the sample.

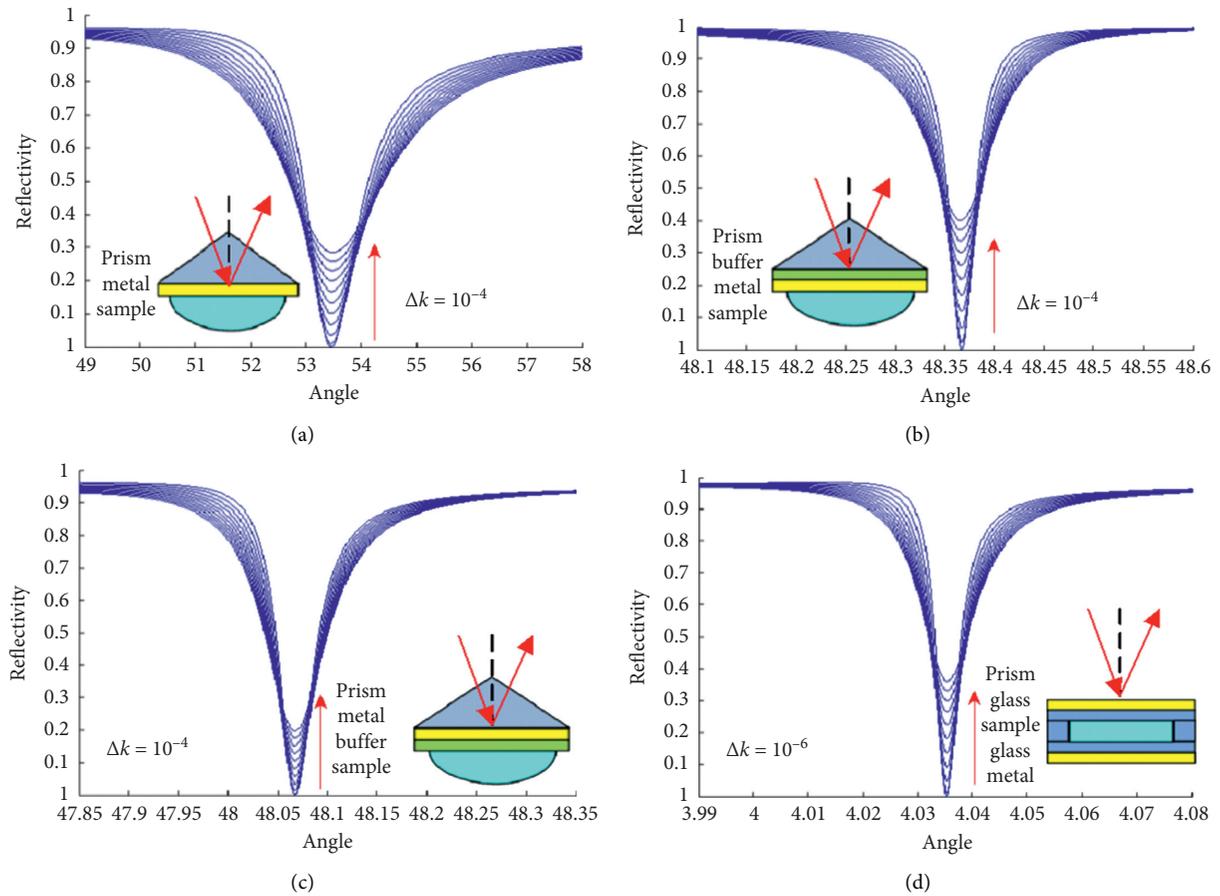


FIGURE 4: Variations of the ATR with the extinction coefficient for the four different sensing technologies. (a) SPR. (b) LSPR. (c) RSW. (d) SMCW.

TABLE 2: The specimens used in the experimental measurements with various concentrations of HBsAg.

Label of the specimen	Concentration (ng/mL)
A	0
B	20
C	40
D	60
E	80
F	100
G	120

gap between the forward and backward screws will induce errors in the measured experimental data, the laser was scanned in the same one direction to acquire the experimental data.

After all the moving parts are ensured in their stable working conditions, sample A was slowly injected into the sample chamber of the sensor chip by using an electric peristaltic pump. After the solution of the sample fully filled the sample chamber of the sensor chip, the peristaltic pump was shut off, and the solution of the sample in the sample chamber was let to stabilize for several minutes

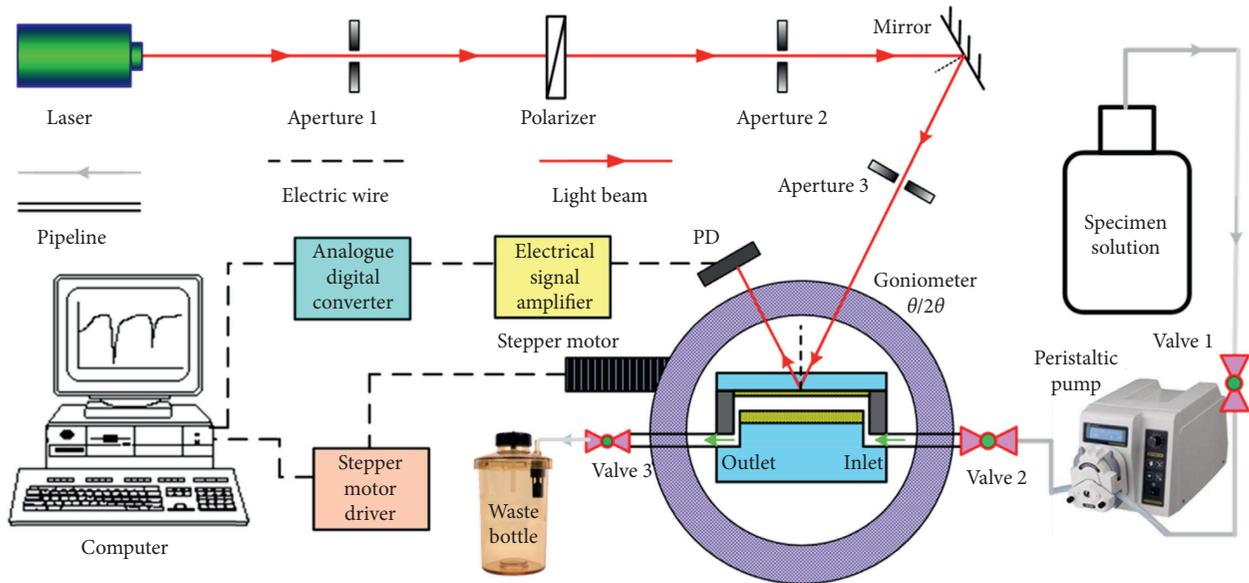


FIGURE 5: Schematic diagram of the experimental setup used for the measurements.

before the measurement started. Computer software was then turned on to control the laser to scan in one direction from the set starting to the set ending incident angles. The absorption peak acquired during each scan was stored as the measurement data. To obtain more reliable experimental data, during the measurements, the components in the experimental setup were maintained stationary, and the laser scanning was made sure to be just in one direction.

The sensor chip must be rinsed with sterile ultrapure deionized water multiple times before it is used to measure the next sample until all the seven samples are measured, and the measurement process for each of the samples was a repeat of the above process to measure sample A.

## 5. Results and Discussion

**5.1. Results.** Figure 6 shows the plot of the minimum value of the absorption peak ( $R_{\min}$ ) versus the concentration of the HBsAg standard sample. The lower right inset in Figure 6 shows all the measured curves of the attenuating total reflection absorption peak, and the upper left inset in Figure 6 is an enlarged view of the bottom peak portion of the measured curves shown in the lower right inset in Figure 6. The plot showed a very good linear relation between measured  $R_{\min}$  and the concentration of HBsAg in the range of the sample concentrations measured. It was also noticed that the minimum value of the ATR absorption peak varied about 0.05 with a variation of 20 ng/mL in the concentration of the HBsAg sample. Due to the resolutions of the data acquisition and computer display, the resolution of this measurement method to differentiate the minimum value ( $R_{\min}$ ) of the ATR peak was not less than 0.004. Considering the influences of other noises in the measurement process, the sensitivity of this optical waveguide detection method will not be lower than 2 ng/mL.

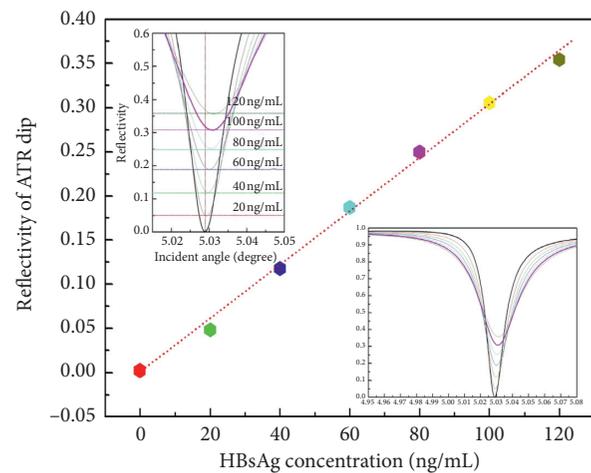


FIGURE 6: Variation of the minimum reflectivity related to different applied HBsAg concentrations.

**5.2. Discussions.** The plot shown in Figure 6 indicates that the minimum value of the ATR absorption peak ( $R_{\min}$ ) is not absolutely equal to zero when the concentration of HBsAg is zero. The reason for this observation is that  $\text{Im}(\beta^0)$  and  $\text{Im}(\Delta\beta^L)$  of the optical waveguide are not absolutely equal since the thickness of the upper coupling metal film layer is not absolutely equal to the optimum thickness at the time of the preparation of the sensor chip, but this will not have any impact on the measurement accuracy of this detection method.

We noticed that, in Figure 6, the attenuating total reflection absorption peaks of different samples were not only shifted with each other in the vertical direction but also slightly shifted in the horizontal direction. These horizontal and vertical shifts of the attenuating total reflection absorption peaks were caused by the real and imaginary parts of the refractive index of the sample, and then the horizontal

and vertical shifts of the attenuating total reflection absorption peaks were totally not related. Therefore, the horizontal shifts of the attenuating total reflection absorption peaks will also not have any influence on the picked measurement parameter  $R_{\min}$  of the proposed detection method.

We also noticed that, in Figure 6, there were some deviations of the experimental data from the theoretical predictions. We believe these deviations of the experimental data from the theoretical predictions were mainly related to the following three factors:

- (1) When the sensor chip was prepared, there existed island structures on the surface of the upper metal film, and then the uniformity of the thickness of the upper metal film was not guaranteed.
- (2) The parallelism and half-width of the output laser of the selected laser model exceeded the theoretical expectations.
- (3) The operation temperature of the sensor chip increased when irradiated by the laser, which caused an inconsistent temperature environment for the sensor chip before and during the experimental measurement. We believe that this issue can be solved if the working device with the sensor chip can be placed in a high-precision incubator.

## 6. Conclusions

Theoretical and experimental results indicate that the lowest resolution of the sensing technology introduced here, based on the symmetrical metal-cladding waveguide configuration, to detect HBsAg at low concentrations can reach 2 ng/mL, which will be useful in the diagnosis and treatment of hepatitis B patients. According to the detection principle of this novel method, this method can be used to detect the quantitative level of any substance as long as the substance can be colored with chromogenic agents, such as hepatitis B e antigen (HBeAg) and C-reactive protein (CRP). The application of optical detection technologies may provide a new technical route for clinical detections.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request and with the permission of funders.

## Conflicts of Interest

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

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 11664043, 81660105, and

11764020), the Science and Technology Landing Project of the Department of Education of Jiangxi Province (Grant no. KJLD14089), the Provincial Natural Science Foundation of Jiangxi (Grant nos. 20192BAB202005 and 20132BAB201048), and the Science and Technology Project of the Department of Education of Jiangxi Province (Grant no. GJJ190862).

## References

- [1] P. Li and B. A. Li, "The clinical application of laboratory tests and diagnosis for common chronic liver diseases Chinese," *Journal of Laboratory Medicine*, vol. 42, no. 3, pp. 151–154, 2019.
- [2] X. Y. Zhang, C. Gu, and X. H. Sun, "Sequence analysis of the HBV S protein in patients with chronic hepatitis B," *Chinese Journal of Laboratory Medicine*, vol. 42, no. 7, pp. 564–568, 2019.
- [3] J. J. Ott, G. A. Stevens, J. Groeger, and S. T. Wiersma, "Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity," *Vaccine*, vol. 30, no. 12, pp. 2212–2219, 2012.
- [4] R. Lozano, M. Naghavi, K. Foreman et al., "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010," *Lancet*, vol. 380, no. 9859, pp. 2095–128, 2012.
- [5] X. Liang, S. Bi, W. Yang et al., "Epidemiological serosurvey of hepatitis B in China—declining HBV prevalence due to hepatitis B vaccination," *Vaccine*, vol. 27, no. 47, pp. 6550–6557, 2009.
- [6] X. Liang, S. Bi, W. Yang et al., "Evaluation of the impact of hepatitis B vaccination among children born during 1992–2005 in China," *The Journal of Infectious Diseases*, vol. 200, no. 7, pp. 39–47, 2009.
- [7] F. M. Lu and H. Zhuang, "Management of hepatitis B in China," *Chinese Medical Journal*, vol. 122, no. 1, pp. 3–4, 2009.
- [8] S. T. Goldstein, F. Zhou, S. C. Hadler, B. P. Bell, E. E. Mast, and H. S. Margolis, "A mathematical model to estimate global hepatitis B disease burden and vaccination impact," *International Journal of Epidemiology*, vol. 34, no. 6, pp. 1329–1339, 2005.
- [9] T. Allice, F. Cerutti, F. Pittaluga et al., "COBAS AmpliPrep-COBAS TaqMan hepatitis B virus (HBV) test: a novel automated real-time PCR assay for quantification of HBV DNA in plasma," *Journal of Clinical Microbiology*, vol. 45, no. 3, pp. 828–834, 2007.
- [10] I. M. Mackay, K. E. Arden, and A. Nitsche, "Real-time PCR in virology," *Nucleic Acids Research*, vol. 30, no. 6, pp. 1292–1305, 2002.
- [11] X. Ding and K. L. Yang, "Development of an oligopeptide functionalized surface plasmon resonance biosensor for online detection of glyphosate," *Analytical Chemistry*, vol. 85, no. 12, pp. 5727–33, 2013.
- [12] L. G. Carrascosa, A. Calle, and L. M. Lechuga, "Label-free detection of DNA mutations by SPR: application to the early detection of inherited breast cancer," *Analytical and Bioanalytical Chemistry*, vol. 393, no. 2, pp. 1733–1182, 2009.
- [13] E. S. Iuliana and K. Lida, "Ultrasensitive DNA microarray biosensing via in situ RNA transcription-based amplification and nanoparticle-enhanced SPR imaging," *Journal of American chemical society*, vol. 133, no. 12, pp. 4271–4243, 2011.

- [14] M. Zourob and S. Mohr, "An integrated optical leaky waveguide sensor with electrically induced concentration system for the detection of bacteria," *Lab on Chip*, vol. 5, pp. 1360–1365, 2005.
- [15] R. Horváth, H. C. Pedersen, and N. B. Larsen, "Demonstration of reverse symmetry waveguide sensing in aqueous solutions," *Applied physics letters*, vol. 81, no. 12, pp. 2166–2168, 2002.
- [16] C.-J. Huang, J. Dostalek, A. Sessitsch, and W. Knoll, "Long-range surface plasmon-enhanced fluorescence spectroscopy biosensor for ultrasensitive detection of *E. coli* O157:H7," *Analytical Chemistry*, vol. 83, no. 3, pp. 674–677, 2011.
- [17] Z. Q. Cao, *Guide-Wave Optics*, Science Press, Beijing, China, 2007.
- [18] P. P. Xiao, *Principle of Double Metal Cladding Waveguide and Its Applications*, Shanghai Jiaotong University, Shanghai, China, 2012.

## Research Article

# Zero-Watermarking Algorithm for Medical Image Based on VGG19 Deep Convolution Neural Network

Baoru Han <sup>1</sup>, Jinglong Du <sup>1</sup>, Yuanyuan Jia <sup>1</sup> and Huazheng Zhu <sup>2</sup>

<sup>1</sup>College of Medical Informatics, Chongqing Medical University, Chongqing, China

<sup>2</sup>College of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing, China

Correspondence should be addressed to Huazheng Zhu; huazhengzhu@cqu.edu.cn

Received 19 February 2021; Accepted 30 May 2021; Published 2 July 2021

Academic Editor: Lingwei Xu

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

Aiming at the security issues in the storage and transmission of medical images in the medical information system, combined with the special requirements of medical images for the protection of lesion areas, this paper proposes a robust zero-watermarking algorithm for medical images' security based on VGG19. First, the pretrained VGG19 is used to extract deep feature maps of medical images, which are fused into the feature image. Second, the feature image is transformed by Fourier transform, and low-frequency coefficients of the Fourier transform are selected to construct the feature matrix of the medical image. Then, based on the low-frequency part of the feature matrix of the medical image, the mean-perceptual hashing algorithm is used to achieve a set of 64-bit binary perceptual hashing values, which can effectively resist local nonlinear geometric attacks. Finally, the algorithm adopts a watermarking image after scrambling and the 64-bit binary perceptual hashing value to obtain robust zero-watermarking. At the same time, the proposed algorithm utilizes Hermite chaotic neural network to scramble the watermarking image for secondary protection, which enhances the security of the algorithm. Compared with the existing related works, the proposed algorithm is simple to implement and can effectively resist local nonlinear geometric attacks, with good robustness, security, and invisibility.

## 1. Introduction

With the construction of hospital normalization, digitization is widely applied in the medical field [1]. Modern medical equipment produces lots of digital medical images every day [2, 3]. Owing to the rapid development of information technology, intelligent medicine and remote diagnosis are becoming more and more mature [4–6]. A large number of medical images are often transmitted through the Internet [7]. Unauthorized persons can easily obtain, store, use, and tamper with medical images on the network [8]. In this scene, sensitive information, such as patient information recorded in medical images, is also easy to leak. With the development of the medical information system, medical image sharing and remote diagnosis technology are becoming more mature [9]. In the application of these techniques, it becomes more and more urgent to protect patient's information, such as personal information in magnetic

resonance imaging (MRI) and other medical images, as well as electronic medical record data from being leaked [10, 11]. Therefore, the problem of medical information security has become an urgent problem.

Medical image watermarking can effectively solve the above problem and provide safe and reliable protection for medical information. Current information security of the medical information system relies on the development of modern information technology [12]. Traditional encryption methods of cryptography have great limitations and high risks. It is unable to identify the copyright of data at the technical level, and it has been unable to adapt to the changing information environment. Digital watermarking is a kind of information security technology, which can be applied to image integrity authentication and image copyright protection [13, 14]. Originally, digital watermarking was used for copyright protection of digital multimedia on the Internet. With the increasing demand for information

security in the medical field, the invisibility and robustness of digital watermarking are used to hide patient's personal information in the medical image to secure transmission on the Internet. By means of digital watermarking, specific identification information is embedded into carrier images, so that the authenticity and integrity authentication of medical images, Electronic Patient Record (EPR) hiding, and copyright protection can be realized. Therefore, medical image watermarking has an extremely important application value, which attracts continuous attention from researchers [15, 16].

Medical image watermarking generally chooses different watermarking algorithms according to invisibility, robustness, and embedding capacity. The invisibility, robustness, and capacity of watermarking restrict each other, but most of the current medical image watermarking algorithms only focus on one of them. Medical images are mostly single channel gray images, and the details contained in the pixels are very important. Any changes may cause distortion, which will affect doctor's diagnosis. However, the application of traditional image watermarking to medical images can easily cause the distortion of medical images, which may lead to potential misdiagnosis. In order to reduce the influence of watermarking on the original information of the medical image, it is very important to design a medical image watermarking algorithm which is lossless to the original image. Since the robust zero-watermarking adopts the way of zero-embedding, the original medical image is not modified. While realizing protection, the robust zero-watermarking takes into account the robustness and capacity of the watermarking and will not destroy the integrity of the medical image. Therefore, it is very suitable for the medical image [17–19].

Robust medical image zero-watermarking mainly focuses on the ability of the algorithm to resist various attacks, that is, the algorithm can completely extract the zero-watermarking information after various attacks [20–22]. The geometric attack will change the position relationship between the attacked medical image and the original medical image, resulting in serious out of synchronization between zero-watermarking information and medical images, which makes it is extremely difficult to extract zero-watermarking. Therefore, the ability of the algorithm to resist geometric attacks has become a research hotspot in the field of medical image zero-watermarking [1, 12]. But these geometric attacks mainly refer to global geometric transformation (rotation, scaling, translation, etc.), which is a kind of linear transformation. However, in practical applications, there is usually another type of geometric attack, which belongs to the local nonlinear geometric attack. At present, there are a few watermarking algorithms that can resist local nonlinear geometric attacks, which make almost all watermarking algorithms invalid. Compared with the global geometric attack, the local nonlinear geometric attack is more difficult to implement.

To solve this problem, a zero-watermarking algorithm based on VGG19 against local nonlinear geometric attacks is proposed in this paper. In this algorithm, VGG19 can extract complex high-level features using a deep structure with

powerful representation and learning ability. At the same time, the network structure has good geometric invariance and can extract image features with high and good robustness. The algorithm uses a perceptual hashing algorithm based on deep features to construct zero-watermarking, which has good robustness against local nonlinear geometric attacks. Hermite chaotic neural network is used to scramble the watermarking image for secondary protection, and it can enhance the security of the algorithm. Experimental results show that the proposed algorithm can effectively resist local nonlinear geometric attacks and display the superior performance of the proposed algorithm compared with other watermarking algorithms.

## 2. The Fundamental Theory

*2.1. Feature Extraction Based on VGGnet.* VGGnet is a representative type of deep convolutional neural network (CNN), which is often used in feature extraction and transfer learning [23]. The most widely used VGGnet is VGG19, which consists of 19 hidden layers (16 convolution layers and 3 fully connected layers), as shown in Figure 1. VGG19 uses a series of  $3 \times 3$  convolution kernels to extract image features and expands the number of feature channels through convolution layers. Let  $W_i$  and  $b_i$  represent the weights and bias of the  $i$ th convolution layer, the feature can be extracted by

$$X_i^{\text{out}} = \sigma(W_i * X_i^{\text{in}} + b_i), \quad (1)$$

where  $X_i^{\text{out}}$  and  $X_i^{\text{in}}$ , respectively, denote the input and output feature maps and  $\sigma(\cdot)$  is the rectified linear unit (ReLU). In each convolution layer, the stride is set to 1. To avoid the explosion of calculation, VGG19 uses max pooling layers to reduce the size of the feature maps.

In fully connected layers, each node of the given layer is connected with all nodes of its previous layers, which can map the distributed feature representation to the sample label space by

$$Y = \text{FC}_3(\text{FC}_2(\text{FC}_1(P(X_{16}^{\text{out}}))))), \quad (2)$$

where  $\text{FC}(\cdot)$  denotes the operation of the fully connected layer and  $P(\cdot)$  refers to the max pooling operation.

At the end of VGG19, a softmax layer produces the classification result of the image:

$$Y_j = \frac{e^{z_j}}{\sum_{c=1}^C e^{z_c}}, \quad (3)$$

where  $Y_j$  is the probability of the  $j$ th node and  $z_j$  and  $C$ , respectively, denote the output of  $j$ th node and the number of the classification.

Compared with other types of CNNs, VGG19 improves the depth of the network and adopts the alternate structure of many convolution layers and nonlinear activation layers, which is beneficial to extract accurate features. In this work, different from image classification tasks, we just use the convolution layers and max pooling layers from the pre-trained VGG19 as the preprocessing method to extract deep

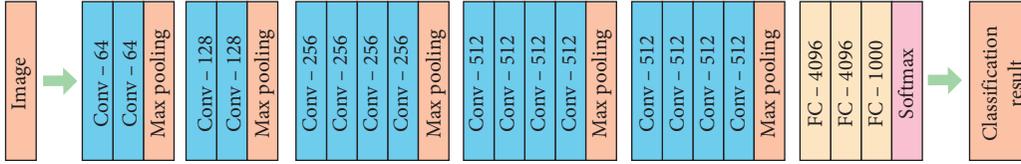


FIGURE 1: The structure of the VGG19 network.

feature maps from medical images. Unlike other medical zero-watermarking algorithms, our method can extract abstract high-level features from medical images to improve the antigeometric attack ability of zero-watermarking.

**2.2. Discrete Fourier Transform.** Discrete Fourier transform plays an important role in the development of signal analysis and processing. Because of its definite physical meaning, discrete Fourier transform is widely used in many fields of signal analysis and processing.

**2.2.1. One-Dimensional Discrete Fourier Transform.** Let  $f(x)$  denote a time domain function of  $x$ , where  $x$  represents a time domain variable and  $u$  is a frequency domain variable. When  $f(x)$  reaches the Dirichlet condition, the

discrete Fourier transform is formula (4) and the inverse transform is formula (5), where  $F(u)$  is a frequency domain function:

$$F(u) = \sum_{x=0}^{N-1} f(x)e^{-j2\pi ux/N}, \quad u = 0, 1, \dots, N-1, \quad (4)$$

$$f(x) = \frac{1}{N} \sum_{u=0}^{N-1} F(u)e^{j2\pi ux/N}, \quad x = 0, 1, \dots, N-1. \quad (5)$$

**2.2.2. Two-Dimensional Discrete Fourier Transform.** For a given image with the size of  $M \times N$ , when it satisfies the Dirichlet condition, the two-dimensional discrete Fourier positive transform is

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)e^{-j(2\pi/M)xu} e^{-j(2\pi/N)yv}, \quad u = 0, 1, \dots, M-1; v = 0, 1, \dots, N-1. \quad (6)$$

Its inverse transformation formula is

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v)e^{j(2\pi/M)xu} e^{j(2\pi/N)yv}, \quad x = 0, 1, \dots, M-1; y = 0, 1, \dots, N-1, \quad (7)$$

where  $x$  and  $y$  are the values in the space domain and  $u$  and  $v$  are the values in the frequency domain.

Two-dimensional discrete Fourier transform transforms an image from the space domain to the frequency domain, which has a clear physical meaning. The frequency of an image is an index that characterizes the intensity of the gray level change in the image and is the gradient of the gray level in the plane space. After the Fourier transform, the changes in the flat area of the image are represented by low-frequency coefficients, and the details of the image are represented by high-frequency coefficients. Figure 2 shows the result of the Fourier transform of a medical image.

After the two-dimensional discrete Fourier transform, the transform coefficient matrix of the image shows that if the origin of the transform matrix is set at the center, the spectral energy is concentrated near the center of the short transform coefficient matrix. If the origin of the two-dimensional discrete Fourier transform matrix is set in the upper left corner, the energy of the image will be concentrated in the four corners of the coefficient matrix. This is

determined by the nature of the two-dimensional discrete Fourier transform itself. It also shows that the image energy is generally concentrated in the low-frequency area. This determines that the two-dimensional discrete Fourier transform is very suitable for image processing. In this paper, a two-dimensional discrete Fourier transform is used to transform the features of the medical image into the Fourier domain to construct zero-watermarking.

**2.3. Perceptual Hashing Algorithm.** Perceptual hashing is considered as a one-way mapping from the multimedia data set to the perceptual content hashing value; that is, a short digital digest uniquely represents multimedia data with the same perceptual content. This digital digest is called perceptual hashing value, so the mapping process is also called the process of hash value generation.

Image perceptual hashing is usually called image digital fingerprint or image digital digest. It can map the image to a group of hash sequences, which greatly reduces the storage of

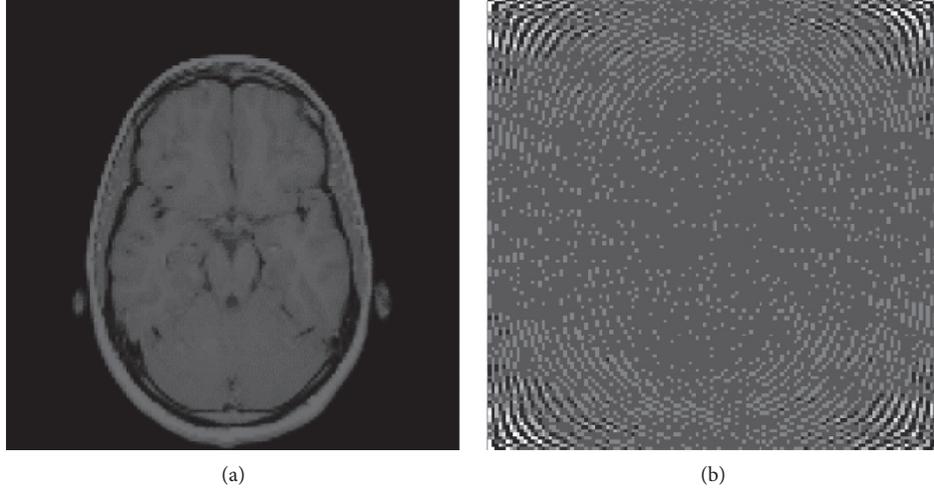


FIGURE 2: Discrete Fourier transform of the medical image. (a) Original medical image. (b) Medical images in the Fourier domain.

the digital image and brings great convenience to image management and maintenance. It has become a research hotspot in the field of multimedia signal processing and security. Perceptual feature extraction is the core of the perceptual hashing algorithm. The effectiveness and reliability of perceptual feature extraction will directly affect the uniqueness and robustness of the perceptual hashing sequence.

A mean-perceptual hashing algorithm is proposed in this paper. The mean-perceptual hashing algorithm uses the mean value of the elements of the medical image feature matrix to generate a hashing sequence. The construction of the feature image is realized by VGG19 and image fusion. Figure 3 describes the flowchart of the mean-perceptual hashing algorithm.

**2.4. Hermite Chaotic Neural Network.** In this work, a new Hermite chaotic neural network is used to scramble the watermarking image [24]. The structure of the network is shown in Figure 4, whose topology is  $m \times n \times 1$ .

The hidden layer neuron input is

$$O_j = w_j x, \quad j = 0, 1, 2, \dots, n-1, \quad (8)$$

where  $w_j$  is the weight between the input layer and hidden layer and  $c_j$  is the weight between the hidden layer and output layer. The hidden layer neuron output is

$$H_j(O_j), \quad j = 0, 1, 2, \dots, n-1, \quad (9)$$

where  $H_j(\cdot)$ ,  $j = 0, 1, 2, \dots, n-1$  represents Hermite orthogonal polynomial terms. The Hermite orthogonal polynomial formula is as follows:

$$\begin{cases} H_0(x) = 1, \\ H_1(x) = 2x, \\ H_{k+1}(x) = 2xH_k(x) - 2kH_{k-1}(x), \quad k = 1, 2, \dots, x \in (-\infty, \infty). \end{cases} \quad (10)$$

The output of the Hermite chaotic neural network is

$$y = \sum_{j=0}^{n-1} c_j H_j(O_j), \quad (11)$$

where  $(T_t, d_t)$ ,  $t = 1, 2, \dots, l$  is the training sample set and  $l$  is the number of training samples. Hermite chaotic neural network input is  $T_t = (x_{1t}, x_{2t}, \dots, x_{mt})$ , and its desired output is  $d_t$ . The backpropagation (BP) algorithm is used to train the network.

The weights of the network are trained according to the following formula:

$$e_t = d_t - y_t,$$

$$E = \frac{1}{2} \sum_{t=1}^l e_t^2,$$

$$\Delta c_j = -\eta \frac{\partial E}{\partial c_j}, \quad (12)$$

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}},$$

where  $t = 1, 2, \dots, l$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, n$ .

The chaotic sequence produced by the logistic chaotic function is used as a sample set. The logistic chaotic function is as follows:

$$x(n+1) = \mu x(n)(1-x(n)), \quad (13)$$

where  $3.5699456 \leq \mu \leq 4$ ,  $x \in (0, 1)$ .

In this work, we set  $\mu = 3.8$ ,  $l = 1000$ ,  $E = 10^{-12}$ ,  $\eta = 0.03$ ,  $n = 3$ , and the maximum number of training to 1500 epochs. The training process of the Hermite chaotic neural network is shown in Figure 5. When the number of training times is 260, the error is  $9.5169e-13$ , which has reached the expected error. When the initial value of the Hermite chaotic neural network is 0.66, the chaotic sequence generated by the network is shown in Figure 6. The chaotic sequence is used to scramble the watermarking image. The

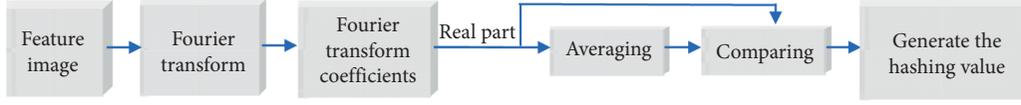


FIGURE 3: The flowchart of mean-perceptual hashing.

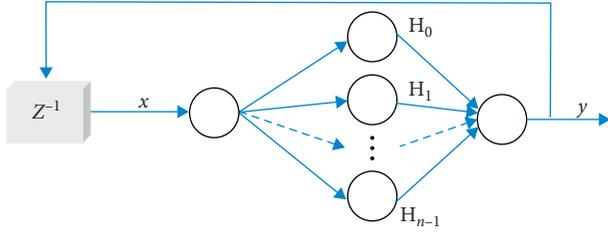


FIGURE 4: The architecture of Hermite chaotic neural network.

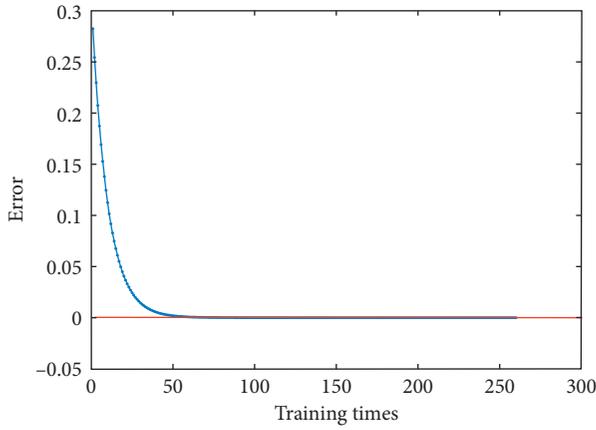


FIGURE 5: Training process.

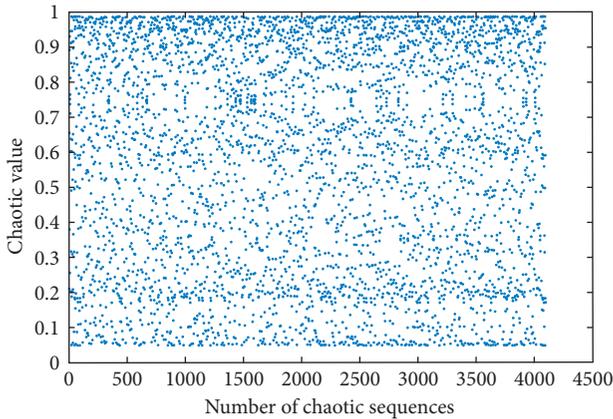


FIGURE 6: Chaotic sequence.

number of the chaotic sequence is determined by the size of the watermarking image.

### 3. Zero-Watermarking Embedding and Extraction Algorithm

3.1. *Embedding Algorithm.* Choosing an image with a specific meaning, our algorithm uses it as the original watermarking

image in the experiment. It is recorded as  $W(i, j) = \{w(i, j) | w(i, j) = 0, 1; 1 \leq i \leq M_1, 1 \leq j \leq M_2\}$ . The gray value of the original watermarking image is represented as  $w(i, j)$ . The original medical image is recorded as  $F(i, j) = \{f(i, j) | f(i, j) \in R, 1; 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$ . Here, the pixel gray value of the original medical image is represented by  $f(i, j)$ . For the convenience of calculation, let  $M_1 = M_2 = 64, N_1 = N_2$ . Figure 7 shows the watermarking embedding procedure.

- (1) Hermite chaotic neural network is used to scramble the position of pixels in the original watermarking image  $W(i, j)$  to get the scrambled watermarking image  $BW(i, j)$ .
- (2) The pretrained VGG19 is used to extract the deep feature maps  $FM(k, l, p)$  of the original medical image  $F(i, j)$ :

$$F(i, j) \longrightarrow \text{VGG19} \longrightarrow \text{FM}(k, l, p), \quad (14)$$

where  $1 \leq k \leq 8, 1 \leq l \leq 8, \text{ and } 1 \leq p \leq 512$ .

- (3) The deep feature maps  $FM(k, l, p)$  are fused to generate feature image  $FI(k, l)$ :

$$FI(k, l) = \sum_{p=1}^{512} FM(k, l, p). \quad (15)$$

- (4) Two-dimensional discrete Fourier transform is used to transform the feature image  $FI(k, l)$ , and the transform coefficients  $FIF(k, l)$  are obtained to construct the feature matrix of medical image  $FIM(k, l)$ :

$$\begin{aligned} FIF(k, l) &= 2D - \text{DFT}(FI(k, l)), \\ FIF(k, l) &\longrightarrow FIM(k, l). \end{aligned} \quad (16)$$

- (5) The hashing sequence of medical image feature matrix  $FIM(k, l)$  is extracted by the mean-perceptual hashing algorithm to generate 64-bit binary hashing sequence  $PH(q)$ :

$$PH(q) = m\text{Hash}(FIM(k, l)), \quad (17)$$

where  $1 \leq q \leq 64$ .

- (6) The 64-bit binary hashing sequence  $PH(q)$  is XORed with the scrambled watermarking image  $BW(i, j)$ . And, the watermarking extraction key  $\text{Key}(i, j)$  is generated to extract the watermarking image:

$$\text{Key}(i, j) = PH(q) \oplus BW(i, j), \quad (18)$$

where  $1 \leq q \leq 64, 1 \leq i \leq 64, \text{ and } 1 \leq j \leq 64$ . The watermarking extraction key  $\text{Key}(i, j)$  can be saved in a third party for later watermarking extraction.

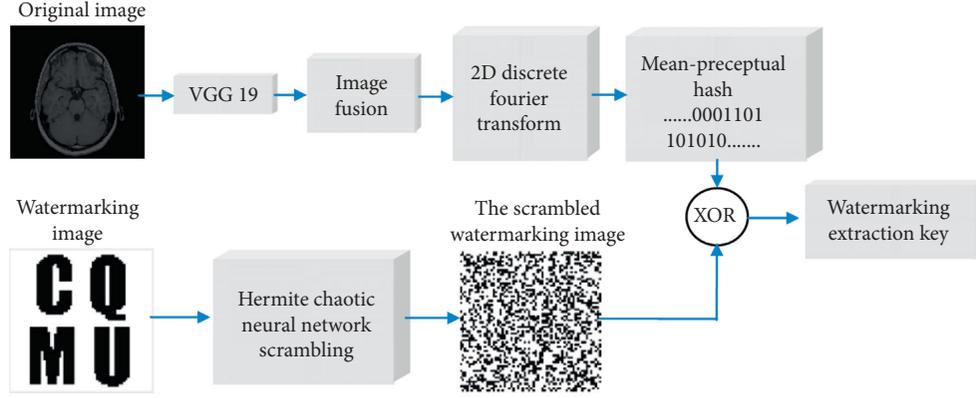


FIGURE 7: Watermarking embedding procedure.

3.2. *Extraction Algorithm.* The medical image to be tested is recorded as  $F'(i, j) = \{f'(i, j) | f'(i, j) \in R, 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$ . The extraction algorithm is similar to the embedding algorithm, and the specific steps are as follows. Figure 8 shows watermarking extraction procedure.

- (1) The pretrained VGG19 is used to extract the deep feature maps  $FM'(k, l, p)$  of the medical image to be tested  $F'(i, j)$ :

$$F'(i, j) \longrightarrow \text{VGG19} \longrightarrow FM'(k, l, p), \quad (19)$$

where  $1 \leq k \leq 8, 1 \leq l \leq 8, 1 \leq p \leq 512$ .

- (2) The deep feature maps  $FM'(k, l, p)$  are fused to generate feature image  $FI'(k, l)$ :

$$FI'(k, l) = \sum_{p=1}^{512} FM'(k, l, p). \quad (20)$$

- (3) Two-dimensional discrete Fourier transform is used to transform the feature image  $FI'(k, l)$ , and the transform coefficients  $FIF'(k, l)$  are obtained to construct the feature matrix of medical image  $FIM'(k, l)$ :

$$\begin{aligned} FIF'(k, l) &= 2D - \text{DFT}(FI'(k, l)), \\ FIF'(k, l) &\longrightarrow FIM'(k, l). \end{aligned} \quad (21)$$

- (4) The hashing sequence of medical image feature matrix  $FIM'(k, l)$  is extracted by the mean-perceptual hashing algorithm to generate 64-bit binary hashing sequence  $PH'(q)$ :

$$PH'(q) = m\text{Hash}(FIM'(k, l)), \quad (22)$$

where  $1 \leq q \leq 64$ .

- (5) The 64-bit binary hashing sequence  $PH'(q)$  is XORed with the watermarking extraction key  $Key(i, j)$  to extract the watermarking image  $BW'(i, j)$  from the medical image to be tested  $F'(i, j)$ :

$$BW'(i, j) = PH'(q) \oplus Key(i, j), \quad (23)$$

where  $1 \leq q \leq 64, 1 \leq i \leq 64, 1 \leq j \leq 64$ .

- (6) The Hermite chaotic neural network is used to inversely scramble the extracted watermarking image  $BW'(i, j)$  to obtain the restored watermarking image  $W'(i, j)$ .
- (7) The normalized correlation coefficient (NC) is used to detect the restored watermarking image  $W'(i, j)$  obtained from the medical image to be tested  $F'(i, j)$ . The normalized correlation coefficient formula is shown in the following formula:

$$NC = \frac{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w(i, j)W'(i, j)}{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w(i, j)w(i, j)}. \quad (24)$$

By comparing the NC value between the original watermarking image  $W(i, j)$  and the restored watermarking image  $W'(i, j)$ , the restored watermarking image  $W'(i, j)$  obtained from the medical image to be tested is evaluated. The larger the NC value, the greater the correlation between the original watermarking image  $W(i, j)$  and the restored watermarking image  $W'(i, j)$ .

The peak signal-to-noise ratio (PSNR) was used to evaluate the quality of the medical image to be tested. The PSNR formula is as follows:

$$PSNR = 10 \lg \frac{\text{MAX}^2}{\text{MSE}}, \quad (25)$$

$$\text{MSE} = \frac{1}{N_1 \times N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} [F(i, j) - F'(i, j)]^2,$$

where MAX is the maximum gray value of the image pixel and MSE is the mean square error between the original image and the test image.

## 4. Experiment and Analysis

In this work, several experiments are conducted by applying local nonlinear geometric attacks to investigate the performance and effectiveness of the proposed zero-watermarking algorithm. Five original medical images of different parts of the human body produced by different medical

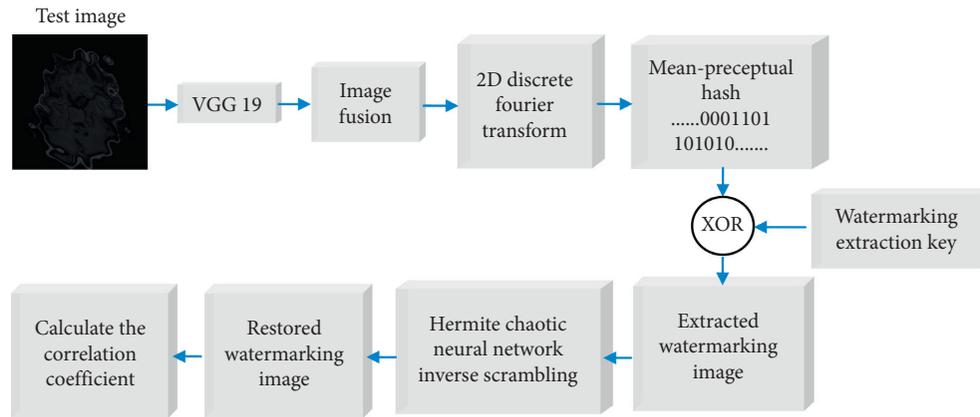


FIGURE 8: Watermarking extraction procedure.

equipment are used to test the effectiveness of the proposed zero-watermarking algorithm. The original medical images with the size of  $128 \times 128$  are shown in Figure 9. The original  $64 \times 64$  watermarking image and the scrambled watermarking image are shown in Figure 10.

The following experiments test the ability of the watermarking algorithm to resist local nonlinear geometric attacks. For the convenience and repeatability of the experiment, we use the filter function of Adobe Photoshop software to implement the following common local distortion attacks. There are four types of local distortion attacks: ripple distortion attack, extrusion distortion attack, spherical distortion attack, and rotation distortion attack.

**4.1. Ripple Distortion Attack.** Ripple distortion attack is a common local nonlinear geometric attack, which transforms the coordinates of image pixels according to different functions. Table 1 shows the experimental results under the ripple distortion attack when the distortion quantity increases from 150% to 750%. Table 2 gives the restored watermarking images with the smallest NC value and their corresponding test medical image. The restored watermarking images can be clearly identified from Table 2, which shows that the proposed algorithm is very effective against ripple distortion attack.

**4.2. Extrusion Distortion Attack.** For extrusion distortion attack, distortion quantity is varied between 10% and 90% for simulation. The experimental results are shown in Table 3. All NC values are greater than 0.74. In Table 4, it is obvious that the restored watermarking image is very clear. The experimental results show that the proposed watermarking algorithm has good robustness against extrusion distortion attacks.

**4.3. Spherical Distortion Attack.** Then, we test the robustness of the proposed method on image spherical distortion. The test medical images were attacked by spherical distortions of 10%, 30%, 50%, 70%, and 90% in turn. Experimental results are listed in Table 5. It can be observed from Table 5 that most of the NC values are close to 1.0. The restored

watermarking image in Table 6 further proves that the proposed algorithm has excellent performance under the spherical distortion attack.

**4.4. Rotation Distortion Attack.** We also evaluate the robustness of the proposed algorithm under rotation distortion attack. Table 7 shows that when the distortion quantity increases, the NC value and the PSNR values gradually decrease. As can be seen from Table 8, even if the NC value is 0.5621, the watermark image can still be accurately extracted. These results demonstrate that the proposed algorithm has better robustness against the rotation distortion attack.

The above experimental results indicate that the proposed algorithm has strong capability against all the four types of local distortion attacks. Because it is a local nonlinear attack, the NC value does not necessarily decrease with the increase of the distortion quantity, and the change of the NC value is not the same as the change of the PSNR value.

## 5. Algorithm Comparison

In this section, the proposed algorithm is compared with two representative medical image watermarking algorithms. The details of these three algorithms are shown in Table 9. These three algorithms are robust medical image watermarking algorithms. Every robust watermarking algorithm can resist all kinds of attacks with good robustness and security and can solve all kinds of information security problems of the medical image. The proposed algorithm and [25, 26] are zero-watermarking algorithms. Their watermarking is based on the important features of the medical image rather than modifying the content of the medical image. They are easy to implement, fast to embed and extract, and have the advantages of small amount of computation and high speed. Therefore, the proposed algorithm and [25, 26] are very suitable for medical images. In these algorithms, DWT, DCT, and DFT are all linear transformations. However, VGG19 is a multilayer perceptron specially designed for image recognition. Its feature extraction is essentially a nonlinear transformation.

In order to further analyze the performance of the proposed algorithm, the algorithm [25, 26] has also carried

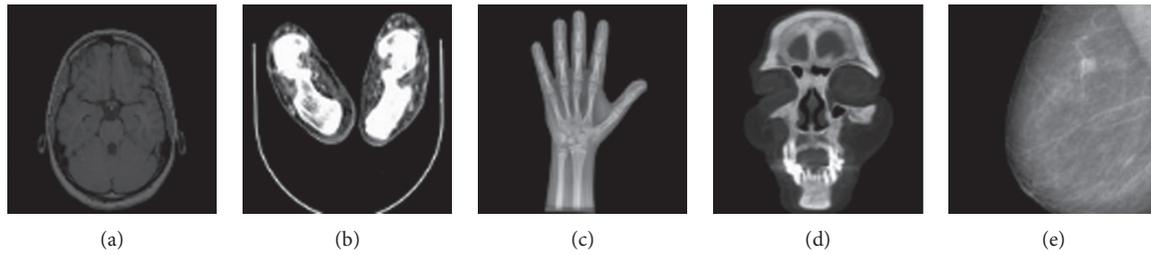


FIGURE 9: The original medical images.

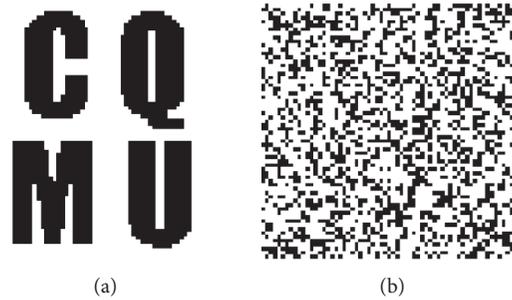


FIGURE 10: The watermarking images. (a) The original watermarking image. (b) The scrambled watermarking image.

TABLE 1: Experimental results under ripple distortion attack.

Distortion quantity (%)		150	300	450	600	750
<i>Medical image A</i>	PSNR (dB)	12.7889	11.9036	11.5765	11.2648	11.0596
	NC	0.87595	0.81021	0.81021	0.81106	0.74818
<i>Medical image B</i>	PSNR (dB)	16.6378	12.8092	12.8092	11.7588	11.0855
	NC	0.87595	0.87595	0.87595	0.68929	0.75075
<i>Medical image C</i>	PSNR (dB)	21.8267	17.8745	15.8378	14.6473	13.8828
	NC	0.81363	0.74989	0.81163	0.65842	0.59668
<i>Medical image D</i>	PSNR (dB)	21.3456	17.7586	16.1687	14.8927	14.0789
	NC	0.87509	0.87138	0.81021	0.81392	0.81392
<i>Medical image E</i>	PSNR (dB)	28.1822	25.2023	23.3165	21.9457	20.5296
	NC	0.87823	0.75104	0.68987	0.68872	0.68872

TABLE 2: The images under ripple distortion attack.

	Medical image A	Medical image B	Medical image C	Medical image D	Medical image E
The smallest NC value	0.74818	0.68929	0.59668	0.81021	0.68872
The restored watermarking image					
Medical image to be tested					

out local nonlinear geometric attack experiments under the same conditions. Their results are compared with the experimental results of the proposed algorithm.

The comparison of the algorithms under four types of local nonlinear geometric attacks is shown in Figure 11. As the blue bar of Figure 11(a) shows, it is obvious that the

TABLE 3: Experimental results under extrusion distortion attack.

Distortion quantity (%)		10	30	50	70	90
<i>Medical image A</i>	PSNR (dB)	13.0494	11.4873	10.6621	10.2615	9.9121
	NC	1.0	1.0	1.0	0.9354	0.9354
<i>Medical image B</i>	PSNR (dB)	21.7239	16.4862	14.1855	12.7081	11.6212
	NC	1.0	0.84222	0.7815	0.81049	0.81278
<i>Medical image C</i>	PSNR (dB)	26.561	19.8342	17.1758	15.695	14.9073
	NC	0.93826	0.93826	0.74818	0.74818	0.81163
<i>Medical image D</i>	PSNR (dB)	24.4511	18.0799	15.8822	14.692	14.0065
	NC	0.93626	0.81021	0.78076	0.78076	0.87395
<i>Medical image E</i>	PSNR (dB)	32.0185	26.229	23.5813	21.8296	20.504
	NC	0.9394	0.9394	0.9394	0.87566	0.87766

TABLE 4: The images under extrusion distortion attack.

	Medical image A	Medical image B	Medical image C	Medical image D	Medical image E
The smallest NC value	0.9354	0.7815	0.74818	0.78076	0.87566
The restored watermarking image					
Medical image to be tested					

TABLE 5: Experimental results under spherical distortion attack.

Distortion quantity (%)		10	30	50	70	90
<i>Medical image A</i>	PSNR (dB)	12.8585	11.4917	11.1459	11.3035	11.3435
	NC	1.0	1.0	0.87166	0.87166	0.87166
<i>Medical image B</i>	PSNR (dB)	18.4985	14.4427	13.0098	11.7524	10.5874
	NC	0.93654	0.87366	0.81221	0.81221	0.74704
<i>Medical image C</i>	PSNR (dB)	27.259	19.872	16.6028	14.5081	13.0453
	NC	0.93512	1.0	1.0	1.0	0.93997
<i>Medical image D</i>	PSNR (dB)	23.9151	17.1131	14.5852	13.2542	12.4597
	NC	1	1	1	1	0.93769
<i>Medical image E</i>	PSNR (dB)	30.64	25.4487	22.91	20.9156	19.2908
	NC	0.9394	1	1	0.81763	0.62984

TABLE 6: The images under spherical distortion attack.

	Medical image A	Medical image B	Medical image C	Medical image D	Medical image E
The smallest NC value	0.87166	0.74704	0.93512	0.93769	0.62984
The restored watermarking image					
Medical image to be tested					

TABLE 7: Experimental results under rotation distortion attack.

Distortion quantity (%)		10	30	50	70	90
<i>Medical image A</i>	PSNR (dB)	13.8738	13.2751	12.8138	12.5202	12.2938
	NC	1.0	1.0	1.0	1.0	0.94054
<i>Medical image B</i>	PSNR (dB)	21.5541	16.336	14.1869	12.7102	11.5936
	NC	0.93883	0.81049	0.87509	0.62698	0.62927
<i>Medical image C</i>	PSNR (dB)	23.2593	16.2611	14.4212	13.6166	13.1196
	NC	0.93626	0.93626	0.93626	0.93626	0.5621
<i>Medical image D</i>	PSNR (dB)	23.6471	17.9937	16.1955	15.1909	14.7219
	NC	0.94054	0.94054	0.90682	0.90682	0.90682
<i>Medical image E</i>	PSNR (dB)	32.9927	27.8346	25.4521	23.7641	22.4491
	NC	0.9394	0.9394	0.87823	0.87823	0.8182

TABLE 8: The images under rotation distortion attack.

	Medical image A	Medical image B	Medical image C	Medical image D	Medical image E
The smallest NC value	0.94054	0.62698	0.5621	0.90682	0.8182
The restored watermarking image					
Medical image to be tested					

TABLE 9: Comparison between the proposed algorithm and other algorithms.

Algorithms	Objective	Method	Medical image content	Transform used
Algorithm [25]	Medical image security	Zero-watermarking	No change	DWT-DFT
Algorithm [26]	Medical image security	Zero-watermarking	No change	DCT
The proposed algorithm	Medical image security	Zero-watermarking	No change	VGG19-DFT

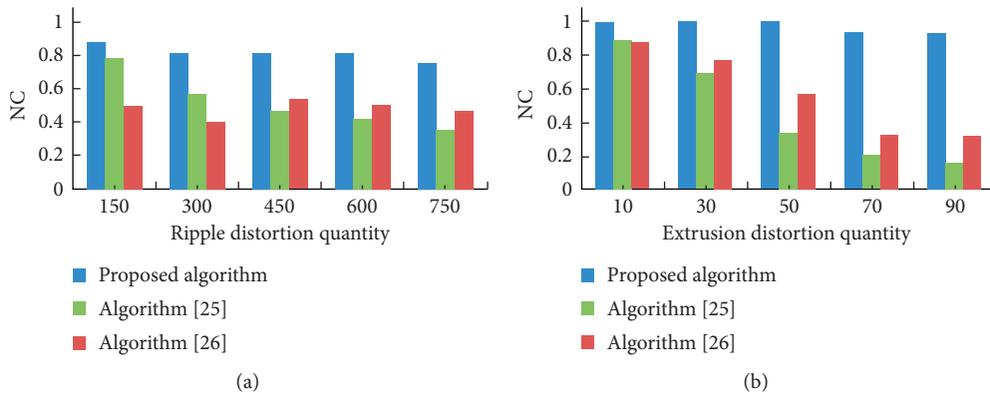


FIGURE 11: Continued.

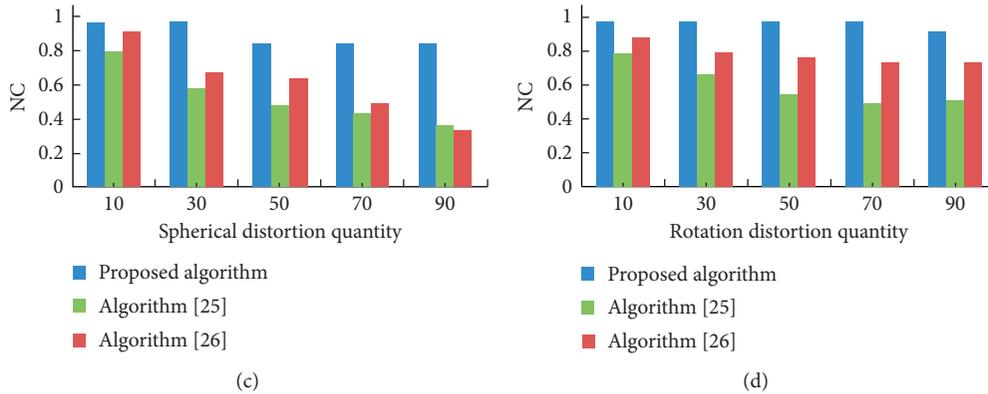


FIGURE 11: Robustness comparison between the proposed algorithm and two algorithms [25, 26] against four types of local nonlinear geometric attacks. (a) Ripple distortion attack. (b) Extrusion distortion attack. (c) Spherical distortion attack. (d) Rotation distortion attack.

ability of the proposed algorithm to resist ripple distortion attack is better than the other two algorithms [25, 26]. The comparison of the experimental results of the extrusion distortion attack is shown in Figure 11(b). Figure 11(b) shows that, with the increase of extrusion distortion quantity, the NC value of the algorithm is close to 1, which is obviously higher than that of the other two algorithms [25, 26]. Especially when the extrusion distortion quantity is more than 70%, the difference of NC value is more significant. Figure 11(c) shows the comparison under the spherical distortion attack. The NC values of the proposed algorithms are all greater than 0.87. This proves that the proposed algorithm is more resistant to the spherical distortion attack. It can be observed from Figure 11(d) that the three algorithms can effectively resist the rotation distortion attack, but the proposed algorithm has a stronger ability.

With the above comparison and analysis of the four local nonlinear geometric attacks, the proposed algorithm is obviously stronger than the other two algorithms in resisting local nonlinear geometric attacks. Therefore, the proposed algorithm has good robustness and can well resist local nonlinear geometric attacks.

## 6. Conclusions

In recent years, the medical image watermarking algorithm against geometric attacks has been a hot and difficult topic in the research of robust watermarking technology. In this paper, a zero-watermarking algorithm based on VGG19 is designed to resist local nonlinear geometric attacks. VGG19 is used to extract deep features of the medical image, and two-dimensional discrete Fourier transform and mean-perceptual hashing algorithm are used to generate zero-watermarking. The design process of the algorithm combines the concepts of deep convolution neural network, Fourier transform, perceptual hashing, cryptography, and zero-watermarking, which solves the problem of watermarking resisting local nonlinear geometric attacks. At the same time, the scrambling of the watermarking image

ensures the security of the algorithm. It has high practical value for medical information protection.

## Data Availability

The data used to 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 regarding this work.

## Acknowledgments

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (no. KJQN201800442) and the General Project of Chongqing Natural Science Foundation (no. cstc2020jcyj-msxmX0422).

## References

- [1] A. Anand and A. K. Singh, "An improved DWT-SVD domain watermarking for medical information security," *Computer Communications*, vol. 152, pp. 72–80, 2020.
- [2] J. J. Garcia-Hernandez, W. Gomez-Flores, and J. Rubio-Loyola, "Analysis of the impact of digital watermarking on computer-aided diagnosis in medical imaging," *Computers in Biology and Medicine*, vol. 68, pp. 37–48, 2016.
- [3] T.-Y. Fan, H.-C. Chao, and B.-C. Chieu, "Lossless medical image watermarking method based on significant difference of cellular automata transform coefficient," *Signal Processing: Image Communication*, vol. 70, pp. 174–183, 2019.
- [4] Z. Ali, M. Imran, M. Alsulaiman, M. Shoaib, and S. Ullah, "Chaos-based robust method of zero-watermarking for medical signals," *Future Generation Computer Systems*, vol. 88, pp. 400–412, 2018.
- [5] X. Wang, L. Wan, M. Huang, C. Shen, Z. Han, and T. Zhu, "Low-complexity channel estimation for circular and non-circular signals in virtual MIMO vehicle communication

- systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3916–3928, 2020.
- [6] V. Natarajan, "Hybrid local prediction error-based difference expansion reversible watermarking for medical images," *Computers & Electrical Engineering*, vol. 53, pp. 333–345, 2016.
- [7] Y. Gangadhar, V. S. Giridhar Akula, and P. C. Reddy, "An evolutionary programming approach for securing medical images using watermarking scheme in invariant discrete wavelet transformation," *Biomedical Signal Processing and Control*, vol. 43, pp. 31–40, 2018.
- [8] A. Sharma, A. K. Singh, and S. P. Ghreera, "Secure hybrid robust watermarking technique for medical images," *Procedia Computer Science*, vol. 70, pp. 778–784, 2015.
- [9] D. Bouslimi and G. Coatrieux, "A crypto-watermarking system for ensuring reliability control and traceability of medical images," *Signal Processing: Image Communication*, vol. 47, pp. 160–169, 2016.
- [10] P. Garg and A. K. Jain, "An invisible based watermarking technique for biometric image authentication," *Materials Today: Proceedings*, 2020.
- [11] M. Vazhora Malayil and M. Vedhanayagam, "A novel image scaling based reversible watermarking scheme for secure medical image transmission," *ISA Transactions*, vol. 108, pp. 269–281, 2021.
- [12] Z. Narima, K. Amine, K. Redouane et al., "A DWT-SVD based robust digital watermarking for medical image security," *Forensic Science International*, vol. 320, Article ID 110691, 2021.
- [13] M. Favorskaya, E. Savchina, and K. Gusev, "Feature-based synchronization correction for multilevel watermarking of medical images," *Procedia Computer Science*, vol. 159, pp. 1267–1276, 2019.
- [14] A. F. Qasim, F. Meziane, and R. Aspin, "Digital watermarking: applicability for developing trust in medical imaging workflows state of the art review," *Computer Science Review*, vol. 27, pp. 45–60, 2018.
- [15] K. Swaraja, K. Meenakshi, and P. Kora, "An optimized blind dual medical image watermarking framework for tamper localization and content authentication in secured telemedicine," *Biomedical Signal Processing and Control*, vol. 55, Article ID 101665, 2020.
- [16] M. Cedillo-Hernandez, A. Cedillo-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, "Improving the management of medical imaging by using robust and secure dual watermarking," *Biomedical Signal Processing and Control*, vol. 56, Article ID 101695, 2020.
- [17] F. Qin, J. Li, H. Li, J. Liu, S. A. Nawaz, and Y. Liu, "A robust zero-watermarking algorithm for medical images using curvelet-dct and RSA pseudo-random sequences," in *Proceedings of the International Conference on Artificial Intelligence and Security*, pp. 179–190, Hohhot, China, July 2020.
- [18] J. Liu, J. Ma, J. Li, M. Huang, N. Sadiq, and Y. Ai, "Robust watermarking algorithm for medical volume data in internet of medical things," *IEEE Access*, vol. 8, pp. 93939–93961, 2020.
- [19] A. Roek, M. Javorník, K. Slaviček et al., "Zero watermarking: critical analysis of its role in current medical imaging," *Journal of Digital Imaging*, vol. 34, pp. 1–8, 2021.
- [20] Z. Xia, X. Wang, C. Wang et al., "Local quaternion polar harmonic Fourier moments-based multiple zero-watermarking scheme for color medical images," *Knowledge-Based Systems*, vol. 216, Article ID 106568, 2020.
- [21] D. Meng, X. Wang, M. Huang, L. Wan, and B. Zhang, "Robust weighted subspace fitting for DOA estimation via block sparse recovery," *IEEE Communications Letters*, vol. 24, no. 3, pp. 563–567, 2020.
- [22] A. Singh and M. K. Dutta, "A robust zero-watermarking scheme for tele-ophthalmological applications," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 8, pp. 895–908, 2020.
- [23] A. Fierro-Radilla, M. Nakano-Miyatake, M. Cedillo-Hernandez et al., "A robust image zero-watermarking using convolutional neural networks," in *Proceedings of the 2019 7th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–5, Cancun, Mexico, May 2019.
- [24] B. Han and J. Li, "A robust watermarking algorithm for medical volume data based on hermite chaotic neural network," *International Journal of Applied Mathematics and Statistics*, vol. 48, no. 18, pp. 128–135, 2013.
- [25] J. Li and Y. Liu, "The medical image watermarking algorithm based on DFT and logistic map," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, pp. 337–340, Hangzhou, China, March 2013.
- [26] M. Sui and J. Li, "Robust watermarking for medical images based on Arnold scrambling and DCT," *Application Research of Computer*, vol. 30, no. 8, pp. 2552–2556, 2013.

## Research Article

# Service Migration Policy Optimization considering User Mobility for E-Healthcare Applications

Xuhui Zhao , Jianguhui Liu , Baofeng Ji , and Lin Wang 

Information Engineering College, Henan University of Science and Technology, Luoyang 471000, China

Correspondence should be addressed to Xuhui Zhao; haustzxh@163.com

Received 30 March 2021; Accepted 8 June 2021; Published 21 June 2021

Academic Editor: Han Wang

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

Mobile edge computing (MEC) is an emerging technology that provides cloud services at the edge of network to enable latency-critical and resource-intensive E-healthcare applications. User mobility is common in MEC. User mobility can result in an interruption of ongoing edge services and a dramatic drop in quality of service. Service migration has a great potential to address the issues and brings inevitable cost for the system. In this paper, we propose a service migration solution based on migration zone and formulate service migration cost with a comprehensive model that captures the key challenges. Then, we formulate service migration problem into Markov decision process to obtain optimal service migration policies that decide where to migrate in a limited area. We propose three algorithms to resolve the optimization problem given by the formulated model. Finally, we demonstrate the performance of our proposed algorithms by carrying out extensive experiments. We show that the proposed service migration approach reduces the total cost by up to 3 times compared to no migration and outperforms the general solution in terms of the total expected reward.

## 1. Introduction

With the development of Internet of Things, E-healthcare applications are emerging and gaining popularity. These E-healthcare applications require low latency and intensive resource (computation or storage resource). However, IoT terminals (sensors or vehicles) have limited resources (low processing speed, small storage capacity, and limited battery life), which is insufficient to complete these applications. To mitigate this limitation, mobile edge computing (MEC) is envisioned as a promising paradigm by providing cloud service at the edge of network [1, 2]. By leveraging cloud service deployed near mobile terminals, the latency of IoT applications can be significantly guaranteed and the quality of experience of mobile user can be greatly improved. Suppose a patient travelling in a vehicle needs monitoring the health parameters continuously such as blood pressure and pulse rate, which are collected by IoT devices and the raw health data are sent to the edge service site through the client E-healthcare application. The edge site processes the data based on the predeployed E-healthcare application

and sends the processed result to the client application. Thus, MEC is assumed a key enabling technology for realizing the E-healthcare visions.

User mobility is common in MEC systems. The edge sites often place cloud resources on the edge access devices such as base stations (BSs) in cellular system and access points (AP) in Wi-Fi. Network virtualization allows heterogeneous networks to coexist on a shared infrastructure, and, hence, it can be considered as a solution for the heterogeneity in MEC [3, 4]. On the other hand, user mobility can result in interruption of ongoing edge services and dramatic drop in QoS, and service migration (SM) has a great potential to address the issues [5, 6]. SM policy should be made such that whether and where to migrate the ongoing edge services as an arbitrary user moves outside the service area of the associated edge cloud site, which involves multiple challenges: first, a user may connect to one edge cloud in one time slot and switch to another in the next. Consequently, communication resources (bandwidth) and computing resources (virtual machine size) may be reconfigured in each time slot, which incur reconfiguration cost due to heterogeneity in

edge cloud. Second, it is hard to predict how each user will move over time in some scenarios, especially in vehicle networks. Under the circumstances with no assumption about user mobility, migration policy still needs to be made in order to ensure service continuity. Third, from the perspective of communication resources, it is most advantageous to migrate the cloud service to the location closest to the user; however, from the perspective of computing resources, this approach has been demonstrated to be inadequate due to the dynamic workloads of the network [7–9].

Despite the extensive existing research on mobility-induced service migration in the edge cloud context in general, fine grained and random mobility features are not considered [5, 8–10], and most of them often assume prior knowledge about user mobility [11–14]. In addition, migration policy design in the research only considers whether and where to migrate. And edge cloud service is always migrated to the cloud server closest to the user. When mobile users have multiple access edge clouds, where to migrate should be considered extensively in terms of fine-grained mobility. The transmission cost and migration cost have been well considered in the existing research; nevertheless, none of them considers the influence of user mobility on reconfiguration induced by service migration.

Based on the above observations, we study the user-mobility-driven optimization of the resource usage, reconfiguration, and migration altogether in edge clouds without mobility model assumption. We use the MDP framework to study service migration jointly considering user mobility and heterogeneity in edge clouds. We make three contributions:

- (1) Our formulation captures multiple types of important costs. The first is resource usage cost associated with service location depending on resource price strategy for the MEC system. The second is the reconfiguration cost accounts for changes in resources. Network reconfiguration in virtualization incurs a cost for the system by remapping a subset of virtual nodes or links to better align the allocation of resources to current network conditions. The third is migration time cost for guaranteeing quality of service. We pursue the optimization of the total cost while serving each user's workload from the perspective of edge cloud operator.
- (2) We provide a mathematical framework to design optimal service migration policies that decide where to migrate without assumption of user mobility in a limited area. We formulate service migration problem into Markov decision process (MDP). We propose Service Migration algorithm based on Value Iteration (SMVI), In-place Value Iteration (SMIVI), and Policy Iteration (SMPI) to resolve the optimization problem given by the above cost model.
- (3) We demonstrate the performance of our proposed algorithms by carrying out extensive experiments. We compare the proposed policy with several baseline strategies that include no migration, migration without considering reconfiguration cost. It is shown that the proposed approach offers

significant gains over those baseline approaches. The results show that, up to 3 times compared to no migration, up to 5% reduction on the total cost can be achieved compared to the approaches typically employed in edge clouds.

The remainder of the paper is organized as follows. In the next section, we review some related service migration in MEC and their problems. In Section 3, we describe the system model and the basic assumptions considered in this paper. Then, we formulate the problem of designing an optimal service migration policy as MDP. We detail the proposed algorithm in Section 4. To validate the proposed study, we provide numerical experiments under various settings in Section 5. Finally, we draw the conclusions in Section 6.

## 2. Related Work

Induced-mobility service migration has been paid much attention in MEC system. The related research work contains optimization of migration policy and execution of migration.

*2.1. Optimization of Migration Policies.* Migration can bring benefits and incur computation and communication resource overheads. Therefore, the decision on whether and where to migrate depends on many aspects, such as user mobility, the number of simultaneous migrations, resource availability at heterogeneous edge clouds, and so on, which is a sophisticated optimization problem. The study in [8] aimed to optimize service migration policy aiming at avoiding the migration if the stable delay is acceptable. They assumed that a required end-to-end latency threshold is known in advance. Meanwhile, they do not consider the workloads of destination edge cloud as well as their reconfiguration costs for the service. The study in [9] proposed a service migration policy to decide the services that are run on the mobile edge nodes to be migrated. This problem is solved by a Lyapunov-based approach. The optimal migration strategy is obtained in closed form, and an online algorithm outputs the number of service replicas (how many) and where these services are to be migrated prior to the handover execution. The proposed mathematical models are based on simple assumptions and thus cannot cope with complex mobility condition and user context information. The authors of [5, 11, 12] studied service migration design under the one-dimensional and two-dimensional mobility model, and the study in [11] proposed stochastic frameworks for dynamic workload migration based on the Lyapunov optimization technique. And references [5, 12] also did not consider the reconfiguration costs for the service, which cannot be directly used in heterogeneity in edge cloud network. Although the study in [11] considered the reconfiguration cost in the problem formulation, it is only assumed that the reconfiguration cost is the function of distance. However, the parameter in the formulation was not easy to determine in practice. In addition, all of them either require statistics information or model on user mobility.

The realization of cloud service depends on virtualization technology represented by virtual machine (VM). The authors of [15] proposed strategies to find suitable bandwidth and precopy iteration count to optimize different performance metrics of VM migration over a WAN. The formulated models were to optimize network resource consumption, migration duration. And a strategy was proposed to determine appropriate the migration bandwidth and number of precopy iterations, and it performed numerical experiments with large number of migration requests. The authors of [16] studied the problem of optimal VM placement and migration to minimize resource usage and power consumption in a data centre. The formulated problem was optimized as a joint multiple objective function and solved by leveraging the framework of convex optimization. A multilevel join VM placement and migration algorithms provided an approximate optimal solution. However, the above work considered the centralized cloud environment instead of edge clouds. In the edge cloud environment, mobile terminals are diversified, and the mobility of different mobile terminals is more complex and has a certain degree of randomness.

Our model captures multiple types of important costs. The first is transmission cost, depending on service location. Edge cloud is usually collocated with an access point (AP, a cellular base station, or a Wi-Fi access point); therefore, different cloud resources may have different prices controlled by resources price strategy for the MEC system [17]. The second is the reconfiguration cost accounts for changes in resources. Although network virtualization is considered as the most promising approach to realize heterogeneous networks to coexist on a shared infrastructure, network reconfiguration in virtualization incurs a cost for the system by remapping a subset of virtual nodes or links to better align the allocation of resources to current network conditions [18]. This reconfiguration cost cannot be neglected to improve the total net gain for the edge cloud service provider [3]. The third is migration time cost. Specifically, the factors affecting VM migration time are the memory size, the number of concurrent VM on destination server, and the available network bandwidth [19]. Without any knowledge on user mobility, our algorithm makes optimal migration decisions while guaranteeing system reward.

**2.2. Execution of Migration.** Execution of migration focuses on how to efficiently execute live migration in a practical MEC system. The framework proposed in [20] aimed at smooth migration of all or only a required portion of an ongoing IP service between a data centre and user equipment. The service migration and continuity were supported by replacing IP addressing with service identification. Recent efforts toward the implementation of service migration in MEC environments have focused on VM migration [21–25]. The study in [21] proposed a technique that initiates the migration of VMs between heterogeneous cloud environments. The author demonstrates heterogeneous VM migration between various cloud platforms built on different architectures. The study in [22] presented a three-layer

framework for migrating active service applications that are encapsulated either in VMs or in containers. It aims at maintaining relatively low service downtime and overall migration time. Especially, the framework applied to both virtual machines and containers. The authors of [23] proposed Follow Me Fog, a framework supporting a new seamless handover timing scheme among different computation access points by designing a job premigration mechanism. The authors of [24] proposed remote loading and redirection to accelerate the service migration. By tracing historic access patterns, the proposed method first generates a loading request list that locates the core codes in the image file of service applications for booting. The study in [25] gave an overview of VM migration and discussed both its benefits and challenges, the studies regarding linking VM migration to user mobility are summarized as well, and further optimizations on live VM migration were listed in the paper. In conclusion, the above research provides a solid research foundation for the implementation of service migration strategy proposed in this paper.

### 3. Model Formulation

Considering the scenario as shown in Figure 1, an edge server contains one or more physical machines hosting several VMs, covering the mobile users in proximity. These edge servers are interconnected with each other via different kinds of network connections supported by network virtualization reconfiguration [3, 18].

Note that we use edge server or edge cloud site as a general term. In our system model, there are  $N$  edge servers that are virtualized and managed by a SM Management Entity (SMME). The edge cloud pointed by red or green arrows denotes the available access edge clouds, and the black one-way arrow indicated the ambulance moved from the current position to another position.

Migration policy is triggered whenever an ambulance is in its migration zone in an access point, as shown in Figure 2. SMME should decide which edge cloud to migrate when mobile users have multiple access edge clouds. In this work, this zone is determined based on a fixed distance to the AP or dynamic distances based on the user's speed and VM size. We assume that the ambulance moves around in a limited area in the city.

We utilize access point A and MEC-A as the source edge site and access point B and MEC-B as the destination edge site. Handover management entity (HME) is used to implement handover policy and handover procedure. The whole mechanism of service migration can be described as follows, which is shown in Figure 3.

Step 1: a patient travelling in a vehicle accesses the E-healthcare application hosted by MEC-A.

Step 2: the access point A sends the position information of the related vehicle to the handover management entity.

Step 3: when the vehicle enters the migration zone of an access point, the HME implements handover

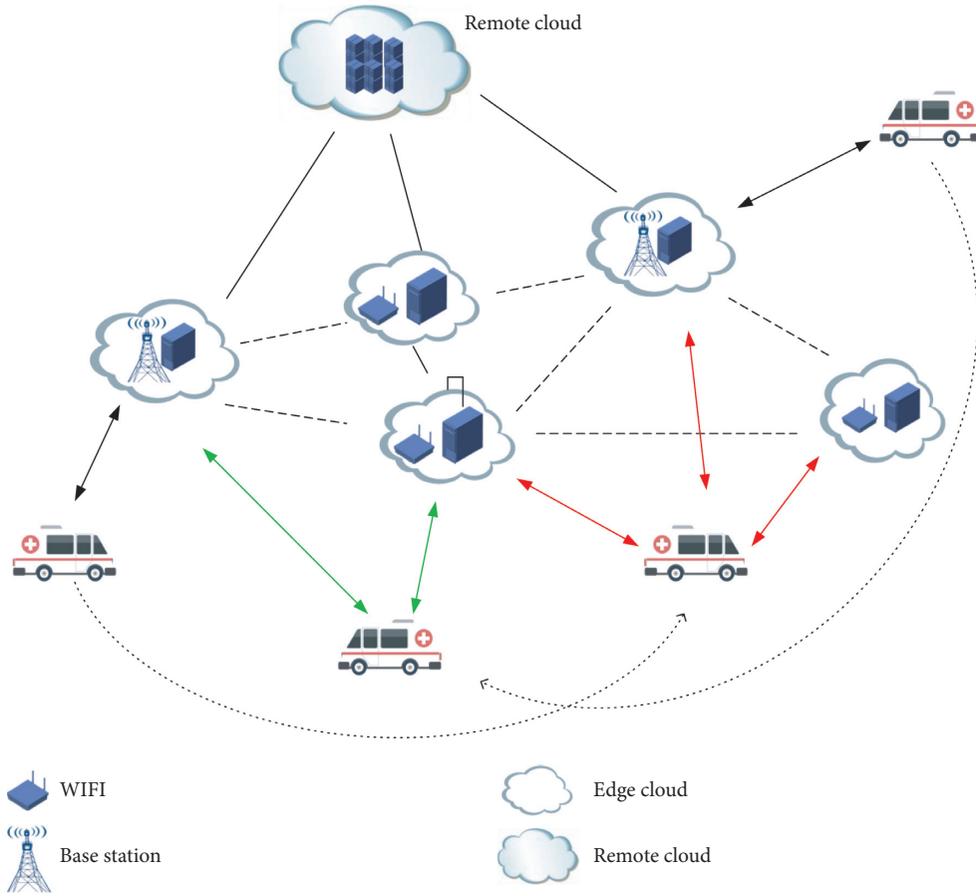


FIGURE 1: A case of service migration in mobile edge computing. The edge cloud pointed by red or green arrows denotes the available access edge clouds, and the black one-way arrow indicates that the ambulance moved from the current position to another.

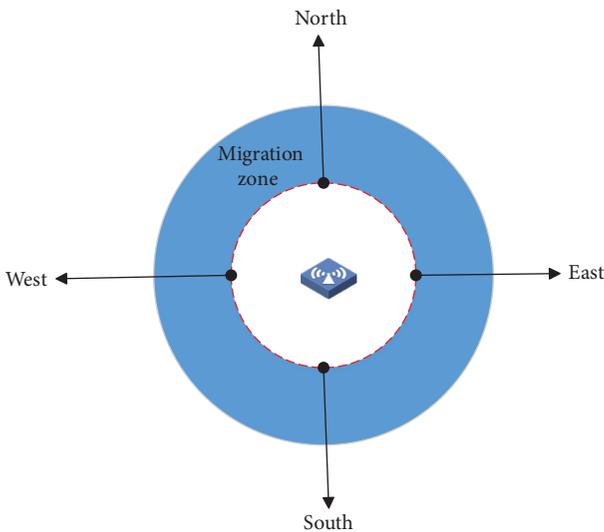


FIGURE 2: Migration zone.

procedure and informs the SMME to make a decision about service migration.

Step 4: the SMME implements service migration policy by our proposed approach, the result of implementing our algorithm is supposed as MEC-B, and then the

SMME informs the MEC-A to implement service migration procedure.

Step 5: the service migration is implemented between MEC-A and MEC-C.

Step 6: the vehicle enters the coverage of access point B.

Step 7: the E-healthcare application service is resumed in the MEC-B.

Step 8: the service migration is completed.

Step 9: all the resources are released including bandwidth, storage, and computation resources hosted by MEC-A and MEC-B.

In the following sections, we describe our model for each component in an edge cloud system; then, we formulate the service migration problem.

**3.1. Edge Cloud System.** We consider an edge cloud system (e.g.,  $N = 5 \times 5$ ) where a vehicle accesses a cloud-based service hosted on the MECs. The set of possible edge cloud locations is given by  $l = (i, j)$ ,  $l \in L$ , where  $L$  is assumed to be finite. An edge cloud is defined as a pool of virtualized computing resources, which is usually collocated with a cellular base station, a Wi-Fi access point, or Mobile Cloudlets. The maximum workload of an edge cloud  $l$  is

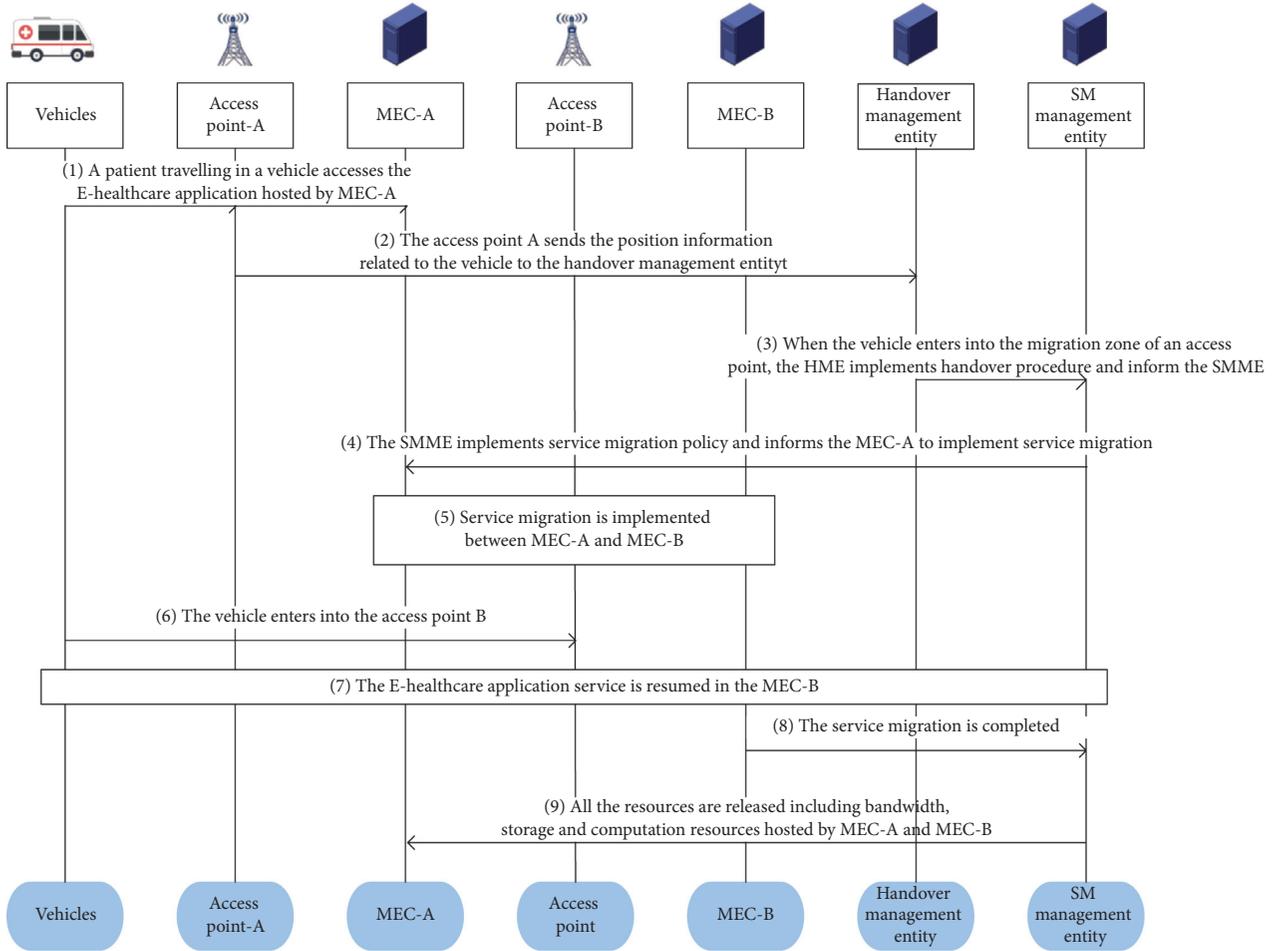


FIGURE 3: The whole mechanism of service migration.

defined as  $c_{i,j}$ . These edge servers are interconnected with each other. The network bandwidth between two edge clouds  $l$  and  $l'$  is given by  $B(l, l')$ . The number of concurrent running workloads of all the edge clouds is denoted by the set  $W = \{w_{ij}\}, \forall (i, j) \in L$ .

**3.2. Mobile Terminals' Mobility and Migration.** We assume an ambulance  $u$  (also called mobile terminal) with IoT devices move around in the limited area that is shown as grids in Figure 2. In a certain time slot, the ambulance connects to an edge cloud  $l \in L$  that covers the vicinity of the ambulance and accesses the E-healthcare application service, incurring an allocated bandwidth  $x_u$  and the number of CPU processing cycles  $m_u$ , and the corresponding virtual machine state data size is denoted by  $d_u$ . When the ambulance leaves the limited coverage of the current edge cloud, the corresponding cloud service can be migrated to one of the available neighbour edge clouds for the purpose of cloud service continuity. As shown in Figure 4, the cloud service at the edge of the grid can migrate to the surrounding two neighbour clouds, while the cloud service at other locations plotted by red line can be migrated to the surrounding four neighbours. For example, edge cloud services in the location of edge cloud

(0, 0) can be migrated to (0, 1) or (1, 0). As another example, services in the edge cloud (2, 2) can be migrated to (1, 2), (3, 2), (2, 1), or (2, 3). SM management entity should decide which edge cloud to migrate as an ambulance moves around in the limited area.

**3.3. Costs.** We consider three aspects of costs in the edge cloud system: the edge cloud usage cost, the migration cost, and the reconfiguration cost. These costs are able to represent the most prominent expenditure from the perspective of the cloud operator.

**3.3.1. Resource Usage Cost.** In this paper, we assume that the resource usage cost refers to bandwidth incurring by data transmission and computation resources usage incurring by task processing, which can be expressed as

$$\text{Cost}_{\text{usa}} = a_l * x^t + b_l * \frac{m^t}{100}, \quad (1)$$

where  $x^t$  denotes the bandwidth at current location  $l$  at time slot  $t$ ,  $x^t < \lambda_u$  and  $a_l$  denotes the unit cost of bandwidth resource at location  $l$ .  $m^t$  measures the number of CPU processing cycles at time slot  $t$ .  $b_l$  represents the cost of using 100 MHz CPU.

(0, 0)	(0, 1)	(0, 2)	(0, 3)	(0, 4)
(1, 0)	(1, 1)	(1, 2)	(1, 3)	(1, 4)
(2, 0)	(2, 1)	(2, 2)	(2, 3)	(2, 4)
(3, 0)	(3, 1)	(3, 2)	(3, 3)	(3, 4)
(4, 0)	(4, 1)	(4, 2)	(4, 3)	(4, 4)

FIGURE 4: Mobile terminals move around in a limited area.

**3.3.2. Reconfiguration Cost.** As an ambulance move around in the limited area, the SMME may migrate the workload from source edge cloud to destination edge cloud decided by migration policy, which results in adapting the amount of resources allocated in destination edge cloud. Such adaptation depends on the number of reallocated virtual links and nodes, which would incur some inevitable cost for preparing the resources. The reconfiguration cost of allocating 100 MHz CPU and 1 Mbps bandwidth is denoted by  $p_l$  for edge cloud  $l$ ; the reconfiguration cost is calculated as

$$\text{Cos } t_{\text{rec}} = p_l \left[ (x^t - x^{t-1}) + \frac{(m^t - m^{t-1})}{100} \right]. \quad (2)$$

**3.3.3. Migration Cost.** This cost represents the delay when we migrate the workload from one edge cloud to another due to data movement. Specifically, the major factors impacting VM migration time are the memory size, memory dirtying rate of the VM to be migrated, and the available network bandwidth, as well as the number of concurrent running VM at the destination server [16]. For simplicity, we approximately calculate the migration time as

$$\text{Cost}_{\text{mig}} = d_u^t * \frac{w_{l'}}{B_{l,l'}}, \quad (3)$$

where  $d_u^t$  denotes the virtual machine's data size at time slot  $t$ ,  $w_{l'}$  denotes the number of concurrent running VM in destination cloud  $l'$ , and  $B_{l,l'}$  denotes network bandwidth between edge cloud  $l$  and  $l'$ .

**3.4. Problem Formulation.** By combining the essential models above, we formulate the migration policy as MDP. In order to obtain the optimal policy, it is necessary to identify the actions, state, and reward functions in our mathematical model, which is given in the following sections.

**3.4.1. System State.** Let  $S^t$  denote the system state at time  $t$ , defined by  $(l^t, o^t)$ , where  $l^t$  is the locations set of cloud service at time  $t$ , and  $o^t$  denotes the workloads at the current location  $l^t$ , where  $o^t = (x_u^t, m_u^t, d_u^t)$ . Consequently, the state vector can be described as  $S^t = (l^t, o^t)$ . Here,  $l^t = (i, j)$ , for  $\forall i < N, j < N$ .

**3.4.2. System Action.** In the system, the SMME decides where to migrate the edge cloud service when a vehicle moves around in the limited area. It is not optimal to migrate the service to a location that is farther away from the user, as one would intuitively expect. Therefore, we assume that when the vehicle is in the migration zone in the current edge cloud, the corresponding cloud service can be migrated to one of the available neighbour edge clouds. The current action is represented by  $a$ ,  $a \in A$ ,  $A = \{0, 1, 2, 3\}$ , where 0 represents migration to the south, 1 represents migration to the west, 2 represents migration to the north, and 3 represents migration to the east.

**3.4.3. Reward Function.** Our goal is to design a policy that uses the E-healthcare application's workload (bandwidth, or CPU consumption) and service location as input, and the policy continuously decides where to migrate the workload for each vehicle as it moves, such that the total cost is minimized over time. Note that the reward function in MDP is the maximum value; we set three mapping variables  $\text{Cos } t'_{\text{usa}}$ ,  $\text{Cos } t'_{\text{mig}}$ , and  $\text{Cos } t'_{\text{rec}}$ , corresponding to the three original variables. Therefore, the reward function is the weighted sum of the resource usage cost, migration cost, and reconfiguration cost, as given by

$$C^t = \alpha * \text{Cos } t'_{\text{usa}} + \beta * \text{Cos } t'_{\text{rec}} + \chi * \text{Cos } t'_{\text{mig}}, \quad (4)$$

where  $\alpha$ ,  $\beta$ , and  $\chi$  are the weights of the three types of cost, respectively.  $\text{Cos } t'_{\text{usa}}$ ,  $\text{Cos } t'_{\text{mig}}$ ,  $\text{Cos } t'_{\text{rec}}$  denotes the corresponding cost after being normalized, and  $\min()$  and  $\max()$  are minimum and maximum functions of a variable.

$$\text{Cos } t'_{\text{usa}} = 1 - \frac{\text{cos } t_{\text{usa}} - \min(\text{cos } t_{\text{usa}})}{\max(\text{cos } t_{\text{usa}}) - \min(\text{cos } t_{\text{usa}})},$$

$$\text{Cos } t'_{\text{rec}} = 1 - \frac{\text{cos } t_{\text{rec}} - \min(\text{cos } t_{\text{rec}})}{\max(\text{cos } t_{\text{rec}}) - \min(\text{cos } t_{\text{rec}})}, \quad (5)$$

$$\text{Cos } t'_{\text{mig}} = 1 - \frac{\text{cos } t_{\text{mig}} - \min(\text{cos } t_{\text{mig}})}{\max(\text{cos } t_{\text{mig}}) - \min(\text{cos } t_{\text{mig}})}.$$

Take expectation with respect to service migration costs in each epoch over the randomized network states  $S^t$  and the actions  $A^t$  induced by a given control policy  $\pi$ . This action causes the system to transition to a new intermediate state  $S^{t+1} = (l^{t+1}, o^{t+1})$ . The expected long-term reward of the MEC conditioned on an initial network state  $S^1$  can be expressed as

$$V(S, \pi) = E_{\pi} \left[ (1 - \gamma) \sum_{t=1}^{\infty} (\gamma)^{t-1} C^t | S^1 = S \right], \quad (6)$$

where  $\gamma \in [0, 1)$  is the discount factor and  $(\gamma)^{t-1}$  denotes the discount factor to the  $(t-1)$ -th power. The objective of the MEC system is to design an optimal control policy  $\pi^*$  that maximizes  $V(S, \pi)$ , for any given initial network state  $S$ , which can be formally formulated as

$$\pi^* = \arg \max_{\pi} V(S, \pi). \quad (7)$$

We denote  $V(S) = V(S, \pi^*)$  as the optimal state-value function.

#### 4. Solving the Optimal Control Policy

Note that standard approaches of solving for the optimal policy of an MDP include value iteration and policy iteration [26]. To derive the service migration policy, we propose SMVI and SMPI algorithm, respectively. The details can be described as follows.

The SMPI algorithm (Algorithm 1) contains two parts: policy evaluation (Algorithm 2) and policy improvement (Algorithm 3). Firstly, the V-function and the policy are initialized randomly. Then, for the current policy, the value function of the state  $s$  under the policy is estimated by the algorithm of iterative policy evaluation, and a new strategy is obtained by greedy policy improvement. The above procedure is looped until the policy remains unchanged. The pseudo codes of the algorithms are as follows:

In order to find the optimal strategy, the SMVI algorithm (Algorithm 4) uses the Bellman optimal equation to iterate. However, there is no explicit strategy in the update process, so the value function in the iteration process may not correspond to any strategy. Additionally, we propose service migration algorithm based on the In-place Value Iterative (SMIVI); the difference between the in-place value iterative method and the general value iterative method is asynchronous update in the algorithm. That is to say, the V update is directly performed during the update process. The marked line with 16 in Algorithm 4 is replaced by sentence  $V[i, j] = np.max(qs)$ , which generate the SMIVI algorithm. The pseudo code of SMVI algorithm is as follows:

**4.1. Evaluation.** The number of actions in MDP formulated by service migration problem is four, because the edge cloud service can only be migrated to the four available neighbour edge clouds. The variable  $N\_action$  is assumed as a constant in all the algorithms. Therefore, the execution time of all the algorithms depends on the execution times implemented by the body code of “for” loop, the time complexity of Algorithm 1, Algorithm 2, Algorithm 3, Algorithm 4 are  $O(n)$ ,  $O(n^2)$ ,  $O(n)$ ,  $O(n^2)$ , respectively.

We validate the performance of the proposed service migration algorithm based MDP in Python. We use PyCharm as the Python integrated development environment. All the experiments are implemented on a computer with Intel Core i5-4200U CPU 2.3 GHz and 4096 MB RAM. We conduct extensive experiments and report the experimental results in this section. In addition to the proposed algorithm, other schemes are taken as the performance reference for comparison on the same topology with the same parameter setting, which are described as follows.

**4.1.1. No Migration.** No matter how the vehicle moves, the system will not take the migration action.

**4.1.2. Other Schemes [5, 12].** The migration policy does not consider the reconfiguration cost.

All the measurements were performed on a computer equipped with Intel Core i5-6500 CPU (3.2 GHz) and 8.0 GB RAM.

#### 5. Experimental Settings

We consider an edge cloud system deployed with 25 edge clouds, as shown in the grid of Figure 2; the allocated bandwidth at different locations follows uniform distributions over (4, 100). The available network bandwidth between the source and destination edge servers was between 20 and 100 MBps. The number of concurrent running VM in each edge cloud server follows random distribution over (1, 25). The VM size was between 800 MB and 1700 MB [27]. The reconfiguration price varies among different edge clouds. We generate the reconfiguration prices following a Gauss distribution with the mean 20-euro and the standard deviation 5-euro. We categorize all the edge clouds into three clusters, each of which is subscribed to one of three bandwidth prices provided by Internet providers. The average prices are 2.49 euro/MBps, 4.86 euro/MBps, and 1.25 euro/MBps, respectively [28].

**5.1. The Optimal Migration Policy under Different Scenarios.** As mentioned before  $\alpha$ ,  $\beta$ , and  $\omega$  denote the weights of the three types of cost, respectively. Three parameter settings are envisioned:

- (1)  $\alpha = 0.8$ ,  $\beta = 0.1$ , and  $\chi = 0.1$  which represents a high resource usage cost compared to migration cost and reconfigure cost if a service migration is launched.
- (2)  $\alpha = 0.1$ ,  $\beta = 0.8$ , and  $\chi = 0.1$  which represents a high migration time cost compared to resource usage cost and reconfigure cost if a service migration is launched.
- (3)  $\alpha = 0.1$ ,  $\beta = 0.1$ , and  $\chi = 0.8$  which represents a reconfiguration cost compared to resource usage cost and migration cost if a service migration is launched.

Figures 5–7 illustrate the optimal policy constructions for the three above-mentioned settings. Arrows in the figure indicate the optimal migration direction. For instance, when  $\alpha = 0.1$ ,  $\beta = 0.1$ ,  $\chi = 0.8$ , the optimal policy recommends service in the location (0, 0) migrating to the east-(0, 1). However, when  $\alpha = 0.1$ ,  $\beta = 0.8$ ,  $\chi = 0.1$ , the optimal policy recommends the same service migrating to the south-(1, 0). This is because reconfiguration cost incurring by migration to the location (0, 1) is greater than that of location (1, 0). The optimal policy recommends migration to the south for increasing total reward. Note that the weight parameters have a great impact on the optimal policy construction, since migration to different location incurs different migration cost and reconfiguration cost. This difference is not negligible in comparison to the achieved gain when migrating a service.

```

(1) env = Env ()//Environment Initialization: MDP mode state  $s$ , action  $a$ , and reward  $r$ 
(2) initialize random policy
(3) for ( $i = 1, i ++, i < \text{max\_iter}$ )
(4)    $V = \text{value\_evaluate}(\text{policy}, \text{env}, \text{max\_step}, \text{tol})$ //evaluate the random policy
(5)    $\text{policy} = \text{policy\_improvement}(\text{env}, V)$ //improve the policy
(6)    $\text{mean\_values.append}(\text{np.mean}(V))$ //store mean value of the policy
(7)    $\text{run\_times.append}(\text{time.time}() - \text{st})$ //store run time
(8)   if  $\text{last\_V}$  is not None and  $\text{np.sum}(\text{np.abs}(V - \text{last\_V})) < \text{tol}$ :
(9)     break
(10)   $\text{last\_V} = V$ //the value function update is small enough, it will stop.
End for
(11) return  $V, \text{mean\_values}, \text{policy}, \text{run\_times}$ # return state value, mean value, the optimal policy and run time

```

ALGORITHM 1: SMPI (policy, env, max\_step = 100, tol = 1e-6).

```

(1) initialization  $V$ 
(2)   for  $i$  in range (max_step):
(3)      $\text{new\_V} = V.\text{copy}()$ 
(4)     for all  $s$  in  $S$ : # for every state, update the value function
(5)        $qs = \text{np.zeros}((N\_ACTIONS), \text{dtype} = \text{np.float32})$ //Initialize value function
(6)       for  $a$  in range (N_ACTIONS): //store the Q value for each action
(7)          $n\_s = \text{env.P}[s, a]$ 
(8)          $r = \text{env.R}[s, a]$ 
(9)          $n\_V = V[n\_s[0], n\_s[1]]$ 
(10)         $qs[a] = r + \text{gamma} * n\_V$ 
(11)         $\text{new\_V}[s] = \text{np.sum}(qs * \text{policy}[i, j])$ 
(12)      End for
(13)    End for
(14)    if  $\text{np.sum}(\text{np.abs}(V - \text{new\_V})) < \text{tol}$ :
(15)      break
(16)     $V = \text{new\_V}$ 
(17)  End for
(18)  return  $V$ 

```

ALGORITHM 2: Value\_evaluate (policy, env, max\_step = 100, tol = 1e-6).

```

(1) Initialize policy
(2)   for all  $s$  in  $S$ :
(3)     Initialize  $qs$ 
(4)     for  $a$  in range (N_ACTIONS):
(5)        $n\_s = \text{env.P}[s, a]$ 
(6)        $r = \text{env.R}[s, a]$ 
(7)        $qs[a] = r + \text{gamma} * V[n\_s[0], n\_s[1]]$ 
(8)      $p = (\text{np.abs}(qs - \text{np.max}(qs)) < 1e-6)$  # greedy strategy
(9)      $p = \text{np.array}(p, \text{dtype} = \text{np.float32}) / \text{np.sum}(p)$  #convert to float type and normalization
(10)     $\text{policy}[i, j] = p$ 
(11)  return policy

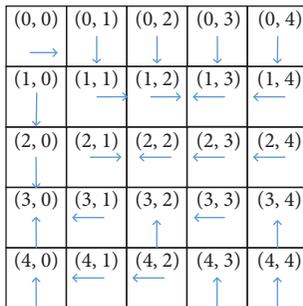
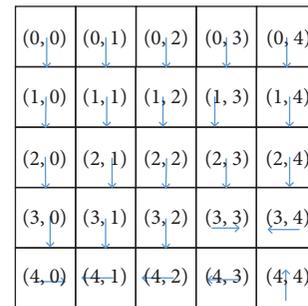
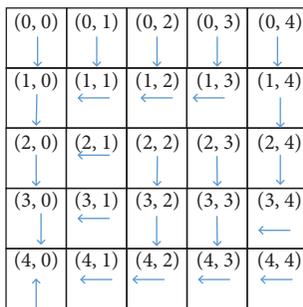
```

ALGORITHM 3: Policy\_improvement (env, V).

```

(1) env = Env ()//Environment Initialization: MDP mode: state  $s$ , action  $a$ , and reward  $r$ 
(2) Initialized  $V$ 
(3) for ( $i = 1, i++, i < \text{max\_iter}$ )
(4)   new_V =  $V$ .copy ()
(5)   update_steps = 0
(6)   for all  $s$  in  $S$ 
(7)     Initialize value function  $V$ 
(8)     for  $a$  in range ( $N\_ACTIONS$ ):
(9)        $n\_s = \text{env.P} [s, a]$ 
(10)       $r = \text{env.R} [s, a]$ 
(11)       $qs [a] = r + \text{gamma} * V [n\_s [0], n\_s [1]]$ 
(12)      update_steps += 1
(13)      new_V [ $i, j$ ] =  $\text{np.max} (qs)$ # update value function base on Bellman's equation
(14)      mean_values.append ( $\text{np.mean} (V)$ )#store the mean value
(15)      run_times.append ( $\text{time.time} () - st$ )#store the run time
(16)      if  $\text{np.sum} (\text{np.abs} (V - \text{new\_V})) < \text{tol}$ :
(17)        break
(18)       $V = \text{new\_V}$ 
(19) return  $V, \text{mean\_values}, \text{run\_times}$ 

```

ALGORITHM 4: SMVI ( $\text{max\_iter} = 100, \text{max\_step} = 100, \text{tol} = 1e-6$ ).FIGURE 5: The optimal policy constructed by SMPI over  $\alpha = 0.1, \beta = 0.1, \text{ and } \chi = 0.8$ .FIGURE 7: The optimal policy constructed by SMPI algorithm over  $\alpha = 0.8, \text{ and } \beta = 0.1, \chi = 0.1$ .FIGURE 6: The optimal policy constructed by SMPI over  $\alpha = 0.1, \beta = 0.8, \text{ and } \chi = 0.1$ .

**5.2. Impact of Weights of the Three Costs on the Expected Reward.** We report also in Figure 8 the impact of the parameter's weights of the three costs on the total expected reward by varying its value in three cases:

Case 1:  $\alpha$  takes a fixed value 0.1,  $\beta$  progressively increases from 0.1 to 0.8, and  $\chi$  progressively decreases from 0.8 to 0.1

Case 2:  $\beta$  takes a fixed value 0.1,  $\alpha$  progressively increases from 0.1 to 0.8, and  $\chi$  progressively decreases from 0.8 to 0.1

Case 3:  $\chi$  takes a fixed value 0.1,  $\alpha$  progressively increases from 0.1 to 0.8, and  $\beta$  progressively decreases from 0.8 to 0.1

It is notable that the maximum reward that each algorithm can achieve may be different when parameters take different values; i.e., the maximum expected reward over  $\alpha = 0.1, \beta = 0.1, \chi = 0.8$  is 9.012, which is the same as parameter settings over  $\alpha = 0.8, \beta = 0.1, \chi = 0.1$ . However, the maximum expected reward over  $\alpha = 0.1, \beta = 0.8, \chi = 0.1$  can only reach 6.8, which is smaller than that of the other two parameter settings.

We observe that when  $\alpha$  takes a fixed value 0.1, the expected system reward progressively decreases. This is because reconfiguration cost and migration cost dominate more system cost. While  $\chi$  takes a fixed value 0.1, the expected system reward progressively increases. This is because increment of weights of resource usage cost can bring more reward under our system context parameter settings. It is

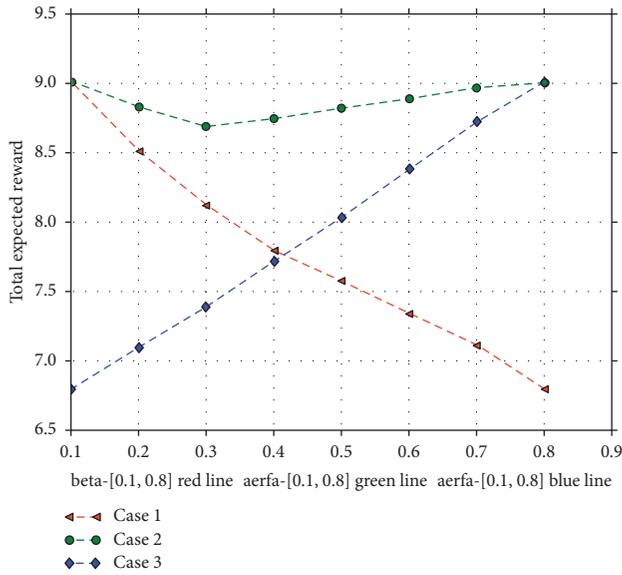


FIGURE 8: The impact of the parameter’s weights of the three costs on the total expected reward.

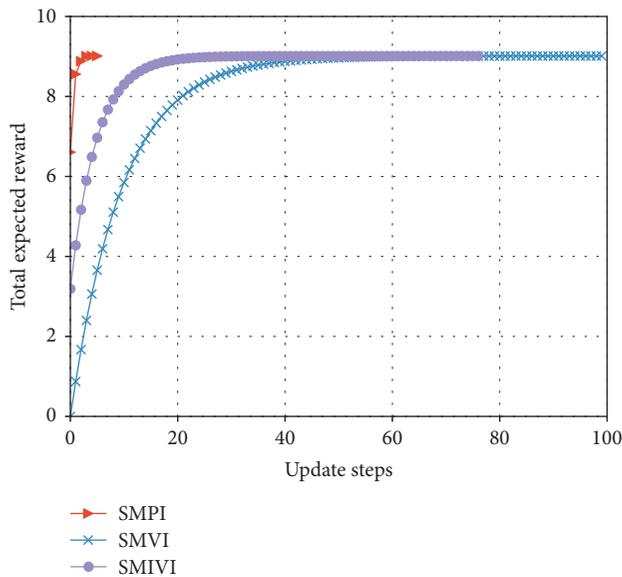


FIGURE 9: The expected reward over  $\alpha = 0.1, \beta = 0.1,$  and  $\chi = 0.8$ .

notable that with the increase of  $\alpha$ , the expected reward of our algorithm declines slightly at the beginning and then increases to a stable level when  $\beta$  takes a fixed value 0.1. And the expected reward over  $\alpha = 0.1, \beta = 0.1, \chi = 0.8$  can reach the maximum value 9.012.

5.3. Comparison of the Convergence Speed of Different Algorithms. In this section, we compare the convergence speed of different algorithms. Here are two settings to be considered.

5.3.1. Total Expected Reward versus the Number of Iteration Update Steps. From Figures 9–11, we observe that our algorithm performs in a similar way under different parameter

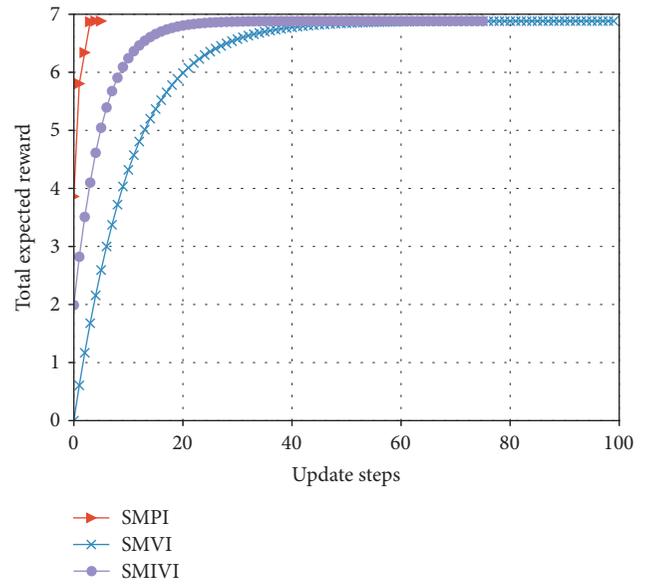


FIGURE 10: The expected reward over  $\alpha = 0.1, \beta = 0.8,$  and  $\chi = 0.1$ .

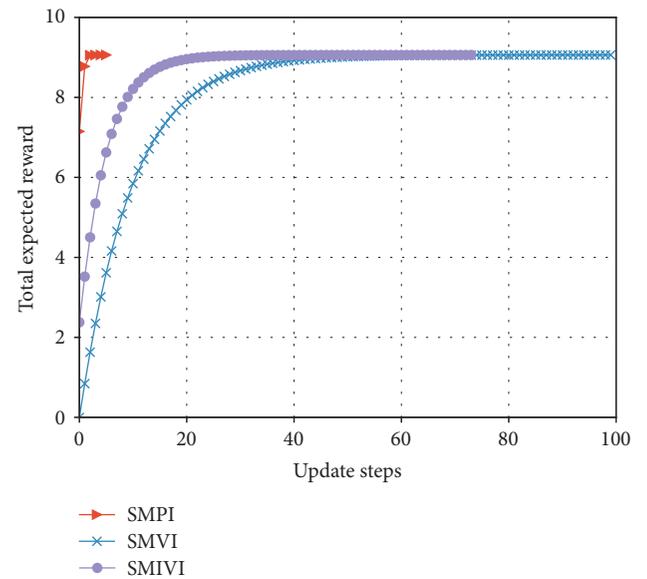


FIGURE 11: The expected reward over  $\alpha = 0.8, \beta = 0.1,$  and  $\chi = 0.1$ .

settings. We can see that when update step takes the same value, the convergence speed of policy iteration algorithm is the fastest and value iteration is the slowest. This is because value iteration finds the optimal policy from the Bellman’s equation iteratively, which may require many iterations before converging to the optimal result. Policy iteration generally requires a smaller number of iterations. Moreover, SMPI finds the exact values of the discounted sum cost for the policy resulting from the previous iteration. SMPI converges at about 7 iterations, while SMVI and SMIVI converge until almost 40 iterations and maintain steady performance thereafter.

5.3.2. Total Expected Reward versus Update Time. As shown in Figures 12 and 13, when update time takes the same value,

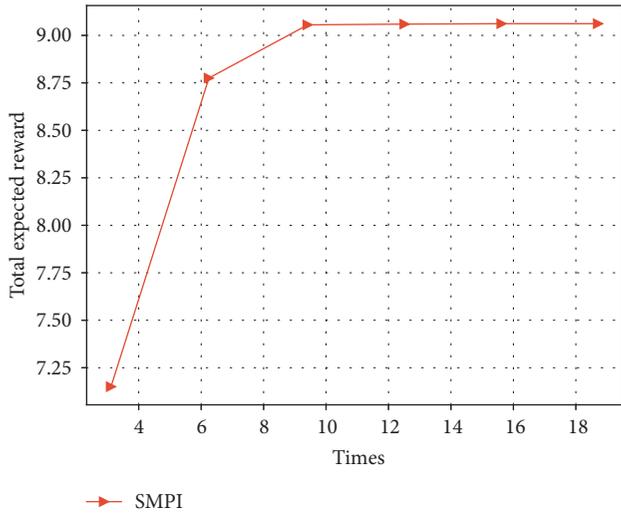


FIGURE 12: The expected reward versus times, implemented by SMPI algorithm with  $\alpha = 0.8$ ,  $\beta = 0.1$ , and  $\chi = 0.1$ .

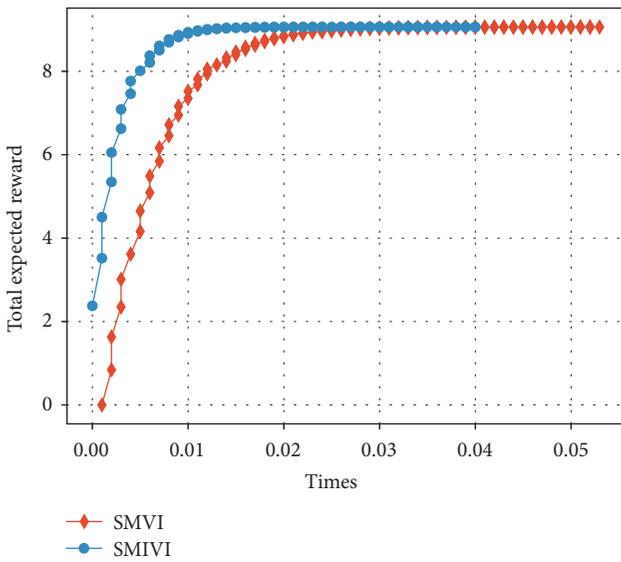


FIGURE 13: The expected reward versus times, implemented by SMVI and SMIVI with  $\alpha = 0.8$ ,  $\beta = 0.1$ , and  $\chi = 0.1$ .

the convergence speed of SMIVI is faster than other algorithms, and SMPI is the slowest (SMVI and SMIVI algorithm converge until almost 0.02 seconds while SMPI algorithm converges at about 5 seconds). This is because SMPI algorithm uses greedy method to improve policy, while the greedy policy improvement generates a lower efficiency.

5.4. Performance Comparison with Other Migration Policies.

In this section, we compare the Proposed Migration Policy considering three costs (PMP) with two benchmark methods in the experiments: the migration strategy only considers Resource Usage cost and Migration Cost strategy (RUMC) [3, 10] and no migration. All the experiments are

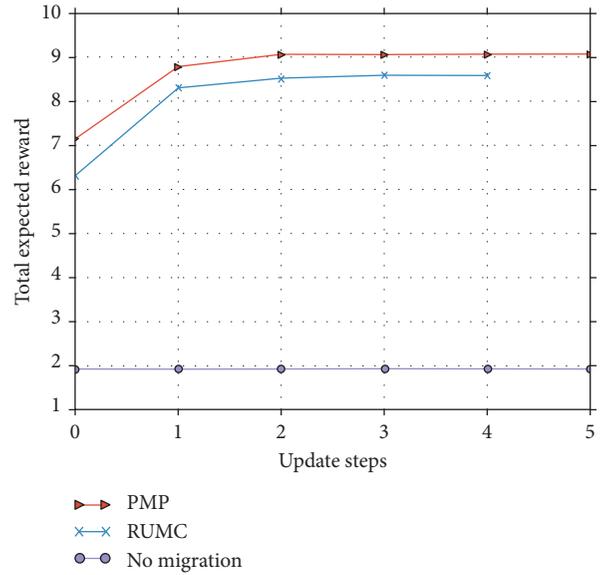


FIGURE 14: The results of implementing SMPI.

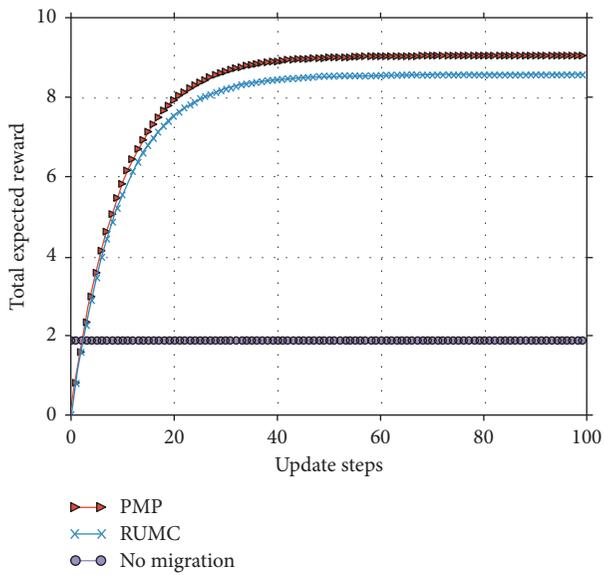


FIGURE 15: The results of implementing SMVI.

implemented under the parameter settings with  $\alpha = 0.8$ ,  $\beta = 0.1$ ,  $\chi = 0.1$ .

It must be noted that the factors considered by different algorithms are not the same, it is difficult to compare directly. For the policy that considers only the communication cost and the migration cost, we first make the corresponding migration decision through the flow of the policy itself and then calculate the reward of this action in the current state according to the function defined in this paper. From the results shown by Figures 14–16, we can see that our PMP implemented by all the three algorithms (SMPI, SMVI, and SMIVI) performs better than RUMC and no migration. That is because RUMC always chooses the destination location that generates the minimum transmission cost and migration cost, while PMP may select one of the four neighbours

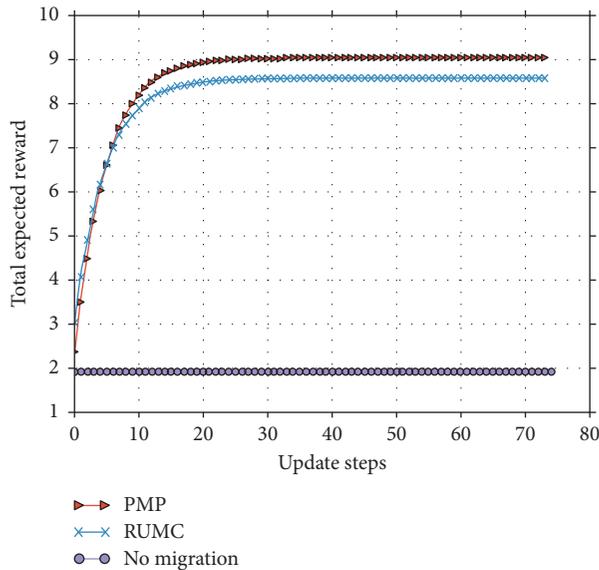


FIGURE 16: The results of implementing SMIVI.

edge servers as destination considering the reconfiguration cost at the destination location. Compared to no migration, PMP can increase reward by three times. Compared to RUMC, PMP can increase reward by 5%.

## 6. Conclusion

In this paper, we design a service migration policy considering mobile terminal's mobility and heterogeneity in edge cloud systems. Our formulation captures general cost models involving transmission cost, migration cost, and reconfiguration cost. We provide a mathematical framework to design optimal service migration policies that decide where to migrate without assumption of mobile terminal's mobility model in a limited area; namely, the policy triggers service migration each time a mobile terminal enters migration zone. We formulate service migration problem into MDP. Value iteration and policy iteration algorithms are used to resolve the optimization problem given by the formulized model. The SMME can optimally decide where a service needs to be migrated. The performance of the main algorithms is demonstrated by extensive experiments. The results show that the proposed service migration mechanism always achieves the maximum expected reward compared to two other policies. The proposed method is suitable for the service migration in limited area. For the large-scale networked MEC systems, the huge state space will lead to the performance degradation of the algorithms. In the future work, we will study live migration method and performance under large-scale networked MEC systems.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC) under grant no. 2002102 and no. 62072121, in part by the Key Technologies R&D Program of Henan Province under grant no. 202102210169, and in part by the Luoyang Major Scientific and Technological Innovation Projects under grant no. 2101017A.

## References

- [1] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: a survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018.
- [2] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: the communication perspective," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.
- [3] P. N. Tran and A. Timm-Giel, "Reconfiguration of virtual network mapping considering service disruption," in *Proceedings of International Conference on Communications 2013*, IEEE, Budapest, Hungary, June 2013.
- [4] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: a survey of the emerging 5 G network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.
- [5] L. Liang, J. Xiao, Z. Ren, Z. Chen, and Y. Jia, "Particle swarm based service migration scheme in the edge computing environment," *IEEE Access*, vol. 8, pp. 45596–45606, 2020.
- [6] S. Wang, J. Xu, N. Zhang, and Y. Liu, "A survey on service migration in mobile edge computing," *IEEE Access*, vol. 6, no. 1, pp. 23511–23528, 2018.
- [7] Z. Liang, Y. Liu, T. M. Lok, and K. Huang, "Multi-cell mobile edge computing: joint service migration and resource allocation," *IEEE Transactions on Wireless Communications (Early Access)*, vol. 1, 2021.
- [8] J. Li, X. Shen, L. Chen et al., "Service migration in Fog computing enabled cellular networks to support real-time vehicular communications," *IEEE Access*, vol. 7, no. 7, pp. 13704–13714, 2019.
- [9] I. Labriji, M. Francesca, C. Davide et al., "Mobility aware and dynamic migration of MEC services for the Internet of vehicles," *IEEE Transactions on Network and Service Management*, vol. 1, no. 18, pp. 570–584, 2021.
- [10] S. Cao, Y. Wang, and C. Xu, "Service migrations in the cloud for mobile accesses: a reinforcement learning approach," in *Proceedings of the 2017 International Conference on Networking, Architecture, and Storage*, pp. 1–10, IEEE, Shenzhen, China, August 2017.
- [11] R. Uргаonkar, S. Wang, T. He, M. Zafer, K. Chan, and K. K. Leung, "Dynamic service migration and workload scheduling in edge-clouds," *Performance Evaluation*, vol. 91, pp. 205–228, 2015.
- [12] S. Wang, R. Uргаonkar, M. Zafer et al., "Dynamic service migration in mobile edge-clouds," in *Proceedings of IFIP Networking Conference 2015*, pp. 1–9, IEEE, Toulouse, France, May 2015.
- [13] D. Wang, Z. Liu, X. Wang, and Y. Lan, "Mobility-aware task offloading and migration schemes in Fog computing networks," *IEEE Access*, vol. 7, no. 7, pp. 43356–43368, 2019.

- [14] K. Ray and A. Banerjee, "Trace-driven modelling and verification of a mobility-aware service allocation and migration policy for mobile edge computing," in *Proceedings of the 2020 IEEE International Conference on Web Services (ICWS)*, pp. 310–317, Honolulu, HI, USA, June 2020.
- [15] U. Mandal, P. Chowdhury, M. Tornatore, C. U. Martel, and B. Mukherjee, "Bandwidth provisioning for virtual machine migration in cloud: strategy and application," *IEEE Transactions on Cloud Computing*, vol. 6, no. 4, pp. 967–976, 2018.
- [16] T. Duong-Ba, T. Tran, T. Nguyen, and B. Bose, "A dynamic virtual machine placement and migration scheme for data centers," *IEEE Transactions on Services Computing*, vol. 14, no. 2, pp. 29–341, 2021.
- [17] S. Kim, "One-on-one contract game-based dynamic virtual machine migration scheme for mobile edge computing," *IEEE Transactions on Emerging Telecommunications Technologies*, vol. 29, no. 1, 2018.
- [18] L. Gao and G. N. Rouskas, "Virtual network reconfiguration with load balancing and migration cost considerations," in *Proceedings of Infocom*, IEEE, Honolulu, HI, USA, April 2018.
- [19] W. Cerroni and F. Esposito, "Optimizing live migration of multiple virtual machines," *IEEE Transactions on Cloud Computing*, vol. 6, no. 4, pp. 1096–1109, 2018.
- [20] K. Ha, "Adaptive VM handoff across cloudlets," CMU-CS-15-113, CMU School of Computer Science, Pittsburgh, PA, USA, 2015, <http://elijah.cs.cmu.edu/Docs/CMU-CS-15-113.pdf>.
- [21] S. Raj, N. Mangal, S. Savitha, and S. Salvi, "Virtual machine migration in heterogeneous clouds-a practical approach," in *Proceedings of the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECT)*, pp. 1–6, Bangalore, India, July 2020.
- [22] A. Machen, S. Wang, K. K. Leung, B. J. Ko, and T. Salonidis, "Live service migration in mobile edge clouds," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 140–147, 2018.
- [23] W. Bao, D. Yuan, Z. Yang et al., "Follow me Fog: toward seamless handover timing schemes in a Fog computing environment," *IEEE Communications Magazine*, vol. 55, no. 11, pp. 72–78, 2017.
- [24] W. Lu, X. Meng, and G. Guo, "Fast service migration method based on virtual machine technology for MEC," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4344–4354, 2019.
- [25] F. Zhang, G. Liu, X. Fu, and R. Yahyapour, "A survey on virtual machine migration: challenges, techniques, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 1206–1243, 2018.
- [26] M. Puterman, *Markov Decision Process: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Hoboken, NJ, USA, 1994.
- [27] A. Strunk, "A lightweight model for estimating energy cost of live migration of virtual machines," in *Proceedings of IEEE Sixth International Conference on Cloud Computing*, pp. 510–517, IEEE, Santa Clara, Santa Clara, CA, USA, July 2013.
- [28] L. Wang, L. Jiao, J. Li et al., "Online resource allocation for arbitrary user mobility in distributed edge clouds," in *Proceedings of IEEE International Conference on Distributed Computing Systems*, pp. 1281–1290, IEEE, Atlanta, GA, USA, June 2017.

## Research Article

# Clinical Effects of Form-Based Management of Forceps Delivery under Intelligent Medical Model

Siming Xin , Zhizhong Wang , Hua Lai , Lingzhi Liu , Ting Shen ,  
Fangping Xu , Xiaoming Zeng , and Jiusheng Zheng 

Department of Obstetrics, Maternal and Child Health Hospital Affiliated to Nanchang University, Nanchang 330006, Jiangxi, China

Correspondence should be addressed to Xiaoming Zeng; 18070038675@163.com and Jiusheng Zheng; zjsheng2012@sina.com

Received 12 April 2021; Revised 13 May 2021; Accepted 20 May 2021; Published 31 May 2021

Academic Editor: Han Wang

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

**Background.** Forceps delivery is one of the most important measures to facilitate vaginal delivery. It can reduce the rate of first cesarean delivery. Frustratingly, adverse maternal and neonatal outcomes associated with forceps delivery have been frequently reported in recent years. There are two major reasons: one is that the abilities of doctors and midwives in forceps delivery vary from hospital to hospital and the other one is lack of regulations in the management of forceps delivery. In order to improve the success rate of forceps delivery and reduce the incidence of maternal and neonatal complications, we applied form-based management to forceps delivery under an intelligent medical model. The aim of this work is to explore the clinical effects of form-based management of forceps delivery. **Methods.** Patients with forceps delivery in Maternal and Child Health Hospital Affiliated to Nanchang University were divided into two groups: form-based patients from January 1, 2019, to December 31, 2020, were selected as the study group, while traditional protocol patients from January 1, 2017, to December 31, 2018, were chosen as the control group. Then, we compared the maternal and neonatal outcomes of these two groups. **Results.** There were significant differences in the maternal and neonatal adverse outcomes such as rate of postpartum hemorrhage, degree of perineal laceration, and incidence of neonatal facial skin abrasions between the two groups, whereas differences in the incidence of asphyxia and intracranial hemorrhage were not significant. **Conclusions.** Form-based management could help us assess the security of forceps delivery comprehensively, as it could not only improve the success rate of the one-time forceps traction scheme but also reduce the incidence of maternal and neonatal adverse outcomes effectively.

## 1. Introduction

During the information era, medical and health field was gradually moving towards informationization and intellectualization. The intelligent medicine was a cross-discipline on the integration of life science and information technology, and it was a new stage of information construction in the healthcare field. By using information management, patients' clinical data could be fully recorded and tell us more about the situation of patients, which could help us to design effective treatments.

Under the environment of high cesarean section rate, we have the responsibility to promote vaginal birth and reduce primary cesarean birth rate [1, 2]. In order to obtain this goal, we need to put intelligent medicine into our daily

medical works for reducing the morbidity of perinatal pregnant women and newborns. Through this way, we could reduce the evaluation errors caused by differences in clinical experience and other factors, so as to reduce the incidence of perinatal pregnant women and newborns [3]. In recent years, intelligent medicine has been widely used in prenatal fetal heart monitoring [4, 5], vaginal trial delivery model prediction after cesarean section [6], prediction model of vaginal birth after cesarean, and premature delivery [7]. It also has been used in postpartum hemorrhage prediction [8, 9]. During the time, it has achieved fairly good results in vaginal midwifery training [10].

Vaginal forceps delivery, one of the surgical vaginal methods, could resolve cephalic dystocia effectively, for example, maternal exhaustion, maternal cardiac disease and

a need to avoid pushing in the second stage of labor, prolonged second stage of labor, and nonreassuring fetal heart rate patterns in the second stage of labor [11]. Under these conditions, forceps delivery could be accomplished more safely and quickly than cesarean. Therefore, the skill of forceps delivery was one of the most important clinical skills required for obstetricians and midwives [12]. Of course, forceps delivery not only requires the above skill but also needs accurate judgment and systematic evaluation of the patient's condition in advance. In recent decades, severe laceration of the birth canal, postpartum hemorrhage, and neonatal injury caused by forceps delivery have been constantly reported [13, 14]. Our hospital began the construction of an intelligent hospital in 2018 and has successively introduced information systems such as integrated platform, HIS, LIS, and HRP. Relying on these information platforms, we were able to implement form-based management of forceps delivery since January 1, 2019. In order to explore the effectiveness of form-based management of forceps delivery in improving the success rate of forceps delivery and reducing the incidence of adverse maternal and fetal outcomes, we were reviewing and analyzing the maternal and infant outcomes of patients with forceps delivery in our hospital during the period from January 1, 2017, to December 31, 2020.

## 2. Materials and Methods

**2.1. Study Population.** This retrospective cohort study was conducted at the Maternal and Child Health Hospital Affiliated to Nanchang University. The study population consisted of pregnant women who underwent forceps delivery from January 1, 2017, to December 31, 2020. Cases with forceps delivery managed by form-based management from January 1, 2019, to December 31, 2020, were established as the study group, and cases with forceps delivery managed by traditional protocol from January 1, 2017, to December 31, 2018, were established as the control group. The inclusion criteria were as follows: single full-term fetus, fetal position being cephalic position, indications for forceps delivery being prolonged second stage of labor or suspicion of immediate fetal distress or maternal complications, and low forceps with Kielland or Simpson forceps. The position of the fetal head in low forceps was that the lowest point of fetal cranial mass was located at or below +3 cm but it did not reach the pelvic floor. The exclusion criteria were huge babies and scarred uterus. The labor process in this study adopted new labor process standards [15]. Eligible women were identified from the hospital data management system, and all study participants were informed of the risks associated with forceps delivery and signed an informed consent form before undergoing forceps delivery.

**2.2. Clinical Data Collection.** Our hospital has introduced a medical record information system and medical record management system. Information between these two systems was completely interoperable. All medical records in this study were retrieved through the medical record management system. The retrieval strategy was: low forceps delivery as the procedure code, procedure time of January 1, 2019, to December 31, 2020, for the study group and January

1, 2017, to December 31, 2018, for the control group, the discharge code was singleton live birth. The delivery data was entered into the hospital data management system immediately after delivery by the midwives responsible for the ongoing care of the woman. Maternal data included age, height, weight, gravidity, parity, gestational age, indication of forceps delivery, instrument used, estimated blood loss, and degree of perineal laceration. Newborn data included birth weight, Apgar scores, umbilical artery blood pH value, intracranial hemorrhage, and facial skin injury.

## 3. Form-Based Forceps Delivery Management Program

From January 1, 2019, to December 31, 2020, we used form-based management of vaginal forceps delivery including a delivery room safe delivery verification form (Table 1) and a forceps delivery verification form (Table 2), which was verified by senior physicians, resident physicians, and midwives to systematically standardize forceps operations with the following steps. In particular, the senior doctors and midwives who were allowed to participate in this study should have the following qualifications: senior doctors should have intermediate or above titles, have at least 2 years of work experience in the delivery room, have passed the hospital assessment, and have been authorized the forceps midwifery technique while midwives were senior aided birth attendants with no less than 5-year midwifery experience.

**3.1. Preoperative Verification.** Prior to forceps implementation, residents reported to the senior physician about the progress of delivery, previous history, the ultrasound results in details, and the current dilemma. Then, the senior physician checked the delivery safety checklist (Table 1) and performed an abdominal examination to reestimate the fetal size and fetal lie, then clarified the fetal station, fetal head descent, and the clinical adequacy of the maternal pelvis by vaginal examination. A discussion was taken among the senior physician, residents, and midwives to determine whether the forceps assisted delivery was reasonable or not, and then an appropriate type of forceps would be selected as a result. The selection criteria of forceps were as following: Simpson forceps were selected when the fetal position was the anterior occipital position or nonoccipital position turned into anterior occipital position by hand, whereas Kielland forceps were selected when the fetal position was transverse and hand rotation fails. Then, the first part of the forceps delivery checklist (Table 2) would be completed.

**3.2. Preoperative Preparation.** After the verification work is finished, the senior physician would tell the pregnant woman and her guardian the necessity and risks of forceps delivery and then indicate them to sign an informed consent form. Meanwhile, the residents and midwives completed the preparation of the midwifery kit, forceps, and neonatal asphyxia resuscitation equipment and completed the second part of the forceps assisted delivery verification form (Table 2).

TABLE 1: The delivery room safe delivery verification form.

Name: _____ Case Number: _____ Age: _____ Gestational age: _____		Delivery time: _____ Number of fetus: _____ Gravidity times: _____ Parity times: _____			
Determining the proximity of labor		Preparing for delivery		2 hours after delivery	
<p>1. Medical history information</p> <p>1.1 History of emergency delivery  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>1.2 History of postpartum bleeding  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>1.3 Uterine scar  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>1.4 Pregnancy complications  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>1.5 Other special circumstances _____</p> <p>1.6 Whether there is special medication?  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>1.7 Any history of drug allergies  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>2. Treatment for pregnant women</p> <p>2.1 Whether glucocorticoids have been used to promote fetal lung maturation?  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>2.2 Whether need antibiotics?  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>2.3 Whether it is necessary to prepare blood in advance?  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>2.4 Whether needing magnesium sulfate and antihypertensive treatment  <input type="checkbox"/> Yes, administered magnesium sulfate  <input type="checkbox"/> Yes, administration of antihypertensive drugs  <input type="checkbox"/> No</p> <p>3. Type of fetal heart monitoring  <input type="checkbox"/> I    <input type="checkbox"/> II    <input type="checkbox"/> III</p> <p>4. Whether informing pregnant women and their families to seek help in case of special signs during delivery?  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p>		<p>1. Abnormal signs maternal or fetal  <input type="checkbox"/> Yes, call for help      <input type="checkbox"/> No</p> <p>2. Whether a pediatrician is needed  <input type="checkbox"/> Yes, already contacted  <input type="checkbox"/> No</p> <p>Make sure you have the necessary supplies at your bedside and prepare for delivery</p> <p>3. For maternity</p> <p>3.1 Estradiol 10U inhalation syringe  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>3.2 Open venous access  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>3.3 Whether other contracting agents are needed at the same time  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>4. For newborns, the following items have been checked for functional status  <input type="checkbox"/> Air-Shields Manual Breathing Unit  <input type="checkbox"/> Negative Pressure Suction</p> <p>4.1 The radiation table is in good function  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>4.2 Neonatal blood gas collection needle  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>4.3 Neonatal pulse oximeter  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>5. Medical staff is in place under the stage  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>6. The delivery is over and the items are counted correctly  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>Number of gauzes before delivery: _____                  Number of gauzes added intraoperatively : _____                  Number of gauzes after delivery: _____</p>		<p>1. Abnormal maternal vitals  <input type="checkbox"/> Yes, ask for help    <input type="checkbox"/> No</p> <p>2. Whether the mother has abnormal vaginal bleeding (the degree of bladder fullness needs to be assessed before the examination)  <input type="checkbox"/> Yes, ask for help    <input type="checkbox"/> No</p> <p>3. Whether the mother needs medication</p> <p>3.1 Needing antibacterial drugs or not  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>3.2 Needing magnesium sulfate and antihypertensive treatment or not  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>4. The need of newborns</p> <p>4.1 Transfer to Pediatrics  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>4.2 Special care and monitoring in the maternity unit  <input type="checkbox"/> Yes, already ready  <input type="checkbox"/> No</p> <p>5. Initiation of breastfeeding and skin-to-skin contact between mother and baby (if the mother and newborn are in good condition)  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p> <p>6. In addition to the midwife's handover, are there any special circumstances that require the doctor's handover?  <input type="checkbox"/> Yes      <input type="checkbox"/> No</p>	
Verification time: _____ Doctor's signature: _____ Midwife's signature: _____		Verification time: _____ Doctor's signature: _____ Midwife's signature: _____		Verification time: _____ Doctor's signature: _____ Midwife's signature: _____	

TABLE 2: The forceps delivery verification form.

Name: _____ Case Number: _____ Age: _____ Gestational age: _____		Delivery time: _____ Number of fetus: _____ Gravidity times: _____ Parity times: _____	
Pre-operative diagnosis:			
Indication of forceps <input type="checkbox"/> Fetal distress <input type="checkbox"/> Prolonged second stage of labor <input type="checkbox"/> Other maternal-fetal complications			
Type of forceps: <input type="checkbox"/> Kielland forceps <input type="checkbox"/> Simpson forceps			
Preoperative verification	Preoperative preparation	Intraoperative verification	Postoperative verification
<p>1.Full opening of uterus <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>2.Fetal position <input type="checkbox"/> Anterior occipital <input type="checkbox"/> Posterior occipital <input type="checkbox"/> Transverse occipital</p> <p>3.Estimated fetal weight</p> <p>4.Touching the fetal head from abdomen <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>5. Caput succedaneum size: __</p> <p>6.Biparietal diameter through the plane of the sciatic spine of the pelvis without contractions <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>7.Skull overlap <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>8.Abnormal intrapelvic measurements <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>9. Ruptured membranes <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>1.Signing informed consent form of forceps delivery <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>2. Call pediatricians <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>3.Prepare for neonatal resuscitation equipment <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>4.Catheterization <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>5.Perineal nerve block <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>6.Episiotomy <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>1.Reconfirm fetal position is occipito-anterior <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>2.Fetal orientation is not occipital anterior position <input type="checkbox"/> Manually rotate the fetal head into anterior occipital <input type="checkbox"/> Manual rotation fails, using Kielland forceps to rotate fetal head into anterior occipital</p> <p>3.Placing forceps on both sides of fetal head and ensuring no soft tissues of the birth canal are clamped <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>4.Reconfirming the correct position of the forceps after clamped forceps <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>5.Pulling along the pelvic axis <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>7.Taking off forceps after crowning of head <input type="checkbox"/> Yes <input type="checkbox"/> No</p>	<p>1.Mother</p> <p>1.1 Injury of birth canal:</p> <p>1.2 Volume of postpartum hemorrhage:</p> <p>1.3 Causes of postpartum bleeding:_____</p> <p>1.4 Other complications:_____</p> <p>2.Newborns</p> <p>2.1 Birth weight:_____</p> <p>2.2 Apgar scores:_____</p> <p>2.3 Umbilical artery pH value: _____</p> <p>2.4 Area of facial indentations or abrasions:</p> <p>2.5 Scalp hematoma:_____</p> <p>2.6 Other complications:_____</p>
Verification time: _____	Verification time: _____	Verification time: _____	Verification time: _____
Doctor's signature: _____	Doctor's signature: _____	Doctor's signature: _____	Doctor's signature: _____
Midwife's signature: _____	Midwife's signature: _____	Midwife's signature: _____	Midwife's signature: _____

**3.3. Intraoperative Verification.** To ensure bladder empty and adequate analgesia, if necessary, a lateral perineal incision would be recommended. The senior physician should reconfirm the fetal position that was occipitoanterior by vaginal examination. If it was posterior or transverse occipital, Sb should transfer them into anterior by manually rotating the fetal head between contractions. Once the rotation failed, Kielland forceps were used to rotate the fetal head into an occipitoanterior position.

**3.3.1. Operations of Kielland Forceps.** Placing forceps on both sides of the fetal head by the one-handed forceps method, then the operator clamped forceps after confirming no soft tissues of the birth canal were clamped. After reconfirming the correct position of the forceps, the operator took a standing position, with the index and middle fingers placed on the two shoulders of the forceps, and then pulled the forceps along the pelvic axis. In the first place, the direction of downward and outward traction was at an angle of 30 degrees below the horizontal plane. As the fetal head was gradually delivered, the handle of the forceps was slowly lifted. When the fetal head was exposed, the direction of traction was changed to horizontal, and the forceps were removed when the fetal head was pulled to the crown. The midwife continued to assist in the delivery of the fetal head and carcass.

**3.3.2. Operations of Simpson Forceps.** Doctor's posture and the traction direction of Simpson forceps were the same as Kielland forceps, while the difference was that in an operation of Simpson forceps when the vaginal opening exposed the forehead of the fetus, the handle of the forceps was gradually lifted up to help the fetal head stretch up. When the mandible of the fetus could be touched, the forceps would be removed, and the midwife continued to deliver the fetal head and carcass.

**3.4. Postoperative Verification.** Once the baby was born, a neonatologist would conduct a careful physical examination. The examination included heart rate, respiration, muscle tension, body reflexes, skin color, facial indentation or abrasion, scalp hematoma, clavicle fracture, and organ dysplasia and then performed Apgar scores. At the same time, the umbilical artery was taken for blood gas analysis to comprehensively evaluate the condition of the newborn. Assessment criteria for neonatal asphyxia: (1) mild asphyxia: Apgar score 1 min  $\leq 7$ , or 5 min  $\leq 7$ , with umbilical artery blood pH  $< 7.2$ , and (2) severe asphyxia: Apgar score 1 min  $\leq 3$  or 5 min  $\leq 5$ , with umbilical artery blood pH  $< 7.0$  [16]. The birth canal was examined by obstetricians and midwives; if there was a laceration of the birth canal, it needed to be sutured. In addition, as forceps midwifery was one of the high-risk factors of postpartum hemorrhage, we gave parturient prophylactic drug treatment to promote uterine contraction immediately after shoulder delivery. If postpartum hemorrhage occurred, the cause of the

hemorrhage should be identified quickly, then therapeutic drugs, surgical hemostasis, and even blood products infusion had to be carried out.

## 4. Traditional Management Scheme of Forceps Delivery

Before implementation of forceps, the delivery process and related auxiliary examination should be checked by the senior physician. After learning fetal head station, fetal position, auricle direction by vaginal examination, the senior physician decided whether it is necessary to carry out forceps delivery or not. Once forceps delivery was decided, the senior physician implemented or instructed residents to perform forceps delivery. The operations of forceps were the same as the study group.

## 5. Observed Indicators

We evaluated the clinical effects of form-based forceps management by the following indicators: success rate of primary forceps traction, rate of perineal laceration, postpartum hemorrhage, neonatal asphyxia, intracranial hemorrhage, and facial skin injury.

## 6. Statistical Analysis

The sociodemographic characteristics and delivery-related data of the subjects were collected from the electronic case system. We analyzed the skewness and kurtosis of the patients' clinical data. Normally distributed data such as age, gestational age, and body mass index (BMI) were expressed as mean  $\pm$  standard deviation, nonnormally distributed data such as gravidity and parity times were expressed as median and interquartile spacing, and qualitative information was expressed as composition ratio. We used an independent sample *t*-test to analyze the potential statistical differences of age, gestational age, and BMI; used Mann-Whitney *U* test to analyze the statistical differences of gravidity, parity times, perineal laceration, and neonatal asphyxia; and used the chi-square test to compare the composition ratio between the two groups. The *P* value was two-sided and the result was considered significantly different at *P*  $< 0.05$ . All the above-mentioned analyses were carried out with SPSS software.

## 7. Results

**7.1. Clinical Characteristics.** During the period from January 1, 2017, to December 31, 2018, the number of forceps deliveries in our hospital was 626, while the number of cumulative deliveries and vaginal deliveries was 44,601 and 24,593, respectively. From January 1, 2019, to December 31, 2020, there were 634 forceps deliveries, 42,409 total deliveries and 23,398 vaginal deliveries. The details are shown in Table 3. Summary statistics of clinical data of the two study groups are shown in Table 4, indicating little statistical difference between the two groups.

TABLE 3: The status of vaginal deliveries and forceps deliveries during the last 4 years.

	01/01/2017–31/12/2018	01/01/2019–31/12/2020	<i>P</i> value	Method
Total deliveries (case)	44601	42409		
Vaginal deliveries (case/rate)	24593 (55.14%)	23398 (55.17%)	0.924	Pearson
Forceps deliveries (case/rate)	626 (2.55%)	634 (2.71%)	0.261	Pearson

TABLE 4: Descriptive statistics of basic information of study population.

Features	Study group ( <i>n</i> = 634)	Control group ( <i>n</i> = 626)	<i>P</i> value	Method
Age (years)	29.33 ± 5.22	28.40 ± 5.72	0.697	Independent sample <i>t</i>
Gestational age (days)	273.29 ± 16.01	270.00 ± 15.03	0.149	Independent sample <i>t</i>
Gravidity (times)	2.00 (1.00–3.00)	3.00 (1.50–5.00)	0.28	Mann–Whitney
Parity (times)	0.50 (0.00–1.00)	0.50 (0.00–1.50)	0.71	Mann–Whitney
BMI (kg/m <sup>2</sup> )	25.79 ± 3.36	25.74 ± 2.03	0.973	Independent sample <i>t</i>

Data are presented as mean ± standard deviation or median (interquartile spacing). BMI: body mass index.

**7.2. Maternal Outcomes.** The success rate of forceps delivery was 100% in both groups. All of the pregnant women in the study group had successful one-time traction, while three patients in the control group had failed in their first traction. The main reason for failure was inaccurate forceps placement due to fetal position error, and we had second-time successful traction after repositioning forceps. Comparing the success rate of disposable forceps traction between the two groups, it was found that the success rate of the study group was slightly higher than that of the control group, although the differences were not obvious. The rates of postpartum hemorrhage, second-degree perineal laceration, and third-degree perineal laceration in the study group were 16.09%, 11.99%, and 0.32%, while those in the control group were 24.92%, 18.85%, and 0.48%. It was found that the rate of postpartum hemorrhage and the degree of perineal laceration in the study group were significantly lower than those in the control group (Table 5).

**7.3. Neonatal Outcomes.** There were 634 newborns in the study group, of which 7 had mild asphyxia and no severe asphyxia. There were 626 newborns in the control group, including 13 cases of mild asphyxia and 1 case of severe asphyxia due to intracranial hemorrhage. There was no significant difference in neonatal asphyxia between the two groups. In the control group, one newborn had intracranial hemorrhage. It was a case of a second traction after repositioning the forceps due to incorrect judgment of fetal position, and the possible cause of intracranial hemorrhage was considered to be excessive compression of the fetal head. In addition, the incidence of facial skin injury was 3.94% in the study group, which was significantly lower than that in the control group (Table 6).

## 8. Discussion

Since 1996, with the improvement of cesarean delivery techniques and the enhancement of pregnant women's awareness of safe delivery, the rate of cesarean section has increased in both developed and developing countries, far exceeding the alert level of the cesarean delivery rate set by

the World Health Organization [17, 18]. However, the cesarean section was not as safe as that we thought. In 2007, a study from Canada showed that the risk of serious maternal illness was three times higher in patients who had a cesarean section than in those who had a vaginal delivery [19]. In the last 2 years, a number of studies had shown that unnecessary cesarean sections increased maternal and neonatal risks, even if it might increase maternal mortality [15, 20]. In short, we believed that vaginal delivery was the safer and more cost-effective way of delivery.

During the second stage of labor, when vaginal delivery became difficult, forceps delivery was an important measure to solve cephalic dystocia. Studies had shown that low forceps or export forceps performed by experienced and trained doctors in the second stage of labor might safely reduce risks of cesarean delivery and that vaginal surgical delivery should be considered a safe and acceptable alternative to cesarean delivery [12, 15, 21, 22]. However, forceps delivery is highly required for obstetricians and midwifery; if performed improperly, it might cause serious birth canal injuries, postpartum hemorrhage, neonatal birth injuries, and other complications. All of the complications would do great harm to the mother and the infant. Therefore, we needed to try to avoid complications.

In 2016, our hospital improved the traction method of Kielland forceps which was consistent with the previously described method. We found that the improved method could reduce the complications of forceps delivery to some extent, but because these technical improvements relied more on the experience of operating physician, and a study showed that the probability of error in judgment of fetal position and parameters was 50%–80% for residents and 36%–80% for attending physicians [23]; therefore, the results we achieved were not significant. Subsequently, we carefully analyzed cases of forceps delivery that had serious complications before January 1, 2019, and we found the main causes of these complications were incomplete preoperative assessment and improper operation, while the underlying cause was the lack of standardized management of forceps delivery.

The workload of the medical staff in the delivery room was heavy. In order to remind the medical staff to pay

TABLE 5: Maternal outcomes of study population.

	Study group ( <i>n</i> = 634)	Control group ( <i>n</i> = 626)	<i>P</i> value	Method
Successful one-time traction (case/rate)	634 (100%)	623 (99.52%)	0.122	Fisher's exact
Postpartum hemorrhage (case/rate)	102 (16.09%)	156 (24.92%)	<0.001	Pearson
Perineal laceration				
Second degree (case/rate)	76 (11.99%)	118 (18.85%)	0.001	Mann-Whitney
Third degree (case/rate)	2 (0.32%)	3 (0.48%)		

TABLE 6: Newborn outcomes of study population.

	Study group ( <i>n</i> = 634)	Control group ( <i>n</i> = 626)	<i>P</i> value	Method
Neonatal asphyxia				
Mild asphyxia (case/rate)	7 (1.10%)	13 (2.08%)	0.116	Mann-Whitney
Severe asphyxia (case/rate)	0 (0)	1 (0.16%)		
Intracranial hemorrhage (case/rate)	0 (0)	1 (0.16%)	0.497	Fisher's exact
Facial skin injury (case/rate)	25 (3.94%)	82 (13.10%)	<0.001	Pearson

attention to the surgical risk of each forceps delivery, we have changed our traditional forceps delivery protocol and have been using a form-based forceps delivery management protocol since 2019. We established a delivery safety checklist and a forceps delivery checklist. The forceps delivery checklist was described in detail according to four parts: preoperative verification, preoperative preparation, intraoperative operation, and postoperative examination. Specific verification requirements were put forward from aspects of prerequisites of forceps midwifery, communication, personnel and facility preparation, key points of operation, examination of postoperative maternal and fetal complications, and so forth. At the same time, senior physicians, residents, and midwives were required to participate in and independently verify the key steps in the process. In this way, we could avoid not only the omission of preparations by medical staff due to fatigue, negligence, and emergency but also incorrect operations due to inexperience and so forth. In addition, we also emphasized the postoperative verification of maternal and infant complications, identified the causes of complications timely, and correct errors early. Through this way, we could improve the forceps delivery continuously.

After carefully checked by senior doctors, residents, and midwives before the forceps delivery, the fetal positions of all fetuses among 634 patients in the study group were accurately judged and the traction was successful at one time. While verification of fetus in the control group only relied on senior doctors, there were three failed tractions at one time due to errors in judging the fetal position, caused by negligence, fatigue, or tension of senior doctors. Also, the incidence of perineal lacerations, postpartum hemorrhage, and neonatal facial skin damage were significantly lower than those of the control group based on no difference in neonatal asphyxia. All of these fully demonstrated that the form-based management of forceps delivery could strengthen cooperation between doctors and midwives, help medical staff comprehensively, and accurately evaluate the necessity and operating conditions of the forceps delivery, aimed to avoid errors in forceps operation and reduce complications of

mother and neonates. At the same time, it might not delay the delivery of high-risk newborns. Therefore, form-based management of forceps delivery was beneficial to obstetric forceps management.

## 9. Conclusions

In this paper, we found that the form-based forceps delivery management could improve the success rate of the one-time forceps traction scheme and reduce the maternal rate of postpartum hemorrhage and risk of perineal laceration under the intelligent medical model. In the context of promoting vaginal delivery and reducing the rate of first cesarean delivery, we need to improve the skill level of obstetric medical staff in forceps delivery and strengthen the management of forceps delivery in the department and homogenize the forceps delivery.

## Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Ethical Approval

The present study followed the tenets of the Helsinki Declaration. Ethics approval was obtained from the Institutional Review Board of Maternal and Child Health Hospital Affiliated to Nanchang University in China.

## Consent

Each participating woman gave informed consent.

## Conflicts of Interest

The authors have declared that no potential conflicts of interest exist.

## Authors' Contributions

Siming Xin and Zhizhong Wang contributed equally to this work.

## Acknowledgments

This work was financially supported by the Science and Technology Project of Jiangxi Province (nos. 20192BBGL70003 and 20203BBGL73130) and the Science and Technology Project of Jiangxi Provincial Health Commission (No. 20195475). All authors are grateful to the patients and the IT department who participated in this study.

## References

- [1] A. D. Bell and S. Joy, S. Gullo, R. Higgins, and E. Stevenson, "Implementing a systematic approach to reduce cesarean birth rates in nulliparous women," *Obstetrics and Gynecology*, vol. 130, no. 5, pp. 1082–1089, 2017.
- [2] D. C. Lagrew, L. K. Low, R. Brennan et al., "National partnership for maternal safety: consensus bundle on safe reduction of primary cesarean births-supporting intended vaginal births," *Obstetrics & Gynecology*, vol. 131, no. 3, pp. 503–513, 2018.
- [3] G. S. Desai, "Artificial intelligence: the future of obstetrics and gynecology," *The Journal of Obstetrics and Gynecology of India*, vol. 68, no. 4, pp. 326–327, 2018.
- [4] J. Balayla and G. Shrem, "Use of artificial intelligence (ai) in the interpretation of intrapartum fetal heart rate (fhr) tracings: a systematic review and meta-analysis," *Archives of Gynecology and Obstetrics*, vol. 300, no. 1, pp. 7–14, 2019.
- [5] P. Iftikhar, M. V. Kuijpers, A. Khayat, A. Iftikhar, and M. D. D. Sa, "Artificial intelligence: a new paradigm in obstetrics and gynecology research and clinical practice," *Cur-eus*, vol. 12, no. 2, Article ID e7124, 2020.
- [6] M. Lipschuetz, J. Guedalia, A. Rottenstreich et al., "Prediction of vaginal birth after cesarean deliveries using machine learning," *American Journal of Obstetrics and Gynecology*, vol. 222, no. 6, pp. 613.e1–613.e12, 2020.
- [7] R. O. Bahado Singh, J. Sonek, D. McKenna et al., "Artificial intelligence and amniotic fluid multiomics: prediction of perinatal outcome in asymptomatic women with short cervix," *Ultrasound in Obstetrics & Gynecology*, vol. 54, pp. 110–118, 2019.
- [8] Y. Miyagi, K. Tada, I. Yasuhi et al., "New method for determining fibrinogen and FDP threshold criteria by artificial intelligence in cases of massive hemorrhage during delivery," *Journal of Obstetrics and Gynaecology Research*, vol. 46, no. 2, pp. 256–265, 2019.
- [9] K. K. Venkatesh, R. A. Strauss, C. A. Grotegut et al., "Machine learning and statistical models to predict postpartum hemorrhage," *Obstetrics & Gynecology*, vol. 135, no. 4, pp. 935–944, 2020.
- [10] R. Moreau, M. T. Pham, X. Brun, T. Redarce, and O. Dupuis, "Simulation of an instrumental childbirth for the training of the forceps extraction: control algorithm and evaluation," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 3, pp. 364–372, 2011.
- [11] M. Black and D. J. Murphy, "Forceps delivery for non-rotational and rotational operative vaginal delivery," *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 56, pp. 55–68, 2019.
- [12] N. Tempest, A. Hart, S. Walkinshaw, and D. K. Hapangama, "A re-evaluation of the role of rotational forceps: retrospective comparison of maternal and perinatal outcomes following different methods of birth for malposition in the second stage of labour," *International Journal of Obstetrics & Gynaecology*, vol. 120, no. 10, pp. 1277–1284, 2013.
- [13] S. Biru, D. Addisu, S. Kassa, and S. Animen, "Maternal complication related to instrumental delivery at felege hiwot specialized hospital, northwest Ethiopia: a retrospective cross-sectional study," *BMC Research Notes*, vol. 12, no. 1, 2019.
- [14] G. M. Muraca, Y. Sabr, S. Lisonkova et al., "Morbidity and mortality associated with forceps and vacuum delivery at outlet, low, and midpelvic station," *Journal of Obstetrics and Gynaecology Canada*, vol. 41, no. 3, pp. 327–337, 2019.
- [15] A. B. Caughey, A. G. Cahill, J.-M. Guise, and D. J. Rouse, "Safe prevention of the primary cesarean delivery," *American Journal of Obstetrics and gynecology*, vol. 210, no. 3, pp. 179–193, 2014.
- [16] Neonatal Resuscitation and Chinese Medical Association, "Expert consensus for diagnosis of neonatal asphyxia," *Chinese Journal of Perinatal Medicine*, vol. 19, no. 1, pp. 3–6, 2016.
- [17] World Health Organisation, *Appropriate Technology for Birth*, World Health Organisation, Geneva, Switzerland, 1985.
- [18] A. P. Betran, M. R. Torloni, J. Zhang et al., "What is the optimal rate of caesarean section at population level? a systematic review of ecologic studies," *Reproductive Health*, vol. 12, no. 1, 2015.
- [19] S. Liu, R. M. Liston, K. S. Joseph, M. Heaman, R. Sauve, and M. S. Kramer, "Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term," *Canadian Medical Association Journal*, vol. 176, no. 4, pp. 455–460, 2007.
- [20] C. Kaboré, V. Ridde, N. Chaillet, F. Yaya Bocoum, A. P. Betrán, and A. Dumont, "DECIDE: a cluster-randomized controlled trial to reduce unnecessary caesarean deliveries in burkina faso," *BMC Medicine*, vol. 17, no. 1, p. 87, 2019.
- [21] B. L. Shaffer and A. B. Caughey, "Forceps delivery: potential benefits and a call for continued training," *Journal of Perinatology*, vol. 27, no. 6, pp. 327–328, 2007.
- [22] L. Hinkson, W. Henrich, and B. Tutschek, "Intrapartum ultrasound during rotational forceps delivery: a novel tool for safety, quality control, and teaching," *American Journal of Obstetrics and Gynecology*, vol. 224, no. 1, pp. 91–93, 2021.
- [23] O. Dupuis, R. Silveira, A. Zentner et al., "Birth simulator: reliability of transvaginal assessment of fetal head station as defined by the American college of obstetricians and gynecologists classification," *American Journal of Obstetrics and Gynecology*, vol. 192, no. 3, pp. 868–874, 2005.

## Research Article

# ASC Performance Prediction for Medical IoT Communication Networks

Fagen Yin <sup>1</sup>, Pingping Xiao <sup>1</sup> and Zefeng Li <sup>2</sup>

<sup>1</sup>College of Physical Science and Engineering, Yichun University, Yichun 336000, Jiangxi, China

<sup>2</sup>Institute of Data Science, City University of Macau, Macau 999078, Macau

Correspondence should be addressed to Fagen Yin; ynx522@163.com

Received 13 April 2021; Accepted 21 May 2021; Published 28 May 2021

Academic Editor: Xingwang Li

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

Wearable devices are gradually entering the medical health field. Medical Internet of Things (IoT) has been widely used in all walks of medical health. With the complexity of medical health application scenarios, the medical IoT communication networks face complex environments. The secure communication issue is very important for medical IoT communication networks. This paper investigates the secrecy performance of medical IoT communication networks. To improve the secrecy performance, we adopt a cooperative communication strategy. We also use the average secrecy capacity (ASC) as a metric, and the expressions are first derived. Then, a secrecy performance intelligent prediction algorithm is proposed. The extensive simulations are used to verify the proposed method. Compared with other methods, the proposed algorithm realizes a better prediction precision.

## 1. Introduction

With explosive growth of medical health applications, the fifth-generation (5G) mobile communication has been widely used in medical Internet of Things (IoT) networks [1, 2]. Different 5G applications [3–5] widely appear in medical IoT communication networks, which can provide quick and convenient user experience and services [6]. However, due to the medical user mobility, the secure communication issue of medical IoT networks is facing many challenges [7].

For medical IoT communication networks, physical layer security is becoming more and more important [8]. With an eavesdropper, the authors [9] investigated the impact of antenna correlation. In [10], the authors developed a code scrambling scheme and analyzed secrecy performance. Considering the physical layer security, Yan et al. [11] studied the resource allocation problem for the cognitive relay networks. In [12], the authors proposed an optimal power allocation to achieve the secure transmission. Considering the cooperative jamming, Lu et al. [13] proposed a secure transmission scheme.

However, analyzing and predicting mobile secrecy performance are very difficult. Recently, machine learning techniques are applied in 5G wireless communications [14, 15]. In pattern classification, classifying the binary data was realized by the support vector machine (SVM) model in [16]. The extreme learning machine (ELM) model was proposed to detect anomaly states in [17]. In [18], the general regression (GR) model predicted the video transmission quality.

To date, no existing studies have considered the secrecy performance prediction of AF relaying medical IoT communication networks. As a consequence, we summarize the main contributions as follows:

- (1) The secrecy performance is analyzed with AF relaying scheme. Then, we use the average secrecy capacity (ASC) to evaluate secrecy performance and derive the exact expressions.
- (2) To realize real-time analysis of ASC, we propose an ASC prediction algorithm based on the BP network. ELM, SVM, and GR methods are examined and compared with the proposed method.

- (3) We verify the derived ASC results under different conditions. Compared with different methods, the proposed algorithm realizes a better prediction precision and a lower time complexity.

## 2. The Medical IoT Communication Network Model

In Figure 1, the medical IoT communication networks have a mobile source (MS), mobile eavesdropper (ME), mobile destination (MD), and mobile relay (MR).  $W_{SR}$ ,  $W_{RD}$ , and  $W_{RE}$  are the relative geometrical gains of MS  $\rightarrow$  MR, MR  $\rightarrow$  MD, and MR  $\rightarrow$  ME links, respectively.

Transmission power is  $E$ , which is allocated by  $K$ . 2-Rayleigh distribution can express the channel coefficient  $h$  [19]. Firstly, MR receives the signal  $r_{SR}$  as [20]

$$r_{SR} = \sqrt{W_{SR}KE}h_{SR}x + n_{SR}, \quad (1)$$

where  $n_{SR}$  is Gaussian noise.

In the second time slot, AF is used at MR. MD and ME receive the signals  $r_{Rki}$ ,  $k \in \{D, E\}$ , as

$$r_{Rk} = \sqrt{c_k E}h_{SR}h_{Rk}x + n_{Rk}. \quad (2)$$

The received  $\text{SNR}_{\gamma_{SRki}}$  is given as

$$\gamma_{SRk} = \frac{\gamma_{SR}\gamma_{Rk}}{1 + \gamma_{SR} + \gamma_{Rk}}, \quad (3)$$

where

$$\gamma_{SR} = W_{SR}K|h_{SR}|^2\bar{\gamma}, \quad (4)$$

$$\gamma_{Rk} = (1 - K)W_{Rk}|h_{Rk}|^2\bar{\gamma}, \quad (5)$$

$$\bar{\gamma}_{SR} = W_{SR}K\bar{\gamma}. \quad (6)$$

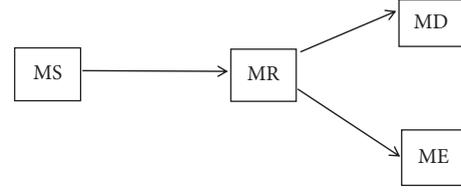


FIGURE 1: System model.

$\gamma_{SRk}$  is hard to calculate exactly. We approximate  $\gamma_{SRki}$  as [21]

$$\gamma_{SRAk} = \frac{\gamma_{SR}\gamma_{Rk}}{1 + \gamma_{SR} + \gamma_{Rk}}, \quad (7)$$

$$\bar{\gamma}_{Rk} = (1 - K)W_{Rk}\bar{\gamma}. \quad (8)$$

With the help of [22], the PDF and CDF of  $\gamma_{SRAk}$  are as follows:

$$f_{\gamma_{SRAk}}(r) = \frac{1}{r}G_{0,4}^{4,0}\left[\frac{r}{\chi_k} \mid 1, 1, 1, 1_-\right], \quad (9)$$

$$F_{\gamma_{SRAk}}(r) = G_{1,5}^{4,1}\left[\frac{r}{\chi_k} \mid 1, 1, 1, 1, 0_1\right], \quad (10)$$

where

$$\chi_k = \frac{\bar{\gamma}_{SR}\bar{\gamma}_{Rk}}{1 + \bar{\gamma}_{SR} + \bar{\gamma}_{Rk}} \quad (11)$$

The instantaneous secrecy capacity is given as [23]

$$C = \max\{\ln(1 + \gamma_{SRA D}) - \ln(1 + \gamma_{SRA E}), 0\}. \quad (12)$$

## 3. Average Secrecy Capacity

The ASC is derived as

$$\bar{C} = \int_0^\infty \int_0^\infty C(\gamma_{SRA D}, \gamma_{SRA E})f(\gamma_{SRA D}, \gamma_{SRA E})d\gamma_{SRA D}d\gamma_{SRA E} = A_1 + A_2 - A_3. \quad (13)$$

$A_1$  is given as

$$\begin{aligned} A_1 &= \int_0^\infty \frac{1}{\gamma_{SRA D}} G_{2,2}^{1,2}(\gamma_{SRA D} \mid 1, 0_{1,1}) G_{0,4}^{4,0}\left[\frac{\gamma_{SRA D}}{\chi_D} \mid 1, 1, 1, 1_-\right] G_{1,5}^{4,1}\left[\frac{\gamma_{SRA D}}{\chi_E} \mid 1, 1, 1, 1, 0_1\right] d\gamma_{SRA D} \\ &= G_{2,2: 0,4: 1,5}^{2,1: 4,0: 4,1} \left[ \begin{array}{c|c|c} 0, 1 & - & 1 \\ \hline 0, 0 & 1, 1, 1, 1 & 1, 1, 1, 1, 0 \end{array} \middle| \frac{1}{\chi_D}, \frac{1}{\chi_E} \right]. \end{aligned} \quad (14)$$

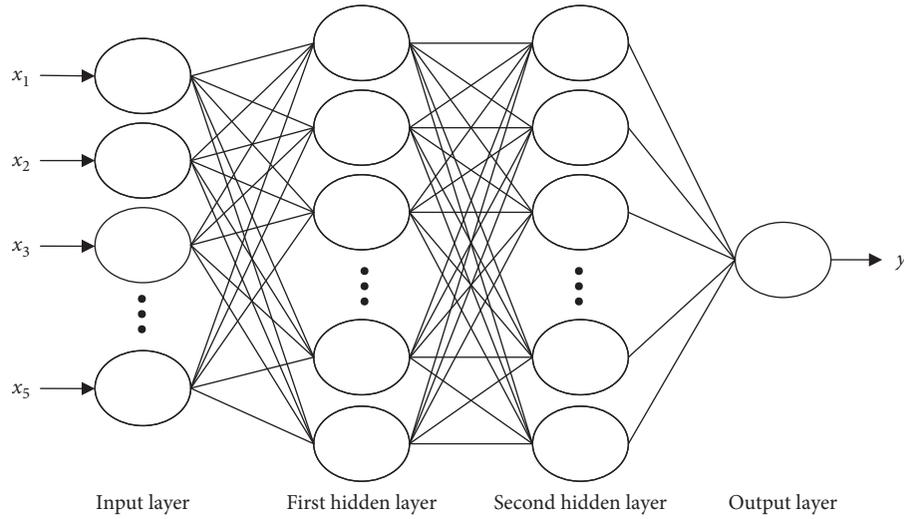


FIGURE 2: The BP structure.

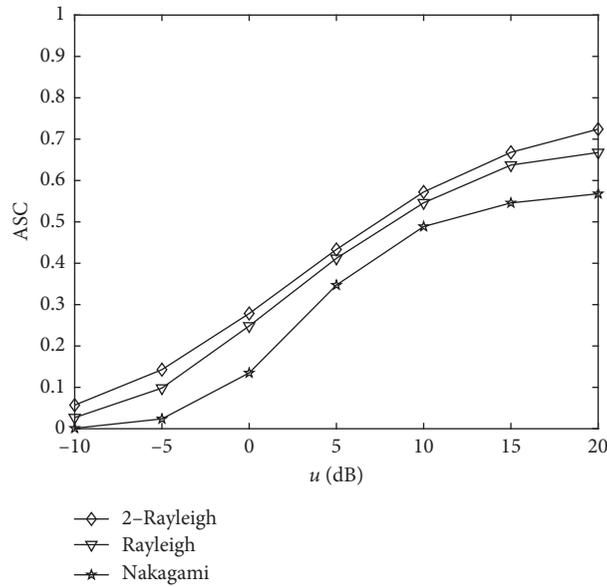


FIGURE 3: The ASC performance versus ( $u$ ).

We obtain the  $A_2$  as

$$\begin{aligned}
 A_2 &= \int_0^\infty G_{2,2}^{1,2}(\gamma_{SRAE}|1, 0_{1,1}) \frac{1}{\gamma_{SRAE}} G_{0,4}^{4,0} \left[ \frac{\gamma_{SRAE}}{\chi_E} | 1, 1, 1, 1_- \right] G_{1,5}^{4,1} \left[ \frac{\gamma_{SRAE}}{\chi_D} | 1, 1, 1, 1, 0_1 \right] d\gamma_{SRAE} \\
 &= G_{2,2: 0,4: 1,5}^{2,1: 4,0: 4,1} \left[ \begin{array}{c|c|c|c} 0, 1 & - & 1 & \\ \hline & & & \frac{1}{\chi_E}, \frac{1}{\chi_D} \\ \hline 0, 0 & 1, 1, 1, 1 & 1, 1, 1, 1, 0 & \end{array} \right].
 \end{aligned} \tag{15}$$

We obtain the  $A_3$  as

$$A_3 = \int_0^\infty \frac{1}{\gamma_{SRAE}} G_{2,2}^{1,2}(\gamma_{SRAE}|1, 0_{1,1}) G_{0,4}^{4,0} \left[ \frac{\gamma_{SRAE}}{\chi_E} | 1, 1, 1, 1_- \right] d\gamma_{SRAE} = G_{2,6}^{6,1} \left[ \frac{1}{\chi_E} | 1, 1, 1, 1, 0, 0_{0,1} \right]. \tag{16}$$

TABLE 1: Simulation parameters.

Parameter	Value
$m$	1, 2, 3
$K$	0.4
$W_{SR}$	5 dB
$W_{RE}$	5 dB

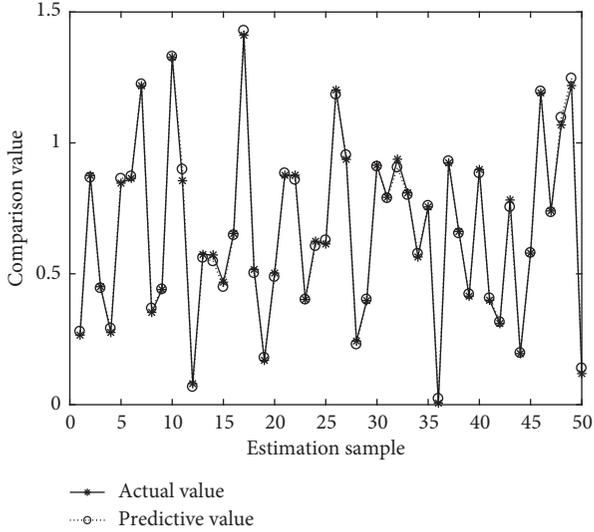


FIGURE 4: Prediction of BP.

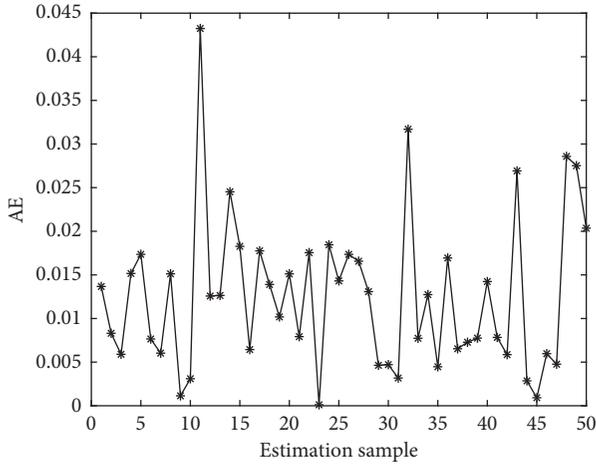


FIGURE 5: AE of BP.

Next, we use the derived ASC expressions to set up the data sets and design the BP prediction model.

#### 4. Secrecy Performance Prediction Method

4.1. Data Sets.  $T_i = (X_i, y_i)$ .  $X_i$  is given as

$$X_i = (x_{i1}, x_{i2}, \dots, x_{i5}). \quad (17)$$

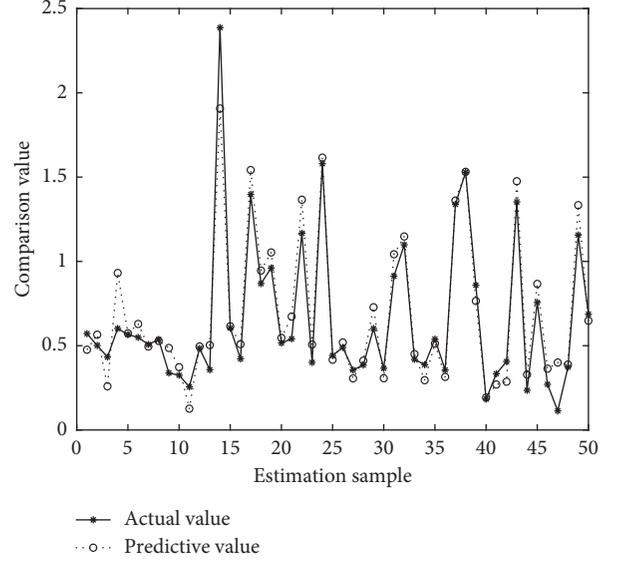


FIGURE 6: Prediction of ELM.

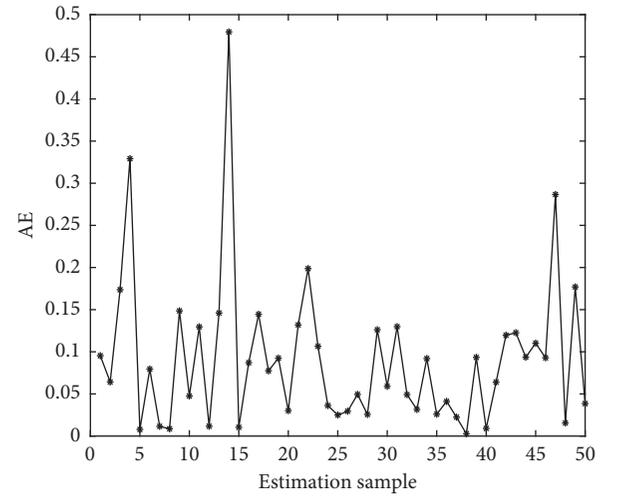


FIGURE 7: AE of ELM.

$X_i$  includes 5 indicators, which are  $W_{SR}$ ,  $W_{RD}$ ,  $W_{RE}$ ,  $K$ , and  $\bar{\gamma}$ . The ASC performance is the output  $y_i$ . By using (13), it can obtain the corresponding  $y_i$ .

4.2. Network Structure. Figure 2 shows the BP structure [24].

4.3. Metrics. Two metrics are MSE and AE. For  $PP$  testing data, they are given as

$$MSE = \frac{\sum_{z=1}^{PP} (d^z - y^z)^2}{PP}, \quad (18)$$

$$AE = |d^z - y^z|. \quad (19)$$

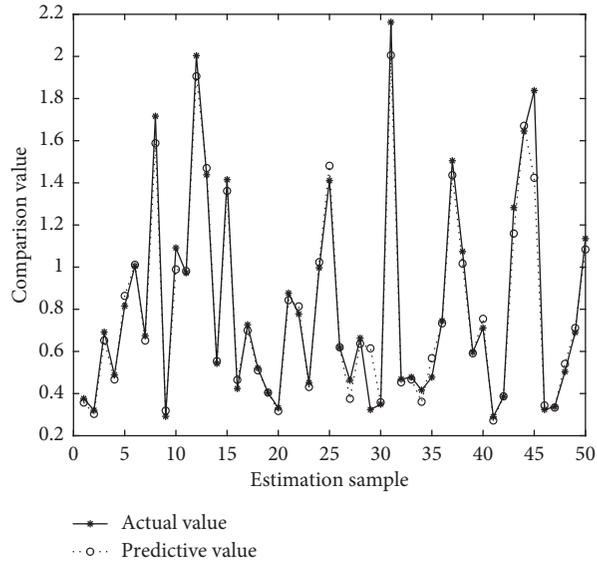


FIGURE 8: Prediction of SVM.

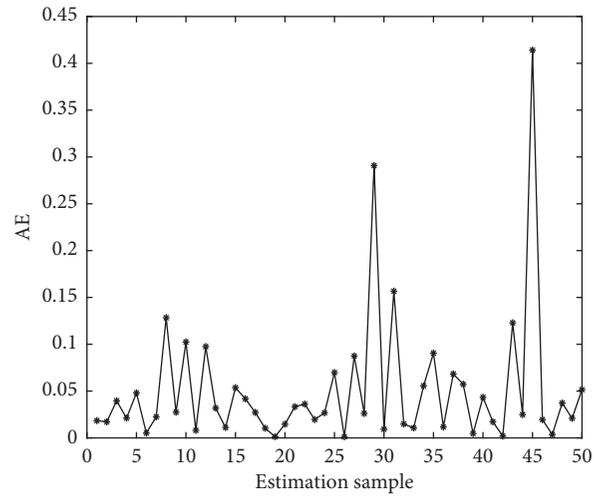


FIGURE 9: AE of SVM.

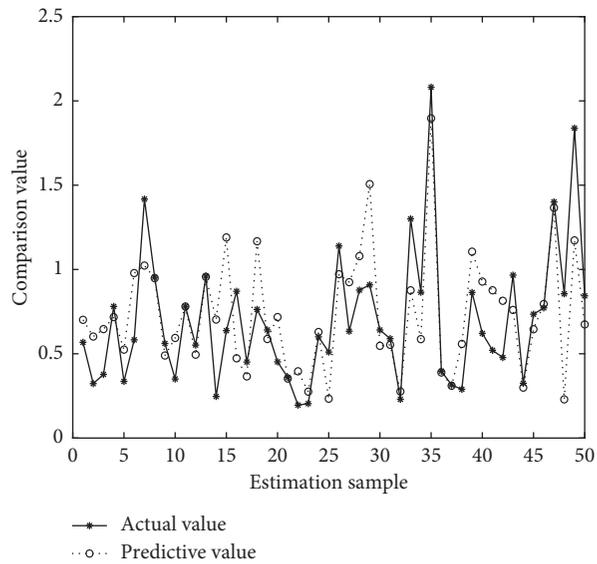


FIGURE 10: Prediction of GR.

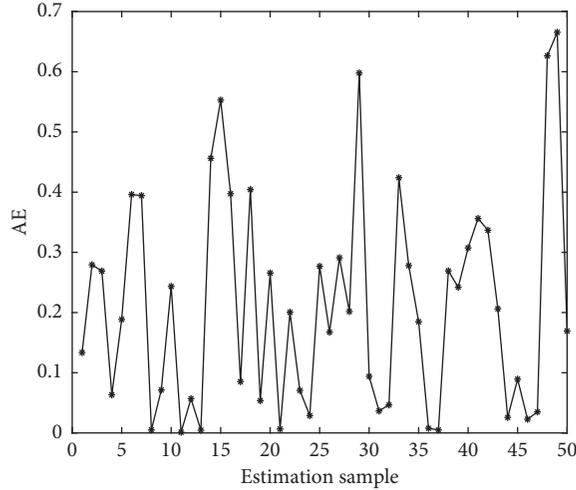


FIGURE 11: AE of GR.

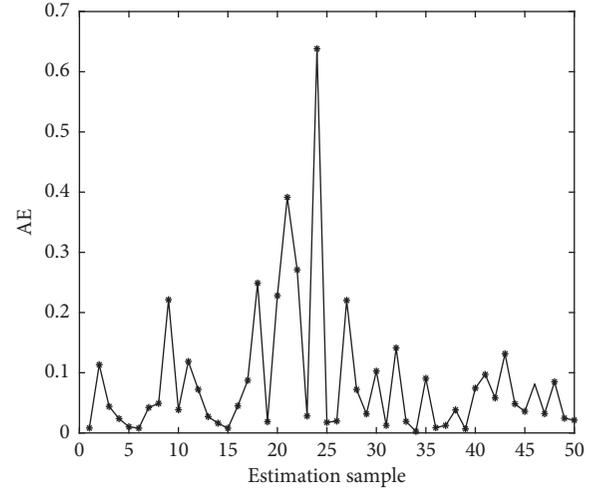


FIGURE 13: AE of RBF.

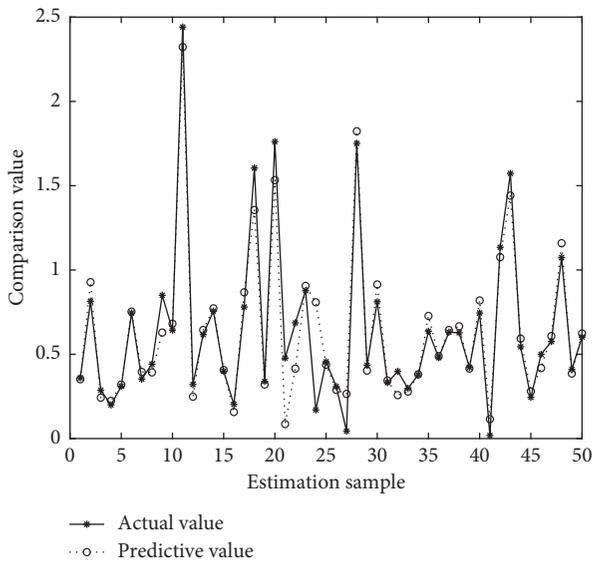


FIGURE 12: Prediction of RBF.

## 5. Simulation Analysis

Here,  $E = 1$ , and  $\mu = W_{RD}/W_{RE}$  (in decibels).

For different channels, we evaluate the ASC performance with  $\bar{\gamma} = 10$  dB in Figure 3. The parameters are given in Table 1. The following observations can be made: (1) increasing  $\mu$  improves the ASC performance; (2) the 2-Rayleigh model can obtain the best ASC performance among the three channels.

In Figures 4–13, we consider SVM, ELM, GR, and RBF [25] methods to compare with the BP network. Table 2 gives the simulation parameters. The MSE and AE of BP are 0.000232889 and 0.04324, which are the lowest MSE and AE in the five methods. This is because BP has the ability to adapt to the time-varying characteristics and enhance the global stability. It has more computing power than other four methods and can be used to solve the rapid optimization problem.

TABLE 2: The parameters for five different methods.

Algorithm	BP	ELM	SVM	GR	RBF
Parameter 1	q: 10 r: 10	q: 18000	c: 128	Spread: 0.054	Spread: 18
Parameter 2	$a_1$ : 0.01		g: 0.011		
Parameter 3	$a_2$ : 0.01				

## 6. Conclusion

An AF relaying scheme was used to improve the ASC performance of medical IoT communication networks in this paper. The ASC expressions were derived. Furthermore, we proposed an intelligent prediction algorithm based on the BP network. The simulation results show that (1) as the  $\mu$  increases, the system's ASC performance becomes better and (2) compared with SVM, ELM, GR, and RBF methods, the proposed BP algorithm can obtain a better MSE and AE.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request and with permission of funders.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 11664043).

## References

- [1] R. Cao, Z. Tang, C. Liu, and B. Veeravalli, "A scalable multi-cloud storage architecture for cloud-supported medical Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1641–1654, 2020.
- [2] H. Wang, L. Xu, Z. Yan, and T. A. Gulliver, "Low-complexity MIMO-FBMC sparse channel parameter estimation for

- industrial big data communications,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3422–3430, 2021.
- [3] D. Meng, X. Wang, M. Huang, L. Wan, and B. Zhang, “Robust weighted subspace fitting for DOA estimation via block sparse recovery,” *IEEE Communications Letters*, vol. 24, no. 3, pp. 563–567, 2020.
- [4] X. Wang, L. Wan, M. Huang, C. Shen, Z. Han, and T. Zhu, “Low-complexity channel estimation for circular and non-circular signals in virtual MIMO vehicle communication systems,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3916–3928, 2020.
- [5] X. Wang, M. Huang, and L. Wan, “Joint 2D-DOD and 2D-DOA estimation for coprime EMVS-MIMO radar,” *Circuits, Systems and Signal Processing*, vol. 40, no. 6, pp. 2950–2966, 2021.
- [6] C. W. Chen, “Internet of video Things: next-generation IoT with visual sensors,” *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6676–6685, 2020.
- [7] P. Gope, Y. Gheraibia, S. Kabir, and B. Sikdar, “A secure IoT-based modern healthcare system with fault-tolerant decision making process,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 862–873, 2021.
- [8] H.-M. Wang, Q. Yang, Z. Ding, and H. V. Poor, “Secure short-packet communications for mission-critical IoT applications,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 5, pp. 2565–2578, 2019.
- [9] M. Torabi and D. Haccoun, “Impact of antenna correlation on the physical layer security of cooperative relaying with OSTBC system,” *IET Communications*, vol. 14, no. 19, pp. 3472–3479, 2020.
- [10] S. Park and H. Son, “Near-Perfect code scrambling with limited key information for wiretap channels,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13410–13423, 2020.
- [11] W. Yang, X. Zhao, and J. He, “Physical layer security and energy efficiency driven resource optimisation for cognitive relay networks,” *IET Communications*, vol. 14, no. 17, pp. 2953–2961, 2020.
- [12] C. Zhang, F. Jia, Z. Zhang, J. Ge, and F. Gong, “Physical layer security designs for 5G NOMA systems with a stronger near-end internal eavesdropper,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13005–13017, 2020.
- [13] X. Lu, W. Yang, X. Guan, and Y. Cai, “DCE-based secure transmission for massive MIMO relay system Against active eavesdropper,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13045–13059, 2020.
- [14] L. Xu, X. Yu, and T. A. Gulliver, “Intelligent outage probability prediction for mobile IoT networks based on an IGWOelman neural network,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1365–1375, 2021.
- [15] L. Xu, H. Wang, and T. A. Gulliver, “Outage probability performance analysis and prediction for mobile IoV networks based on ICS-BP neural network,” *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3524–3533, 2021.
- [16] X. J. Peng, “A spheres-based support vector machine for pattern classification,” *Neural Computing & Applications*, vol. 31, no. 1, pp. 379–396, 2019.
- [17] T. Wu, W. L. Xue, H. Z. Wang et al., “Extreme learning machine-based state reconstruction for automatic attack filtering in cyber physical power system,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1892–1904, 2021.
- [18] L. Xu, H. Wang, H. Li, W. Lin, and T. A. Gulliver, “QoS intelligent prediction for mobile video networks: a GR approach,” *Neural Computing and Applications*, vol. 33, no. 9, pp. 3891–3900, 2021.
- [19] G. K. Karagiannidis, N. C. Sagias, and P. T. Mathiopoulos, “Nakagami: a novel stochastic model for cascaded fading channels,” *IEEE Transactions on Communications*, vol. 55, no. 8, pp. 1453–1458, 2007.
- [20] A. Pandey and S. Yadav, “Physical layer security in cooperative AF relaying networks with direct links over mixed Rayleigh and double-Rayleigh fading channels,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10615–10630, 2018.
- [21] F. K. Gong, Y. Wang, N. Zhang, and P. Ye, “Cooperative mobile-to-mobile communications over double Nakagami-m fading channels,” *IET Communications*, vol. 6, no. 18, pp. 3165–3175, 2012.
- [22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic, San Diego, CA, USA, 2007.
- [23] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, “Wireless information-theoretic security,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2515–2534, 2008.
- [24] X. Liu, Y. Zhou, Z. Wang, and X. Chen, “A BP neural network-based communication blind signal detection method with cyber-physical-social systems,” *IEEE Access*, vol. 6, pp. 43920–43935, 2018.
- [25] X. Zeng, H. Peng, and F. Zhou, “A regularized SNPOM for stable parameter estimation of RBF-AR(X) model,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 4, pp. 779–791, 2018.

## Research Article

# Efficient Algorithms for E-Healthcare to Solve Multiobject Fuse Detection Problem

Ijaz Ahmad <sup>1</sup>, Inam Ullah <sup>2</sup>, Wali Ullah Khan,<sup>3</sup> Ateeq Ur Rehman <sup>2,4</sup>,  
Mohammed S. Adrees,<sup>5</sup> Muhammad Qaiser Saleem,<sup>5</sup> Omar Cheikhrouhou <sup>6</sup>,  
Habib Hamam,<sup>7,8</sup> and Muhammad Shafiq <sup>9</sup>

<sup>1</sup>School of Pattern Recognition and Intelligent System, Shenzhen Institute of Advance Technology (Chinese Academy of Science), Shenzhen, China

<sup>2</sup>College of Internet of Things (IoT) Engineering, Hohai University (HHU), Changzhou Campus, Changzhou 213022, China

<sup>3</sup>School of Information Science and Engineering, Shandong University, Qingdao 266071, China

<sup>4</sup>Department of Electrical Engineering, Government College University, Lahore 54000, Pakistan

<sup>5</sup>College of Computer Science and Information Technology, Al Baha University, Al Baha, Saudi Arabia

<sup>6</sup>College of CIT, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

<sup>7</sup>Faculty of Engineering, Université de Moncton, Moncton NB E1A3E9, Canada

<sup>8</sup>International Institute of Technologie (IIT), Sfax, Tunisia

<sup>9</sup>Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

Correspondence should be addressed to Muhammad Shafiq; shafiq@ynu.ac.kr

Received 7 April 2021; Revised 23 April 2021; Accepted 13 May 2021; Published 27 May 2021

Academic Editor: Han Wang

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

Object detection plays a vital role in the fields of computer vision, machine learning, and artificial intelligence applications (such as FUSE-AI (E-healthcare MRI scan), face detection, people counting, and vehicle detection) to identify good and defective food products. In the field of artificial intelligence, target detection has been at its peak, but when it comes to detecting multiple targets in a single image or video file, there are indeed challenges. This article focuses on the improved K-nearest neighbor (MK-NN) algorithm for electronic medical care to realize intelligent medical services and applications. We introduced modifications to improve the efficiency of MK-NN, and a comparative analysis was performed to determine the best fuse target detection algorithm based on robustness, accuracy, and computational time. The comparative analysis is performed using four algorithms, namely, MK-NN, traditional K-NN, convolutional neural network, and backpropagation. Experimental results show that the improved K-NN algorithm is the best model in terms of robustness, accuracy, and computational time.

## 1. Introduction

E-healthcare is a broad term consisting of improvements in medical services detected by digital technology. In E-healthcare, there are various diseases including diabetes, cancer, and stroke, as well as machine learning systems for diagnosing these diseases and integrating AI and so forth. Recently, deep learning classifiers have had excellent target detection performance in various electronic healthcare applications such as diagnosis of heart disease based on heart image, detection of cancer, and various EEG data sets

classification including chest X-ray, diabetic retinopathy, and skin cancer. FUSE-AI, a startup company based in Hamburg, has developed a system that can detect and classify tumors in MRI scans. Hamburg has developed the FUSE-AI system, which can classify tumors based on MRI scans using machine learning classifiers.

Object detection is an artificial intelligence technology related to image processing and computer vision, which can detect various objects (vehicles, buildings, and people) in specific categories in digital videos and images. In-depth study of object detection areas includes pedestrian detection,

face detection, and traffic signal detection [1–7]. Use cases ranging from personal safety to work efficiency subdivide object detection into a wide range of areas [8–13].

Although huge innovations are taking place in the field of healthcare, many problems must still be solved, especially heterogeneous data fusion, mobile data transmission, and analysis. Facial recognition is a form of face detection that can be used as a high security measure that allows only certain people to enter, such as highly sensitive areas in government buildings. Many applications of multitarget detection and classification exist, such as face detection [14], people counting [15], vehicle detection [16], and identification of good and bad food [17].

The combination of image processing and artificial neural network is a huge combination that can be used endlessly for various purposes. In the past few years, a lot of work has been done in these two fields. Therefore, this technology has become the focus of any artificial intelligence company, and it is also very necessary for the governments of many countries/regions in the world [18]. In addition, researchers are also focusing on the modernization of other scientific fields such as artificial intelligence and information technology [19–31].

Object detection methods are usually divided into two methods, machine learning and deep learning. Machine learning methods include support vector machine (SVM) classification strategies and deep learning classification methods used by various neural networks such as convolutional neural networks (CNN), neural backpropagation networks (or backscatter communication networks), and K-nearest neighbor network (KNN) [32]. The object detection model can be divided into two parts, extraction function and classification [33–35].

In the visualization of objects, feature extraction needs to extract various visual features to provide a reliable representation [18, 33, 34, 36, 37]. In fact, these features represent similarities with the human brain and complex cells [12]. It is difficult to find a powerful feature extractor to extract all the features of an object and construct it manually. However, in classification, the classifier distinguishes the target object from other representative features or categories for visual recognition. Generally, support vector machine (SVM) [38], deformable part-based model (DPM) [39], and AdaBoost [40] are good classifiers.

This article compares three different neural networks to conduct CNN, BP, and K-NN experiments to find out which is the best in fusion detection and can also evaluate training and output time and preprocessing time. Most of the researches in the field of target detection and surveillance in wireless sensor networks (WSN) focus on single or multiple target detection. However, there are few studies aimed at monitoring and detecting fuse objects. The purpose of this research is to design a robust algorithm to effectively classify fuse objects in a single image. First, preprocess the data and prepare it for each algorithm, because each algorithm uses the data differently, especially the backpropagation network. Then train the neural network to correctly classify each fuse. We used deep learning in our research work.

The following are the key contributions of this work:

- (i) We apply the new method by adding robust neighbor to the training samples to modify the K-NN model.
- (ii) A comparison study is accomplished by using modified K-NN with classical K-NN, B-PN, and CNN models based on accuracy and time complexity in E-healthcare object (fuse) detection. Our study proved that, compared with other models, the modified K-NN model has higher accuracy and less time complexity.
- (iii) We propose using the modified K-NN model for future research in E-healthcare object detection applications.

The rest of this article is structured as follows. Section 2 introduces related work. Section 3 introduces the methodology. Section 4 introduces the modified K-nearest neighbor (M-KNN) algorithm. Section 5 discusses the experimental results, and Section 6 discusses the comparison results. Finally, Section 7 summarizes the paper and provides prospects for future research work.

## 2. Related Work

**2.1. CNN.** CNN is an artificial neural network with similar architectures to ordinary neural networks [41]. CNNs have three sections of architectures, input neurons, learnable weights, and biases. CNN has been widely applied in the area of handwritten character recognition. CNN is an artificial neural network with an architecture similar to ordinary neural networks [41]. CNN has a three-part structure, namely, input neurons, learnable weights, and deviations, and is widely used in the field of handwritten character recognition [42–46] and face recognition [47]. The explicit assumption that the image is input is made by the CNN architecture, which allows us to encode certain attributes into the architecture. Make the forwarding function more effective and greatly reduce the number of network parameters. The general structure has special benefits for using neural networks to solve different problems. CNN has a clear biological structure; the early work of Hubel and Wiesel is based on the ANN model of cat's visual cortex application [42, 45, 48].

CNN has been successfully applied to many classification applications such as traffic signal detection to detect various traffic signals. CNN uses supervised learning algorithms, so it can predict better classification results. The CNN architecture is shown in Figure 1.

**2.2. BP.** Backpropagation neural network (BPN) is a neural network used as a neural forwarding network [49–51]. The BP algorithm has a multilayer architecture mapping, which can transmit information between the forwarding and output layers. Therein, it transmits the signal through the hidden layer and adjusts the weight based on the delta rule between the actual output of the ANN and the predicted output, thereby minimizing the error rate. Any nonlinear function with arbitrary precision can usually be

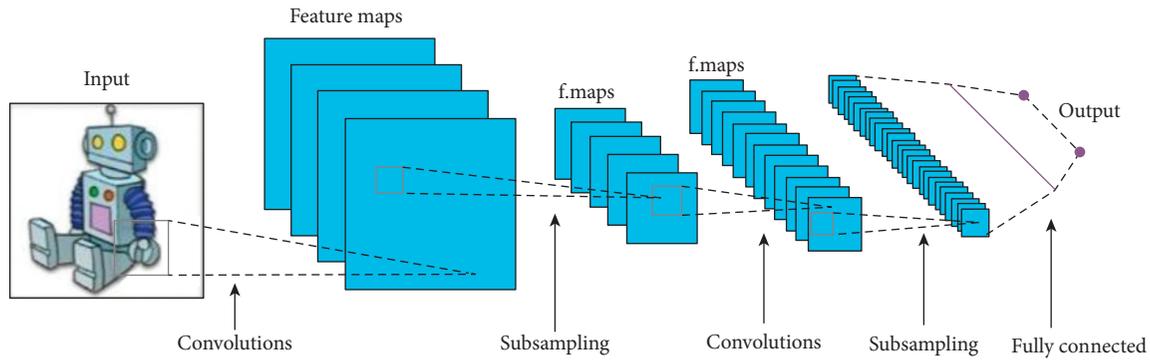


FIGURE 1: CNN architecture.

approximated by a hidden BPN layer [52]. For predicting complex nonlinear systems, this feature makes BPN famous. BP algorithm is a classification model with two gradient descent methods and mean square error. When the minimum error is reached, the square number and connection weight will be adjusted. In the process of BP algorithm, the input signal is provided to the network first, and then the sample training process is carried out. Then calculate the gradient of the error signal value [53].

The architecture of BPN consists of two parts called forward propagation (or output layer) and backpropagation (or error signal). In forward propagation, the input layer is propagated to the input signal and forwarded to the hidden layer through the output layer. The weight and offset values remain constant throughout the process. The next layer is affected by the operation signal. If the predicted output layer and the output layer do not match, the operation signal will shift to the backpropagation of the error signal. On the other hand, backpropagation is defined as the difference between the expected output value and the actual output layer value. The weight and offset value will be constantly changed and added to reduce the difference between the actual output layer value and the predicted output layer value. In this step, the error signal is propagated and used as the input signal. The BP architecture is shown in Figures 2 and 3, respectively.

**2.3. K-NN.** In object detection, the K-NN algorithm has been successfully used for classification and regression, which is an unassisted method. In the K-NN algorithm, the input depends on the value of  $K=0$  and  $K=n$  in the training process room. Due to its high accuracy and simplicity and being easy to understand and easy to implement, the K-NN algorithm has been successfully used in many data analysis applications, including pattern recognition databases, information retrieval, and machine learning [54, 55]. Therefore, this is why the top ten algorithms for data mining in recent decades have been K-NN algorithms [56]. This algorithm is used for data processing, data classification, and clustering. D. Vijayalaksmi applies the K-NN algorithm to the classification and clustering of diabetes in the medical application of diabetes. Based on the accuracy and purity of the sample, the solution was compared between the K-NN and K-means methods. The implementation results of these models show that the K-NN solution is

more effective and reliable than K-means. Satheesh and Patel's dynamic K-NN is proposed as a data classifier, combined with the principle of object-oriented programming [57]. This question is classified in a single consolidated form during the training phase. Compared with traditional K-NN, it implements and classifies solutions more effectively. The K-NN architecture is shown in Figure 4.

All object detection methods must consider the inherent uncertainty of the size, position, and structure of the target in the natural scene image. Viola Jones detector [58], histogram of oriented gradients (HOG) detector [34], and deformable part-based model (DPM) [59] are just some traditional methods of object detection in the natural image scenes.

These methods mainly rely on manually extracted object features to determine the parameters of the algorithm. However, due to the rapid development of deep learning in recent years, many object detection methods based on this advanced technology have been developed [60]. By properly training their network architecture, these methods have proven that they can accurately locate object regions in natural image scenes. The object detection method based on R-CNN, the object detection method based on SSD, and the object detection method based on YOLO are three types of object detection methods.

Researchers have proposed a variety of target detection algorithms. For example, an algorithm based on multiscale deformable CNN is proposed in [61]. The authors compare and study mainstream object detection algorithms to solve the problems faced by current methods. This study confirmed results that are comparable to or better than the latest methods. Deep convolutional networks usually are used to obtain multiscale features and solve geometric transformations by integrating deformable convolutional structures. These networks fuse multiscale features through sampling to introduce the final object recognition and region regression.

In [62], the authors introduced another object detection through flowing and fusion. The authors propose an end-to-end deep neural network (DNN) and flow fuse tracker (FFT) tracking technology, which solves two methods of tracking problems, such as target fusion and target flow. To be more precise, the FlowTracker DNN module obtains an unlimited number of target directional motions from the pixel-level optical flow in the target stream. On the other hand, the FuseTracker DNN module refines and tracks the target

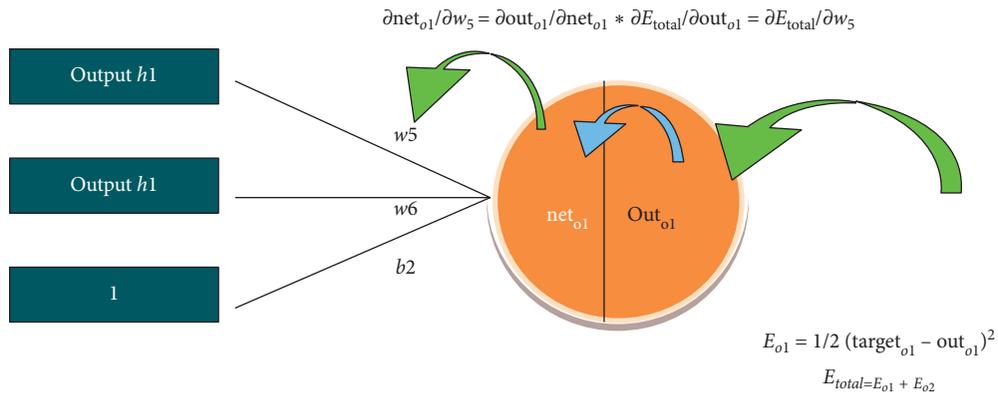


FIGURE 2: BP architecture (output layer).

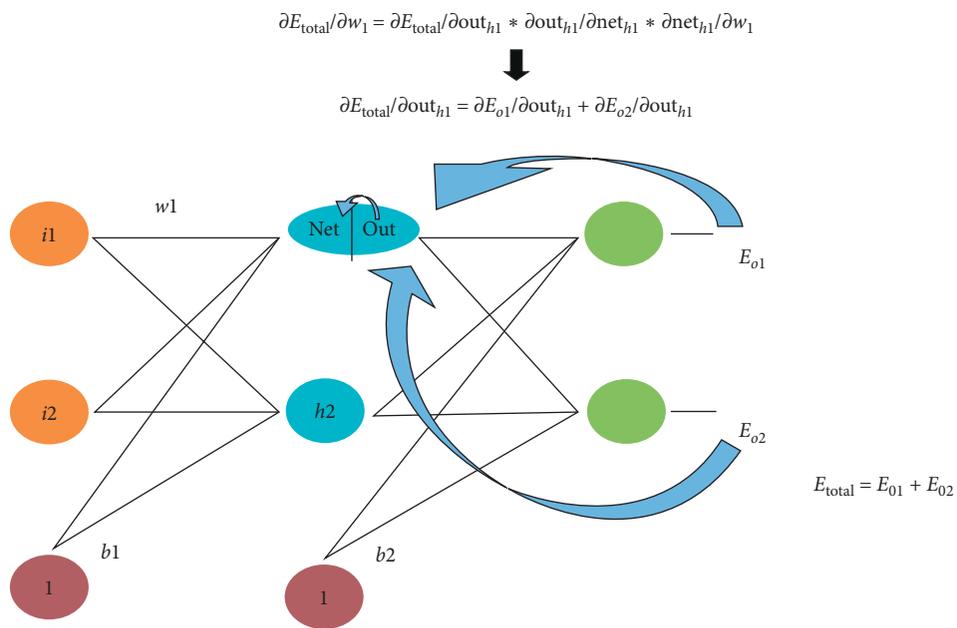


FIGURE 3: BP architecture (hidden layer).

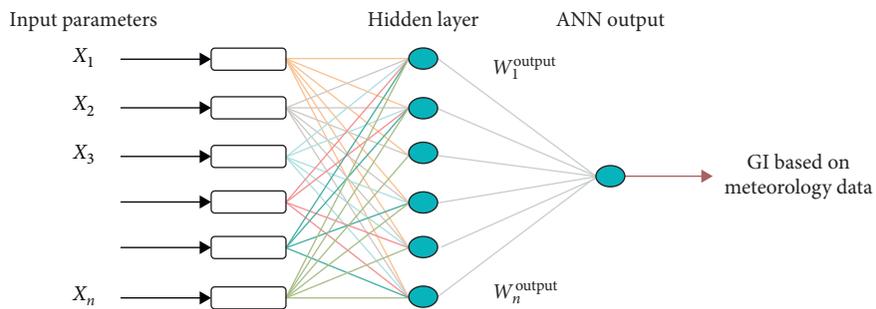


FIGURE 4: Typical K-NN architecture.

proposed by FlowTracker and integrates frame-by-frame object detection in target fusion, instead of trusting one of the two incorrect sources of target recommendations. Because FlowTracker can detect complex target motion patterns, FuseTracker can improve and integrate the targets in FlowTracker and the detector.

### 3. ANN Models Evaluation Methodology

In this section, three different neural networks are evaluated for CNN, BP, and K-NN experiments to find which is better in fusion detection and to determine training time, execution time, or preprocessing time. The processing process of

some networks is not simple, requiring a lot of preprocessing to adjust our data and then put it on the network. The whole process is shown in Figure 5. The data is preprocessed first, and then the demand for the network is evaluated, because each network can be processed differently in different ways. For example, the input signal required by CNN may be simple and straightforward, but processing BP may not be the case.

With the latest developments in computer vision models that focus on deep learning, object detection applications are easier to create than ever. In addition to significant performance improvements, these methods also utilize massive image data sets to reduce the need for large data sets. The current method focuses on a complete end-to-end pipeline and allows real-time use, so performance has been greatly improved. If you prepare the data in vector form, provide the same data set for each network and start the training process. After the training is completed, give the network a record from the main vector data set, and let the network predict the performance. First check the single fuse detection, and if the result is good, provide the entire test data set to the network to predict the fuse and the empty fuse slot, and allow the network difference between the two items. Finally, evaluate the accuracy and timing of each network.

The following describes in detail the experimental process from preprocessing to output detection for fuse detection.

**3.1. Preprocessing.** In the preprocessing process of fuse target detection, it is important to understand the basis of the specific neural network which our experiment can effectively perform. In this study, the overall results of these three models are evaluated for the first time based on 126 images of a small data set. It contains a fuse artifact and an empty fuse slot. There are several methods of detecting objects. The experiment tested three models of CNN, BP, and K-NN [9]. There are more models that can also be tested, but these three models were selected for the research work because these models have been widely used in many major applications around the world. In order to deal with these models first, the processing steps must be divided into the following questions.

Training the model with more and more images and determining the best accuracy is a difficult task. As the number of images increases from 500 to 1,000, the accuracy and time complexity of the model also increase. The model takes some time to load the data set. As a result, training time will increase. To balance the two, we use 500 images to train the model.

- (1) Good data set of images: in order to train any neural network model, good experimental results require a large number of images to provide the model. The “good number” has no specific number, only the model can be trained and tested, and the number of images continues to increase until the model produces more reliable results and fewer errors. But increasing the number of images is not free. If the computer is not powerful enough, it will cost us more

training time. Therefore, the collection of image data sets is a difficult task.

- (2) Basic knowledge of each model: it is necessary to understand the basic knowledge of each model, because this will allow us to choose which model is the most effective to complete our work, thus spending relatively less training time and better performance. The size of the image needs to be adjusted, and usually the entire size of a single image or the entire data set needs to be changed, so it is very important to have basic knowledge of image processing; and, as mentioned above, each network takes data as input in a different format or size, in order to understand the internal data of the image and how to use it.

Before continuing to compare results, you must have basic knowledge of how neural networks work, especially how convolutional neural networks and deep neural networks work and how to train these levels. What are activation and loss functions? How do they work and how to get different results for different functions? Another important parameter in a neural network is the learning rate, which directly affects the model learning process in any neural network. During training, the average learning rate and the average failure rate should be opposite to each other. For example, if the loss is less than 0.05 and the learning rate is close to 1, it is assumed that the model is well trained and the training process can be stopped.

**3.2. Data Set (Input Data).** The fuse object data comes from the Natural Science Foundation of Jilin Province. This data set includes 3000 images of fuses. In order to train a custom model, the experiment needs to prepare the data set of the required target and in some cases also needs to prepare other classes and categories, so that the network can easily identify the required target and nontarget objects. Each network requires different forms and sizes of data to be fed into the network first. It is necessary to understand how this particular network works, and then do some data preprocessing and prepare for the network.

Usually, the problem of object detection related to the X problem is raised. There is no need to implement the model. The important view is how many images are needed in the training phase. Each type of representative image must have a large number (e.g., >100 and possibly >1000). The special context is required for the image perspective in pattern recognition. In traffic signal detection, it will be required to meet the image requirements under each condition, such as camera conditions and different weather conditions. The training process may be affected by these contexts, causing the model to fail to train correctly, if we have a small amount of data.

The following is a fuse board containing at least 39 fuse objects, as shown in Figure 6; the fuse board extracted fuse objects from different images and prepared our small data set, which mixed an empty fuse slot.

First, extract each fuse object, convert it to grayscale, and then adjust it to  $200 \times 200$  in a small image to save training time.



starting MKNN adds a strong neighbor to the training and calculation of the sample. Then define the value after assigning the weights. Figure 7 shows the pseudocode of the K-NN modified algorithm. Validation error rate and training error rate are two parameters, and their values are between  $k=1$  and  $k=7$ . If we increase the value of  $k=1$  to  $k=7$ , the error rate and verification error decrease, so we use  $k=7$  to get the best results.

## 5. Experimental Results

This section introduces the experimental methods used to detect fuses, how to develop the data set, and how to transform the data before feeding it to each network and then discusses the training process and finally see the results. The training process is gradually repeated in CNN, BP, and KNN to detect objects. In addition, it also explains how the training process is carried out, how much time is spent on each network, and the percentage of the fuse recognition rate, which means whether the network is a fuse or an empty fuse slot determined by the percentage.

*5.1. Classification of K-Nearest Neighbour Network.* MK-NN is a very-easy-to-implement and simple architecture. It determines the closest K neighbor based on the minimum distance between the query text and the training sample. After collecting K-nearest neighbors, most of these K-nearest neighbors will be used to predict problem events. The classifier MK-NN performs the two following steps. First, the entire computational data set is run between each training observation. Then, in the training data, name the K point closest to the range. Note that K is usually an odd number to prevent this from happening. Second, estimate the conditional likelihood of each category, that is, the score of the AA midpoint with the category name (note that  $I(x)$  is an indicator function; when the statement  $xx$  is true, its value is 1; otherwise it is 0). Finally, our input  $xx$  is assigned to the most probable class. Another way to look at K-NN is to treat it as a measure of the decision boundary (i.e., the boundary of two or more categories) and then divide the data into training, testing, and verification, as shown in Table 1.

- (1) Training process (K-NN): before starting training, the process needs to have various  $K$  values to see the value of  $K$  at which the network performs well. For each value of  $K$ , train the model and return the accuracy results, as shown in Table 2. If  $K=1$ , the accuracy is as high as 98%, while, for  $K=7$  and  $K=9$ , it remains at 90% and so on, as shown in Table 3.

It is difficult to find the value of  $k$  in MK-NN. A smaller value means that noise will have a greater impact on efficiency, while a higher value will make it computationally expensive. If the number of categories is 2, the data scientist will usually choose an odd number, and another easy way to choose  $k$  is to set  $k = \text{sqrt}(n)$ . The nearest neighbor algorithm uses a very basic classification method. When comparing with the new example, look at the training data to find the  $k$  training example that is the closest to the

new example. Thereafter, the most common class tags (in these  $K$  training examples) are assigned to test cases. Therefore,  $K_i$  is just the number of "voting" neighbors of the test example class. If  $k=1$ , the test example has the label as the closest example in the training set. If  $k=3$ , it will check whether there are three closest categories in the label and will assign the most common label (i.e., at least twice) and so on to get larger  $K_s$ .

- (2) Preparing data (MK-NN): preparing data for MK-NN is the same as the two networks listed above, but the only difference is that MK-NN obtains data in a two-dimensional (2D) array format. For CNN, the data needs a tensor format, which can have a larger size, including color scheme parameters. But MK-NN will get the data in (sample number, dimension) format. This is why once the data size changes, the process needs to take additional steps to convert the entire training and test data set to this format.
- (3) Predictions (MK-NN): if the training process for MK-NN has been completed, the process needs to verify the network data set and determine how many accurate predictions the network makes. The sample results of  $Y$  prediction are shown in Table 4. The following prediction shows a perfect fit with the test results. The first value array shows the test results, and the second value array shows the values expected by the model.
- (4) Complete predictions on test data set (MK-NN): we evaluate the entire data set on the web and view the overall results. If the model is wrong, or if it is uncertain at a certain stage whether the model is the desired target, the vertical line will reflect the trustworthiness of the model. The following is the result of the complete test data set shown in Figure 8. The predictions of the test data set are perfect, and none of the predictions are lower than 98%. Overall, the model works well, and there is no single expected result with a confidence level of less than 98%.

The accuracy is  $\text{tp} = (\text{tp} + \text{FP})$  ratio, where the number of true positives is  $\text{tp}$  and the number of false positives is  $\text{fp}$ . The ability of the classifier to intuitively not mark negative samples as positive is what we called accuracy. Recall the ratio of  $\text{tp} = (\text{tp} + \text{fn})$ , where the true positive number is  $\text{tp}$  and the wrong negative number is  $\text{fn}$ . Intuitively, recall is the ability of the classifier to identify all positive samples. The  $F1$  score can be defined as the weighted average of precision and recall, where the highest score of  $F1$  is 1, and the worst is 0. The relative accuracy and recall rate of the  $F1$  grades are comparable. The formula for  $F1$  score is as follows:  $F1 = 2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall})$ .

- (5) Precision and recall (K-NN): the precision and recall results based on the above perfect results are listed in Table 5.

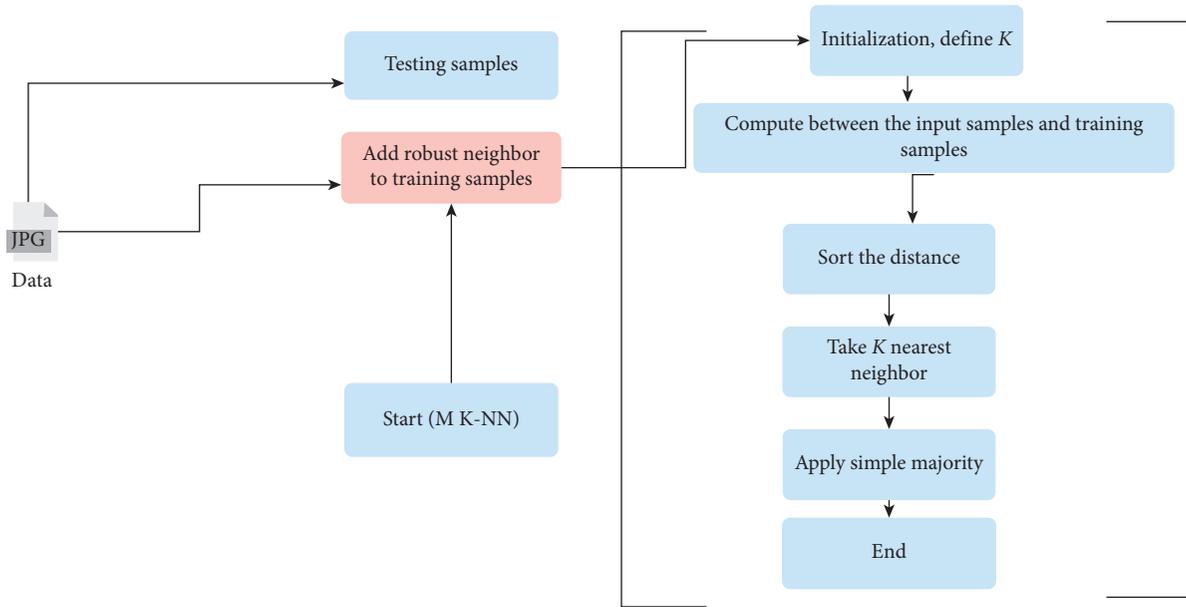


FIGURE 7: Proposed work model (modified K-NN).

TABLE 1: Training test and validation split.

Training data points	Training data labels	Validation data points	Test data labels
90	90	10	10

TABLE 2: Predicted labels.

True label	Predicted label	Predicted items	Detected class
[ 0 1 ]	[ 0.0087723 0.991227 ]	0.0 1.0	Not target
[ 0 1 ]	[ 9.99775e - 01 2.244791e - 04 ]	1.0 0.0	Target
[ 0 1 ]	[ 9.99931e - 01 6.88999e - 05 ]	1.0 0.0	Target
[ 0 1 ]	[ 9.998918e - 01 1.081199e - 04 ]	1.0 0.0	Target
[ 0 1 ]	[ 9.991754e - 01 8.245995e - 04 ]	1.0 0.0	Target

TABLE 3: Training process (K-NN).

K-value	Accuracy
K=1	Accuracy = 98%
K=1	Achieved highest accuracy of 98% on validation data
K=3	Accuracy = 98%
K=1	Achieved highest accuracy of 98% on validation data
K=5	Accuracy = 98%
K=1	Achieved highest accuracy of 98% on validation data
K=7	Accuracy = 98%
K=1	Achieved highest accuracy of 98% on validation data

TABLE 4: Predictions accuracy on test data (K-NN).

Test data set items	Correctly predicted	Incorrectly predicted	Accuracy (%)
26	25	1	98

5.2. *Classification of Convolutional Neural Network.* CNN is the simplest network to implement all image processing models. Therefore, CNN is always the programmer's first choice in terms of basic object recognition. In the process of preprocessing, training, and displaying the results, when a simple item is found, there are fewer problems with the K value. Each process will be discussed in more detail below. For the size of the output tensor (image) of the regular layer,  $O$  is the size or width of the output image.  $I$  represents entering the size or width of the image.  $K$  is the kernel size or width used in the traditional layer.  $N$  is kernel number.  $S$  is stride of the convolution operation.  $P$  is padding. The size of the output image ( $O$ ) is determined by

$$O = \frac{I - K + 2Px^2}{S} + 1. \quad (1)$$

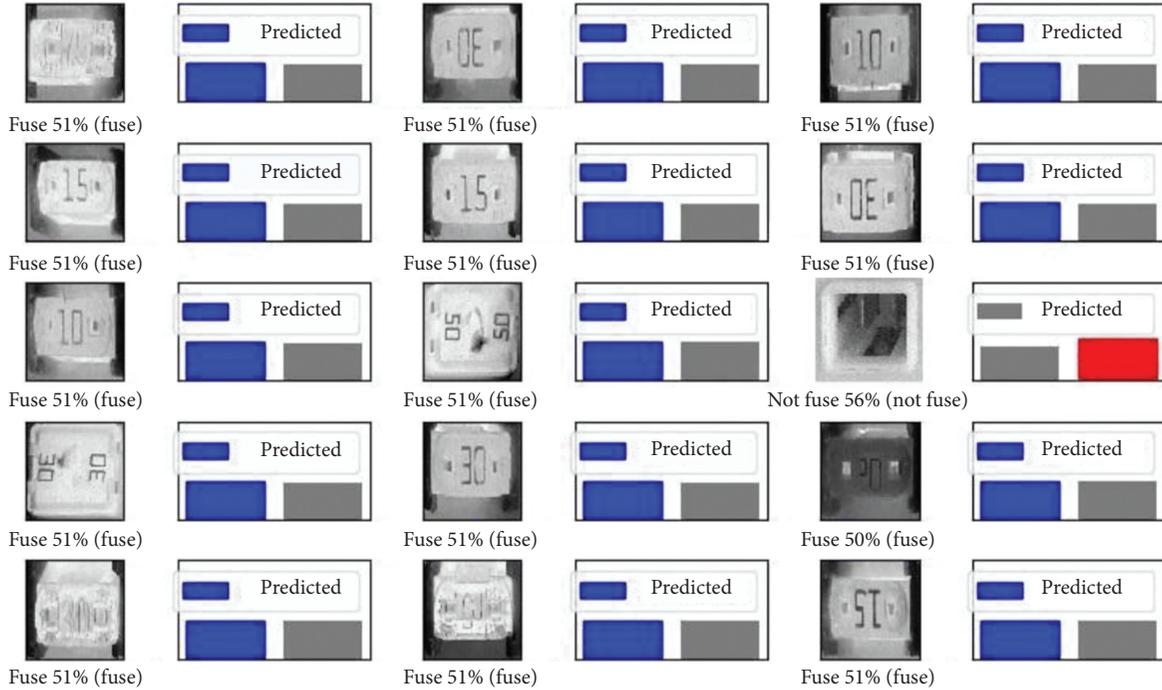


FIGURE 8: Complete predictions on test data set (BP).

TABLE 5: Precision and recall.

Label	Precision	Recall	F1-source	Support
0	1.00	1.00	1.00	3
1	1.00	1.00	1.00	23
Av/total	1.00	1.00	1.00	26

The number of channels in the output image is equal to the number of cores.

The input image is  $227 \times 227 \times 3$ , and the 96 kernels in the first convolutional layer are  $11 \times 11 \times 3$ . The path is 4 and the padding is 0. Therefore, the ratio of the output image is just after the first convolutional layer.

$$O = \frac{277 - 11 + 2x^0}{44} + 1 = 55. \quad (2)$$

So, the output image size is  $55 \times 55 \times 96$  (one channel for each kernel).

- (1) Preparing data set (CNN): in this step, prepare a common data set for all models, because you will not be able to compare our results because you want to change the size of the data set or image. First, delete each fuse object from the key image, and then save it to the data set. The next step is to resize it to  $200 \times 200$  in the small image, and then convert the RGB to grayscale to save training time, as shown in Figure 9. First load the data set and build the marker. But, before creating these systems, we need to create labels first.

A sample of the image vector is shown in Figure 10. Next, make a paste image; for example, from the first image to 104 images, the target object must be

marked as 0 or 1 or as needed. Mark the remaining images from 104 to 126 as 1 or 0 as untargeted images. Select the first group of images as 0; the remaining images must be labeled as 1. There is a simple data division and a clear boundary between the target object and no target object. The division of the training test is 20% of the data used for testing and 80% of the data used for preparation. Figure 11 shows the pseudocode of the modified K-NN.

Once the data set preparation process is completed using data tags, the data and labels will be changed to change where the fuse objects are located and the order in which no data set objects are merged. It has greater significance to the network and contributes to stronger and more effective preparations. This method will repeat these steps to ensure that each of our models remain straight so that the results can be easily matched. Finally, the data set is divided into training data set and test data set. The training data set will be provided to the network for training purposes, and the test data set will be provided to the network to verify the quality of the data on the network.

**5.3. Training Process (CNN).** The provided data can be used during the training process to train and execute the model using the fit () function. Before the training process, we should understand the number of layers required to enable this feature, as shown in Figure 12. Then compile and call the function.

On the other hand, the accuracy value should be equal to 1, and the error value should be changed to 0. During the training process, failure and accuracy must be reversed. The same is true for the lack of legitimacy and accuracy of

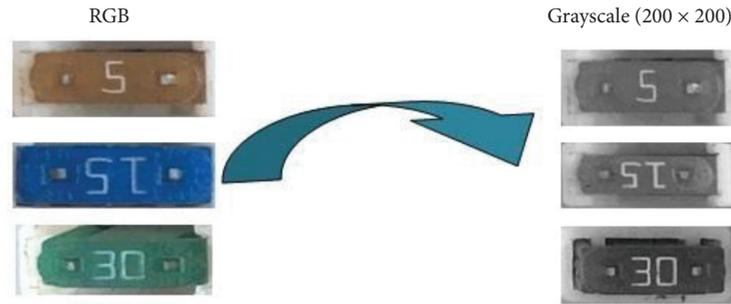


FIGURE 9: Preparing data for CNN.

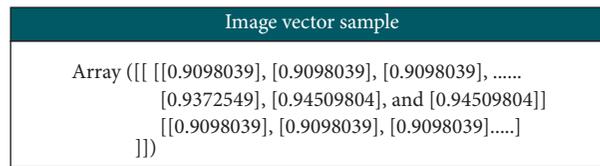


FIGURE 10: Image vectors and labels.

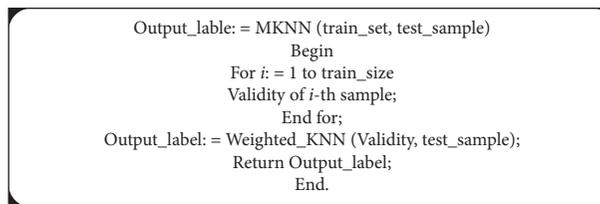


FIGURE 11: Pseudocode (modified K-NN).

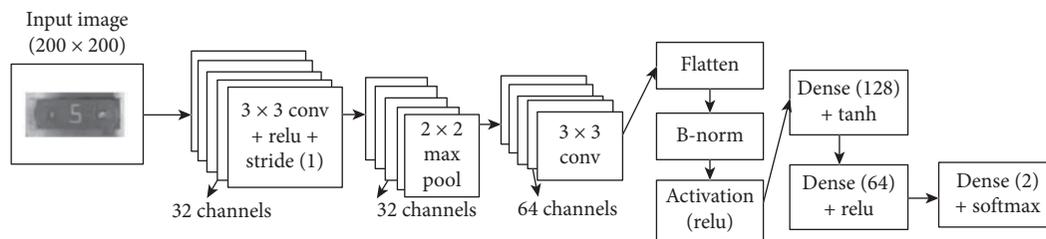


FIGURE 12: Preparing for training and defining the network.

verification. As long as the loss is reduced during the training phase, the checkpoint/weight will be stored as CNN.hdf5. Next time use these weights to prevent the lengthy training phase and start using the new weights saved locally to make predictions on our artifacts.

5.4. *Predictions (CNN)*. If the CNN training is complete, provide the test data set to the network and determine how many accurate predictions the network has made. An example prediction screen is shown in Table 2. The following table shows the real label [0 1], for example, the predicted label [0.008 0.991] and the predicted object [0 1], and finally shows the detected category [not target]. Repeat this process for each item in the test data set and evaluate the results as shown below. As shown in Table 2,

the time used for each period/prediction should be 1.23 seconds.

5.5. *Complete Predictions on the Test Data Set (CNN)*. If a prediction is made through a single image test, then it is time to test the entire data set on the network and see the overall results. If the model is wrong, or it is uncertain whether the model is the desired goal at a certain stage, the vertical line will reflect the trustworthiness of the model. For example, if the model determines that 96% is the required target, the first bar will be almost complete, and the second bar will be almost invisible in the forecast chart. However, if the model predicts that the object accounts for 51% of the appropriate object, the height of the two bars will be almost the same. The following is the result of the complete test data set shown in Figure 13. The

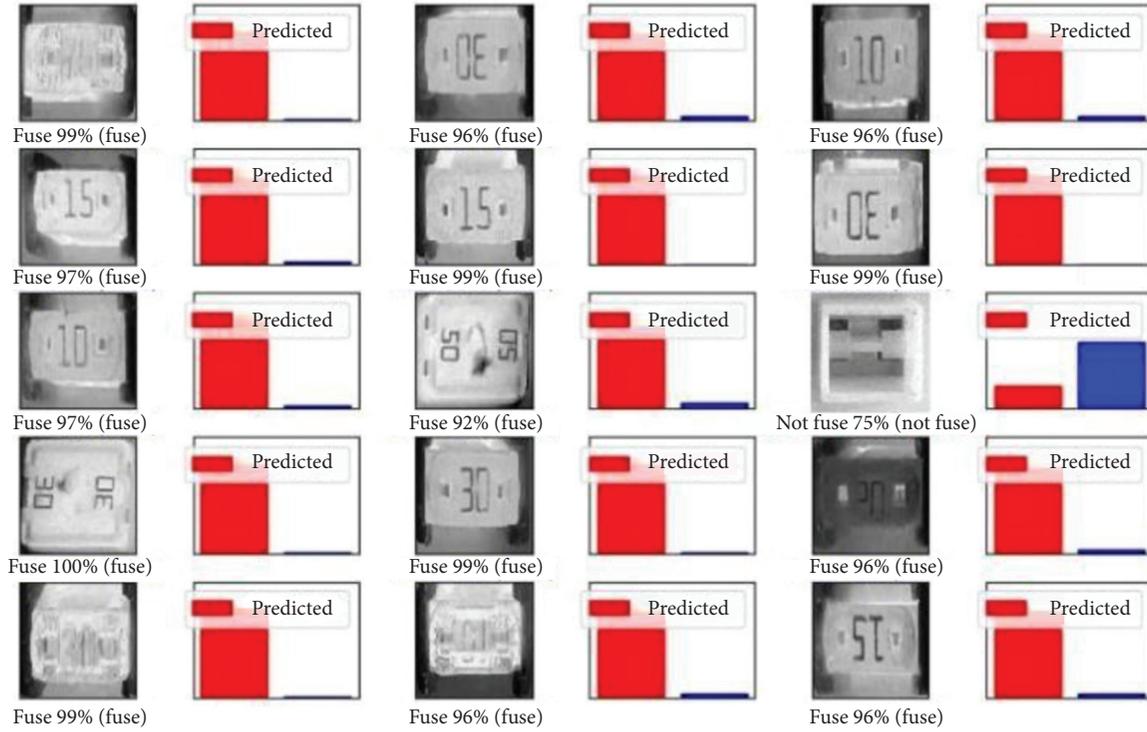


FIGURE 13: Complete predictions on the test data set (CNN).

prediction of the test data set is satisfactory because, on average, all artifacts required have a confidence level of more than 96%.

**5.6. Accuracy and Loss Check (CNN).** Figures 14(a) and 14(b) show the accuracy and failure of the test and training data check. The accuracy of the training data is almost 0.8, and the accuracy of the test data is 0.9. Of course, the loss is initially high and gradually decreases for training and test results. If the number of periods increases and the size of the data set increases, the results will of course be updated. Or at least increase the number of periods to further improve efficiency. In the training process, more data and more opportunities, direct loss, and accuracy provide more conditions for the model. In this way, the network can learn and create more spaces between the target and the nontarget.

**5.7. Classification of Backpropagation Network.** Backpropagation may be the second choice of most python programmers, because python does not have a built-in framework or library to implement it easily. However, python scripts must be run to implement the backpropagation model. In other words, BP is not a simple network. Most of the steps taken are the same as CNN's image preparation, training process, and evaluation process. The difference is that the size of the input node during the network feed is  $200 \times 200$  to accommodate the input size data of each image. The key steps taken to train the network are as follows. The overall BP equation for this problem is given as follows:

$$\delta^{Lj} = \frac{\partial C}{\partial a^{Lj}} \sigma'(Z^{Lj}),$$

$$\delta^1 = ((w^1 + 1)T\delta^1 + 1) \sigma'(Z^1),$$

$$\frac{\partial C}{\partial b^{1j}} = \delta^{1j},$$

$$\frac{\partial C}{\partial w^{1jk}} = a^{1-1} k^{\delta^{1j}},$$

(3)

where  $k$  represents overall nodes in the layers above node  $j$ ,  $dj$  is the expected output of node  $j$ , and  $yj$  is the actual output. In addition,  $w$  is the weight between the inputs, and  $\delta L$  is the only data required to calculate the weight gradient of layer  $L$ , and then we can calculate the previous  $\delta^{L-1}$  layer and replicate it recursively, and  $\sigma$  is the error term of node  $j$ .

- (1) Preparing data (BP): just like preparing data for CNN, the same measures are taken to prepare data for BP network. Take the same data set and the same dimensions; nothing has changed. The division of the training test is 20% of the data used for testing and 80% of the data used for preparation. First, you need to extract each fuse object from the main photo and save it to our data set. The next step is to resize it to a small image of  $200 \times 200$ . Therein, we convert RGB to grayscale to save training time.

If the command is executed, python will display the progress of the training process. Each epoch will show the learning rate and the elapsed time (in

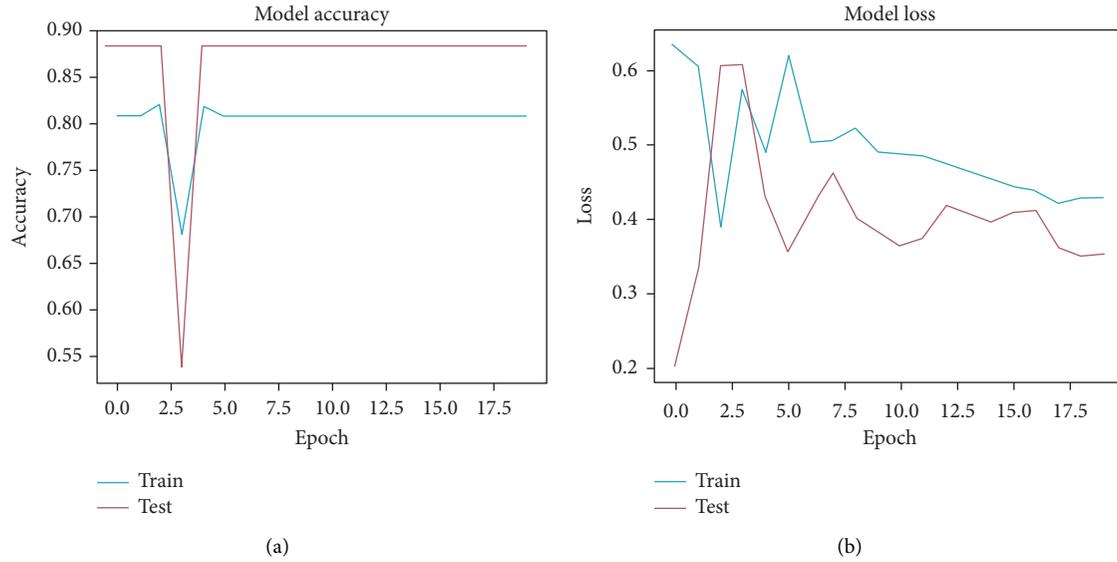


FIGURE 14: (a) Accuracy of the model. (b) Loss of the model.

seconds). The BP network must automatically change the learning rate in each period because it must fit the input vector and output, and then measure the error. Based on the error, the learning rate will be adjusted and the network will be tested in the next period. The training phase is shown in Table 6.

- (2) Predictions (BP): if the training process is completed, BP will use the network test data set to test how many accurate predictions the network produces. The results of the predicted samples are shown in Table 7. The following prediction shows that these two objects are not fusion objects, and the rest are fuse objects on the network.
- (3) Complete predictions on test data set (BP): once the identification of a single image test is completed, it is now possible to test the entire data set on the network and view the overall results. If the model is wrong, or at a certain stage it is uncertain whether the model is the desired goal, the vertical line will reflect the trustworthiness of the model. For example, if the model determines that 96% is the required target, the first bar will be almost complete, and the second bar will be almost invisible in the forecast chart. However, if the model predicts that the object accounts for 51% of the appropriate object, the height of the two bars will be almost the same. The entire result of the test data set is shown in Figure 8. The prediction on the test data set is good because the confidence of all necessary items is not too high, and the average is only slightly higher than 51%.
- (4) MSE and learning rate (BP): Figures 15(a) and 15(b) show the relationship between MSE (mean square error), learning rate, and accuracy of training results. When the learning rate is maintained at 0.5, the MSE

TABLE 6: Sample epochs from training (BP).

Epoch	Learning rate	Elapse time
1/60	0.000683	207.932893
3/60	0.000676	213.214195
7/60	0.000669	218.458495
10/60	0.000662	223.785800
13/60	0.000656	229.071102
16/60	0.000649	234.325403
18/60	0.000643	239.591704
20/60	0.000636	244.842004
25/60	0.000630	250.073303
30/60	0.000624	255.384607

TABLE 7: Predictions on the test data set.

True label	Predicted label	Predicted items	Detected class
[1 0]	[0.50812526 0.49187474]	[1.0 0.0]	Fuse
[1 0]	[0.50812513 0.49187487]	[1.0 0.0]	Fuse
[1 0]	[0.50812734 0.49187266]	[1.0 0.0]	Fuse
[1 0]	[0.50801862 0.49198138]	[1.0 0.0]	Fuse
[1 0]	[0.50801227 0.49198773]	[1.0 0.0]	Fuse
[1 0]	[0.50800363 0.49199637]	[1.0 0.0]	Fuse
[1 0]	[0.50801598 0.49198402]	[1.0 0.0]	Fuse
[1 0]	[0.5080938 0.4919062]	[1.0 0.0]	Fuse
[1 0]	[0.44388368 0.55611632]	[0.0 1.0]	Not fuse
[1 0]	[0.50800662 0.49199338]	[1.0 0.0]	Fuse
[1 0]	[0.50801372 0.49198628]	[1.0 0.0]	Fuse

decreases rapidly and then becomes unstable but still remains high and low. In addition, since no single object is detected, the accuracy is 0%. When the learning rate drops to 0.01, the MSE drops rapidly and remains consistent throughout the rest of the period. At the same time, because the model can detect target and nontarget objects at the same time,

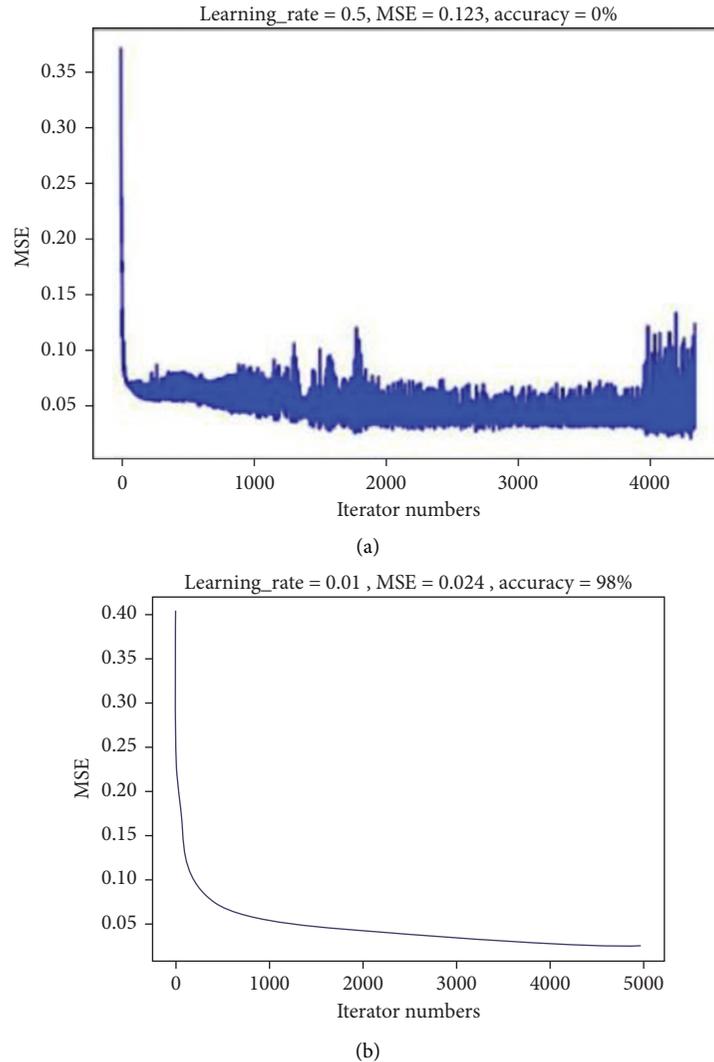


FIGURE 15: (a) MSE versus L-Rate = 0.5. (b) MSE versus L-Rate = 0.01.

the accuracy is increased to 98%. The accuracy of the training data is almost 0.61, and the test data is about 0.51. Initially, of course, for both the training and the test results, the loss is very high and gradually decreases, but, by increasing the number of periods, the performance will be further improved.

## 6. Results and Discussion

Based on the above data, it is not difficult to conclude that K-NN has the best model among the three models. According to training time, prediction time, and prediction accuracy, KNN performed well and provided a result of 98%. As you can see, each model in the table below has several parameters. The first is whether the model needs to preprocess the image, and the second parameter is the training time, which is one of the most important parameters in neural network training. The third key parameter is estimated time. This parameter is very important for real-time systems, because the prediction must be made in real time, so delayed prediction may cause

a big problem; and combine all the parameters in the final parameters with the percentage accuracy. Therefore, looking at all the data below, it is easy to assume that the performance of K-NN is very effective and that the predicted result is much higher than the planned performance. In addition, from an application point of view, processing time, memory, and resource requirements play a key role in selecting a classifier, and K-NN can provide results comparable to results that are computationally inexpensive and easy to classify.

For the computationally demanding CNN and BP, when selecting a classifier, applications where processing time and resource requirements are key considerations should also consider hardware capabilities. Table 8 and Figure 16 show the overall comparative analysis of these three models. Throughout the research process, we came to the conclusion that, compared with traditional K-NN, CNN, and BP, in terms of accuracy (%), time complexity (min), and processing high dimensionality, performance-based modified K-NN is a more effective model data set.

TABLE 8: Overall comparisons of the three algorithms with modified K-NN in fuse detection.

Deep learning models	Accuracy (%)	Time complexity (mins)	Required data processing	Handle high-dimensional data set
Modified K-NN	98–100	15	Yes	Yes
Classical K-NN	96–98	20	Yes	Yes
CNN	96	30	No	No
BP	51	25	No	No

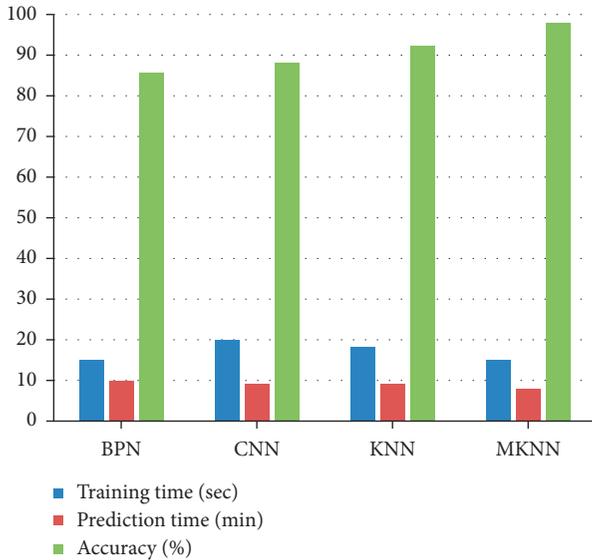


FIGURE 16: Overall comparisons of CNN, BP, and K-NN.

## 7. Conclusion

This paper proposes an improved K-NN (MK-NN) model to improve object detection in E-healthcare applications. We analyzed the proposed MK-NN model through CNN, BP, and traditional K-NN models to find out which model is effective after multiple training and testing rounds. Our analysis results show that, based on the key parameters of the network, training time, prediction time, and prediction accuracy, MK-NN is the best model among the four models. MK-NN performs well in training time and prediction time (20 seconds, 20 minutes) and provides 98% accuracy. Future research should discover and incorporate the proposed Model Architects (MK-NN) model and compare the performance in other computer vision applications. Similarly, an interesting implementation is to use multilayer output to fully connect the layers to increase the vector size of the function to improve performance.

## Data Availability

Since the funding project is not closed and related patents have been evaluated, the simulation data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data, upon the approval of patents after project closure, will be considered by the corresponding author.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

Omar Cheikhrouhou acknowledges Taif University for its support under the project Taif University Researchers Supporting Project (TURSP-2020/55), Taif University, Taif, Saudi Arabia.

## References

- [1] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time Chinese traffic sign detection algorithm based on modified yolov2," *Algorithms*, vol. 10, no. 4, p. 127, 2017.
- [2] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012.
- [3] I. Ullah, Y. Liu, X. Su, and P. Kim, "Efficient and accurate target localization in underwater environment," *IEEE Access*, vol. 7, pp. 101415–101426, 2019.
- [4] J. Zhang, Q. Huang, H. Wu, and Y. Liu, "A shallow network with combined pooling for fast traffic sign recognition," *Information*, vol. 8, no. 2, p. 45, 2017.
- [5] B. Moustapha, "The effect of propagation delay on the dynamic evolution of the bitcoin blockchain," *Digital Communications and Networks*, vol. 6, no. 2, pp. 157–166, 2020.
- [6] Y. Wu, Y. Liu, J. Li, H. Liu, and X. Hu, "Traffic sign detection based on convolutional neural networks," in *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Dallas, TX, USA, 2013.
- [7] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: a multi-class classification competition," in *Proceedings of the 2011 International Joint Conference on Neural Networks*, IEEE, San Jose, CA, USA, 2011.
- [8] M. T. Sadiq, X. Yu, Z. Yuan et al., "Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain-computer interfaces," *IEEE Access*, vol. 7, pp. 171431–171451, 2019.
- [9] H. H. Aghdam, E. J. Heravi, and D. Puig, "A practical approach for detection and classification of traffic signs using convolutional neural networks," *Robotics and Autonomous Systems*, vol. 84, pp. 97–112, 2016.
- [10] I. Ullah, S. Qian, Z. Deng, and J.-H. Lee, "Extended Kalman filter-based localization algorithm by edge computing in wireless sensor networks," *Digital Communications and Networks*, 2020, In press.
- [11] X. Yao, F. Farha, R. Li, I. Psychoula, L. Chen, and H. Ning, "Security and privacy issues of physical objects in the IOT: challenges and opportunities," *Digital Communications and Networks*, 2020, In press.
- [12] L. Bourdev and J. Malik, "Poselets: body part detectors trained using 3D human pose annotations," in *Proceedings of the 2009*

- IEEE 12th International Conference on Computer Vision*, IEEE, Kyoto, Japan, 2009.
- [13] P. Ren, W. Fang, and S. Djahel, "A novel Yolo-based real-time people counting approach," in *Proceedings of the 2017 International Smart Cities Conference (ISC2)*, IEEE, Wuxi, China, 2017.
- [14] X. Zhao, Y. Ni, and H. Jia, "Modified object detection method based on Yolo," in *Proceedings of the CCF Chinese Conference on Computer Vision*, 2017.
- [15] Y. LeCun, B. Boser, J. S. Denker et al., "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [16] S. Yu, J. Liu, J. Wang, and I. Ullah, "Adaptive double-threshold cooperative spectrum sensing algorithm based on history energy detection," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 4794136, 12 pages, 2020.
- [17] X. Su, I. Ullah, X. Liu, and D. Choi, "A review of underwater localization techniques, algorithms, and challenges," *Journal of Sensors*, vol. 2020, Article ID 6403161, 24 pages, 2020.
- [18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [19] F. Jameel, W. U. Khan, M. A. Jamshed, H. Pervaiz, Q. Abbasi, and R. Jantti, "Reinforcement learning for scalable and reliable power allocation in SDN-based backscatter heterogeneous network," in *Proceedings of the 2020 IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE, Toronto, Canada, 2020.
- [20] W. U. Khan, X. Li, M. Zeng, and O. A. Dobre, "Backscatter-enabled NOMA for future 6G systems: a new optimization framework under imperfect SIC," *IEEE Communications Letters*, vol. 25, no. 5, pp. 1669–1672, 2021.
- [21] F. Lin, J. Liu, Q. Wu et al., "FMNet: feature mining networks for brain tumor segmentation," in *Proceedings of the 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, Portland, OR, USA, 2019.
- [22] W. U. Khan, X. Li, A. Ihsan, M. A. Khan, V. G. Menon, and M. Ahmed, "NOMA-enabled optimization framework for next-generation small-cell IoT networks under imperfect SIC decoding," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [23] X. Li, Y. Zheng, W. U. Khan et al., "Physical layer security of cognitive ambient backscatter communications for green internet-of-things," *IEEE Transactions on Green Communications and Networking*, p. 1, 2021, In press.
- [24] W. U. Khan, F. Jameel, N. Kumar, R. Jantti, and M. Guizani, "Backscatter-enabled efficient V2X communication with non-orthogonal multiple access," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1724–1735, 2021.
- [25] F. Jameel, W. U. Khan, N. Kumar, and R. Jantti, "Efficient power-splitting and resource allocation for cellular V2X communications," *IEEE Transactions on Intelligent Transportation Systems*, p. 1, 2020, In press.
- [26] W. U. Khan, J. Liu, F. Jameel, V. Sharma, R. Jantti, and Z. Han, "Spectral efficiency optimization for next generation NOMA-enabled IoT networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15284–15297, 2020.
- [27] F. Jameel, S. Zeb, W. U. Khan, S. A. Hassan, Z. Chang, and J. Liu, "NOMA-enabled backscatter communications: toward battery-free IoT networks," *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 95–101, 2020.
- [28] W. U. Khan, F. Jameel, M. A. Jamshed, H. Pervaiz, S. Khan, and J. Liu, "Efficient power allocation for NOMA-enabled IoT networks in 6G era," *Physical Communication*, vol. 39, Article ID 101043, 2020.
- [29] Z. Ali, W. Farooq, W. U. Khan, M. Qureshi, and G. A. S. Sidhu, "Artificial intelligence techniques for rate maximization in interference channels," *Physical Communication*, vol. 47, Article ID 101294, 2021.
- [30] W. U. Khan, F. Jameel, T. Ristaniemi, S. Khan, G. A. S. Sidhu, and J. Liu, "Joint spectral and energy efficiency optimization for downlink NOMA networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 645–656, 2019.
- [31] S. Yu, W. U. Khan, X. Zhang, and J. Liu, "Optimal power allocation for NOMA-enabled D2D communication with imperfect SIC decoding," *Physical Communication*, vol. 46, Article ID 101296, 2021.
- [32] W. U. Khan, N. Imtiaz, and I. Ullah, "Joint optimization of NOMA-enabled backscatter communications for beyond 5G IoT networks," *Internet Technology Letters*, vol. 4, no. 2, p. e265, 2021.
- [33] W. Li, Y. Wang, Z. Jin, K. Yu, J. Li, and Y. Xiang, "Challenge-based collaborative intrusion detection in software defined networking: an evaluation," *Digital Communications and Networks*, 2020, In press.
- [34] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, IEEE, San Diego, CA, USA, 2005.
- [35] R. A. Naqvi, M. Arsalan, A. Rehman, A. U. Rehman, W.-K. Loh, and A. Paul, "Deep learning-based drivers emotion classification system in time series data for remote applications," *Remote Sensing*, vol. 12, no. 3, p. 587, 2020.
- [36] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proceedings of the 2002 International Conference on Image Processing*, vol. 1, IEEE, Rochester, NY, USA, 2002.
- [37] I. Ullah, X. Su, J. Zhu, X. Zhang, D. Choi, and Z. Hou, "Evaluation of localization by extended Kalman filter, unscented Kalman filter, and particle filter-based techniques," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8898672, 15 pages, 2020.
- [38] C. Cortes and V. Vapnik, "Support vector machine," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [39] Y. Freund and R. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," in *Proceedings of the 1995 European Conference on Computational Learning Theory (EuroCOLT)*, Barcelona, Spain, 1995.
- [40] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models. Sion multimodal latent dirichlet allocation for image annotation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, 2010.
- [41] E. Sackinger, B. E. Boser, J. Bromley, Y. LeCun, and L. D. Jackel, "Application of the ANNA neural network chip to high-speed character recognition," *IEEE Transactions on Neural Networks*, vol. 3, no. 3, pp. 498–505, 1992.
- [42] Y. Bengio, Y. LeCun, and D. Henderson, "Globally trained handwritten word recognizer using spatial representation, convolutional neural networks, and hidden Markov models," in *Advances in Neural Information Processing Systems*, pp. 937–944, MIT Press, Cambridge, MA, USA, 1994.

- [43] M. M. Kamal, I. Ullah, A. Ashraf, and N. Ullah, "Designing band notch features in ultra-wideband antenna," in *Proceedings of the 2017 7th IEEE International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE)*, IEEE, Xi'an, China, 2017.
- [44] R. R. Swain, P. M. Khilar, and T. Dash, "Multifault diagnosis in WSN using a hybrid metaheuristic trained neural network," *Digital Communications and Networks*, vol. 6, no. 1, pp. 86–100, 2020.
- [45] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.
- [46] B. Fasel, "Robust face analysis using convolutional neural networks," in *Object Recognition Supported by User Interaction for Service Robots*, vol. 2, pp. 40–43, IEEE, 2002.
- [47] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, no. 3, pp. 574–591, 1959.
- [48] I. Khan, Y.-b. Tian, I. Ullah, M. M. Kamal, H. Ullah, and A. Khan, "Designing of E-shaped microstrip antenna using artificial neural network," *International Journal of Computing, Communications and Instrumentation Engineering*, vol. 5, no. 1, 2018.
- [49] S. K. Haider, A.-M. Jiang, S. Ashraf, and I. Ullah, "Study on detection of p300 ERP component in EEG signals and algorithms," in *Proceedings of the 3rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017)*, Atlantis Press, Gangzhou, China, 2017.
- [50] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Proceedings of the 2011 International Joint Conference on Neural Networks*, IEEE, San Jose, CA, USA, 2011.
- [51] I. Khan, Y.-B. Tian, Inamullah, H. Vllah, S. U. Rahman, and M. M. Kamal, "Design annular ring microstrip antenna based on artificial neural network," in *Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, IEEE, Xi'an, China, 2018.
- [52] I. Ullah, Y. Shen, X. Su, C. Esposito, and C. Choi, "A localization based on unscented Kalman filter and particle filter localization algorithms," *IEEE Access*, vol. 8, pp. 2233–2246, 2019.
- [53] A. Aslanargun, M. Mammadov, B. Yazici, and S. Yolacan, "Comparison of arima, neural networks and hybrid models in time series: tourist arrival forecasting," *Journal of Statistical Computation and Simulation*, vol. 77, no. 1, pp. 29–53, 2007.
- [54] K. Zhou and Y. Kang, *Neural Network Model and Matlab Simulation Program Design*, Publishing House of Tsinghua University, Beijing, China, 2005.
- [55] E. Blanzieri and F. Melgani, "Nearest neighbor classification of remote sensing images with the maximal margin principle," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 6, pp. 1804–1811, 2008.
- [56] J. Chen and J. Shao, "Jackknife variance estimation for nearest-neighbor imputation," *Journal of the American Statistical Association*, vol. 96, no. 453, pp. 260–269, 2001.
- [57] A. Satheesh and R. Patel, "Dynamic nearest neighbours classifier for integrated data using object oriented concept generalization," *International Journal of Simulation-Systems, Science & Technology*, vol. 11, no. 1, 2010.
- [58] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [59] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Anchorage, AK, USA, June 2008.
- [60] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014.
- [61] D. Cao, Z. Chen, and L. Gao, "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks," *Human-Centric Computing and Information Sciences*, vol. 10, p. 14, 2020.
- [62] J. Zhang, S. Zhou, X. Chang et al., "Multiple object tracking by flowing and fusing," 2020, <https://arxiv.org/abs/2001.11180>.

## Research Article

# Cross-Database Micro-Expression Recognition Exploiting Intradomain Structure

Yanliang Zhang , Ying Liu , and Hao Wang 

*School of Physics and Electronic Information, Henan Polytechnic University, Jiaozuo 454150, China*

Correspondence should be addressed to Hao Wang; [roihaio@163.com](mailto:roihaio@163.com)

Received 10 February 2021; Revised 31 March 2021; Accepted 7 May 2021; Published 24 May 2021

Academic Editor: Lingwei Xu

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

Micro-expressions are unconscious, faint, short-lived expressions that appear on the faces. It can make people's understanding of psychological state and emotion more accurate. Therefore, micro-expression recognition is particularly important in psychotherapy and clinical diagnosis, which has been widely studied by researchers for the past decades. In practical applications, the micro-expression recognition samples used in training and testing are from different databases, which causes the feature distribution between the training and testing samples to be different to a large extent, resulting in a drastic decrease in the performance of the traditional micro-expression recognition methods. However, most of the existing cross-database micro-expression recognition methods require a large number of model selection or hyperparameter tuning to select better results from them, which consumes a large amount of time and labor costs. In this paper, we overcome this problem by exploiting the intradomain structure. Nonparametric transfer features are learned through intradomain alignment, while at the same time, a classifier is learned through intradomain programming. In order to evaluate the performance, a large number of cross-database experiments were conducted in CASMEII and SMIC databases. The comparison of results shows that this method can achieve a promising recognition accuracy and with high computational efficiency.

## 1. Introduction

Micro-expression is a spontaneous expression that is associated with self-defense mechanisms that occur when a person attempts to conceal an internal emotion and can neither be faked nor suppressed [1]. A spontaneous facial expression was first identified by Haggard and Isaacs [2] in 1966. In 1969, Ekman and Friesen [3] found that a depressed patient who smiled a lot had occasional frames of very painful expressions during repeated viewing of a conversation. The researchers call these quick, unconscious, spontaneous facial movements that people make when they experience strong emotions micro-expressions. Therefore, micro-expressions have a lot of potential utilization value for emotion recognition tasks, e.g., clinical diagnosis [4], marital relationship prediction [5], communication negotiation [6], and teaching assessment [7, 8].

It has been found that micro-expression, a unique and common tiny facial movement that lasts for a very short

time, is usually only 1/25 to 1/5 second [9]. Since it is almost hard to detect and recognition with the naked eye, it is necessary to carry out automatic recognition of micro-expressions by computer. So far, although the micro-expression recognition experiments done by the researchers have achieved some results, the micro-expression samples for training and testing come from the same database. In many practical applications, in fact, the micro-expression recognition samples for training (source) and testing (target) come from different databases. Due to the differences in race, gender, age, camera equipment, and recording environment between the two databases, the original consistency of feature distribution in traditional micro-expression recognition was seriously damaged, resulting in most micro-expression recognition methods being unsatisfactory. Previously, the recognition of cross-database micro-expression related to it has been widely followed by most researchers, and many relevant and effective methods have been

proposed. Selective transfer machine (STM) [10, 11], a new method, is discovered by using the target sample to study a group of weights of the source sample and pick the most appropriate parameter values  $\lambda$ ,  $\varepsilon$ ,  $\beta$ , etc., where the weighted source sample has the same or similar distribution of features as the target sample. In the work [12], a dictionary is jointly learned through samples from source domain and target domain, balancing the parameters using the trade-offs and optimizing the model, so that the source and target domains have the same or similar feature distribution. Zheng and Zhou [13] and Zheng and Zhou [14] discovered a transformed subspace learning framework to handle cross-pose and cross-database cases in facial expression recognition and selected the optimal solution by adjusting parameters. To address the challenge of cross-database micro-expression recognition (CDMER), Zong et al. [15] first proposed the use of domain adaptive methods and introduced some new methods such as the domain regeneration framework and target sample regenerator. The experiment yielded good results but still needs to tune the individual parameters. Li et al. proposed target-adapted least-squares regression (TALSr) to learn the regression coefficient matrix based on the source domain data information and optimize TALSr to fit the target micro-expression database by weighing parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  [16]. Although some existing methods are effective for the task of cross-database micro-expression recognition, it is worth noting that most of them require parameter tuning as well as model selection. Some adaptive adjusting algorithms can be used for implementing the parameter tuning and model selection such as the recursive algorithms [17–25] and the iterative algorithms [26–32]. However, due to the inability to determine the best model and its optimal hyperparameters, it usually takes a lot of labor cost to use parameter grid search method or cross-validation strategy to search. In addition, such selection is somewhat subjective and often performs radially very differently on different datasets. Applying it to real situations remains a challenge. Therefore, how to find a more efficient way to carry out this work is particularly important.

In this paper, we propose the Intradomain Structure Domain Adaptation (IDSDA) method to solve this problem. Nonparametric transfer features are learned by aligning the source and target domain subspaces, and then, a classifier is learned through intradomain programming to predict the sample labels of the target domain, without model selection and hyperparameter tuning. For one thing, this method saves time and labor costs. For another, it improves the recognition accuracy and practicality of CDMER.

In general, this paper contains the following main contributions:

- (1) IDSDA is proposed to carry out cross-database micro-expression recognition. IDSDA mainly learns nonparametric transfer features through intradomain alignment and a classifier through intradomain programming, which is simple to operate and enhances practicability

- (2) Different from the current popular CDMER, this method does not require model selection and hyperparameter tuning, saving labor cost and greatly shortening experiment time
- (3) We have conducted extensive CDMER experiments, and the experimental results prove that our method has advantages in micro-expression recognition, and the recognition accuracy is improved

The rest of this paper is arranged as follows. Section 2 reviews the work about unsupervised domain adaptation and state-of-the-art CDMER. In Section 3, we describe the Intradomain Structure Domain Adaptation (IDSDA) for CDMER in detail. For better evaluation of this method, extensive experiments and analyses on SMIC and CASMEII are shown in Section 4. At last, the conclusion and the future planning of this paper are drawn in Section 5.

In addition, a large number of symbols are used in this paper. For clarity, we show the frequently used notations and corresponding descriptions in Table 1.

## 2. Related Work

Domain adaptation is actually a method of transfer learning. The principle of domain adaptation is a learning process in which models acquired in the old domain are applied to the new domain. Since the source domain has sufficient data information, it usually uses all the data to fully learn the internal structure of the data. Therefore, how to effectively reduce the distribution difference between the source domain and target domain is the essence of domain adaptation. Jing et al. [33] proposed Adaptive Component Embedding (ACE), for resolving large domain discrepancies. In the work of [34], Maximum Density Divergence was proposed to minimize the interdomain divergence and maximize the intradomain density. In [35], Heterogeneous Domain Adaptation (HAD) method is found, which optimizes both feature discrepancy and distribution discrepancy in a uniform objective function. In addition, Li et al. [36] discovered a new method, Locality Preserving Joint Transfer, which considers the knowledge transfer between the feature and sample level. In this way, the distribution divergence between the two domains is reduced, while preserving the neighborhood relationship of samples and making it robust to outliers. Similar to the domain adaptive classification, according to whether the target domain has the label information or not, cross-database micro-expression recognition is usually divided into unsupervised and semisupervised categories. The former can only use the source domain's sample label information for training, while the latter can combine some existing label information of the target domain with the known sample information of the source domain, so the information of the latter is sufficient. Obviously, the former has more practical applications than the latter, so this paper focuses on unsupervised cross-database MER.

For unsupervised cross-database MER, although methods such as domain regeneration in the original label space (DRLS) and domain regeneration in the original

TABLE 1: Notations and corresponding descriptions.

Notations	Descriptions
$H$	Coefficient matrix
$d$	The dimension of the feature vector
$\lambda$	The trade-off parameter to balance objective function
$\mu$	The trade-off parameter to control the sparsity of the coefficient matrix
$X_s$	The feature matrix of source domain micro-expression samples
$X_t$	The feature matrix of target domain micro-expression samples
$n_s$	The number of source domain samples
$n_t$	The number of target domain samples
$Q$	The whitened source domain features
$A^s$	Regenerated source domain micro-expression samples features
$A^t$	Regenerated target domain micro-expression samples features
$M$	Probability annotation matrix
$c$	The class of the label
$l_c$	The center of class $c$

feature space with unchanged target domain (DRFS-T) have achieved good results by Zong et al. [15], they all require hyperparameter tuning. Taking the DRFS-T method as an example. Firstly, a regenerator  $Z(\cdot)$  is learned, which is able to regenerate the source and target domain samples. Then, the classifier (SVM) can be trained according to the data information in the source domain, and the target domain sample class can be predicted based on the classifier. But the constraints of the regenerator are as follows.

First, the regenerated target domain samples remain unchanged in the feature space:

$$\min_Z \|X_t - Z(X_t)\|_F^2. \quad (1)$$

$X_t \in \mathbb{R}^{d \times n_t}$  and  $X_s \in \mathbb{R}^{d \times n_s}$  are the target domain and source domain micro-expression samples, and  $n_t$  and  $n_s$  are the number of target and source domain samples. Besides,  $d$  represents the dimension of the feature vector.

Second, the regenerated source domain and target domain samples have the same feature distribution:

$$\min_Z \|X_t - Z(X_t)\|_F^2 + \lambda f_Z(X_s, X_t). \quad (2)$$

For the purpose of regulating the balance of the two terms in the objective function of DRFS-T, the trade-off parameter  $\lambda$  is introduced, while  $f_Z(X_s, X_t)$  is the regular term.

Then, Zong et al. [15] exploited the characteristics of regenerative Hilbert space and its MMD distance and finally optimized them as

$$\min_{H \in \mathbb{R}^{(n_s+n_t) \times d}} \|X_t - H^T K_t\|_F^2 + \lambda \left\| \frac{1}{n_s} H^T K_s 1_s - \frac{1}{n_t} H^T K_t 1_t \right\|_2^2 + \mu \|H\|_1. \quad (3)$$

$1_s$  and  $1_t$  are column vectors with dimensions  $n_s$  and  $n_t$ , respectively. In addition,  $K_s$  and  $K_t$  are the kernel matrices of the source and target domain samples.  $\|H\|_1$  is the L1 parametric constraint on the coefficient matrix  $H$ , where  $\mu$  is the sparsity trade-off parameter controlling  $H$ .

To sum up, DRFS-T involves hyperparameters  $\lambda$  and  $\mu$ , which requires huge time cost to pick the optimal result. However, IDSDA can solve this problem well. The IDSDA method mainly utilizes the subspace method for intradomain alignment followed by nonparameter intradomain programming, so as to achieve the purpose of improving the recognition accuracy of CDMER.

### 3. IDSDA for Cross-Database Micro-Expression Recognition

**3.1. Problem Definition.** Domain adaptation is specifically defined as follows: given a labeled source domain  $\Omega_s = \{(x_j^s, y_j^s)\}_{j=1}^{n_s}$  and an unlabeled target domain  $\Omega_t = \{x_i^t\}_{i=1}^{n_t}$ .  $x_j^s$  and  $x_i^t$  are the samples in the source and target domains, respectively, while  $n_s$  and  $n_t$  are the number of their samples. It is assumed that the feature space, label space, and conditional distribution are the same, i.e.,  $\mathcal{X}_s = \mathcal{X}_t$ ,  $\mathcal{Y}_s = \mathcal{Y}_t$ , and  $Q_s(y_s|x_s) = Q_t(y_t|x_t)$ , but the margin distribution is different  $P_s(x_s) \neq P_t(x_t)$ . The core idea is to learn classifiers using  $\Omega_s$  with sufficient information for predicting the labels of the target domain.

As shown in Figure 1, Figure 1(a) indicates the different covariance of the sample in source and target domains. Figure 1(b) shows that while the target domain remains unchanged, the feature correlations of the source domain are removed. Figure 1(c) draws the correlations from the target domain into the whitened source domain so that their source and target distributions are aligned. Figures 1(a)–(c) are for source and target domain distribution alignment. Figure 1(d) depicts the performance of learning a classifier on the source domain data by intradomain programming using the intradomain structure. In the next section, we will describe each step in turn.

**3.2. Intradomain Alignment.** The IDSDA learns nonparametric transfer the feature through intradomain data alignment, mainly by transforming the statistical features of the data for alignment, thereby reducing the difference between the training domain and testing domain. CORrelation ALIGNment (CORAL) [37] is one of the better

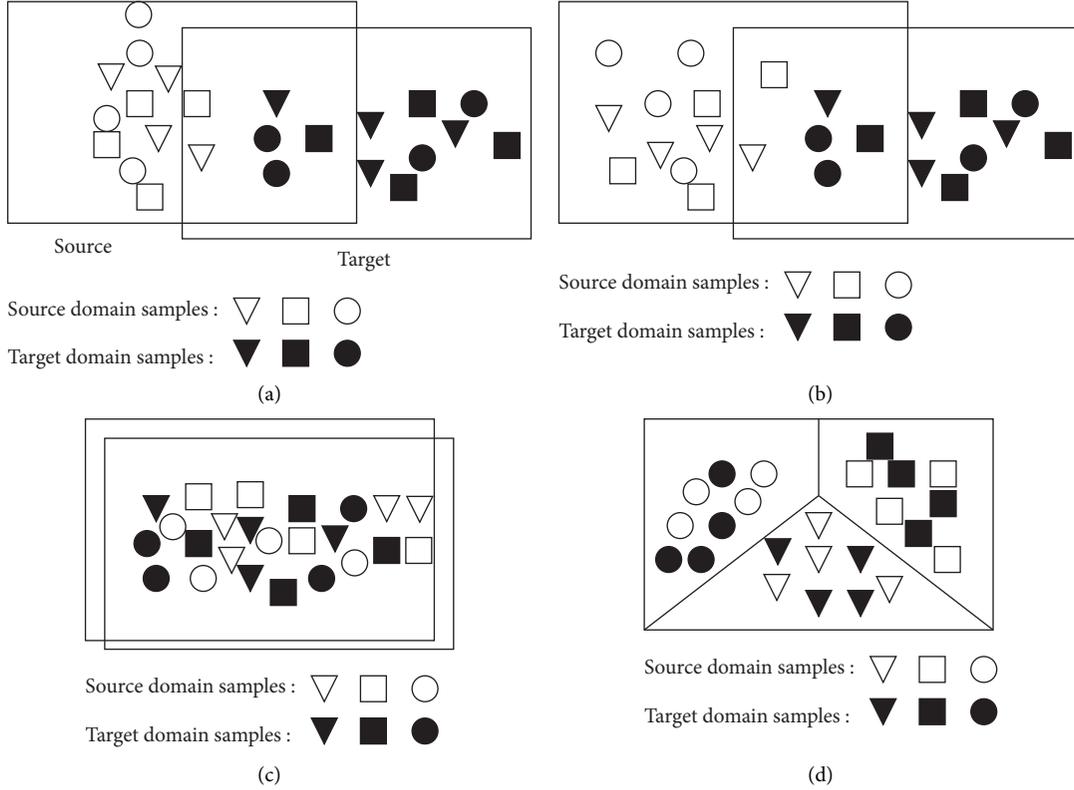


FIGURE 1: IDSDA for cross-database MER. The flowchart of IDSDA is as follows. (a) Original features. (b) Whitening source. (c) Regenerated source. (d) Intradomain programming.

subspace alignment methods, it is computationally efficient, and it does not include other parameters to adjust. Inspired by this method, the intradomain alignment process for the IDSDA method is as follows:

$$Q = X_s \cdot (\text{cov}(X_s) + \text{eye}(\text{size}(X_s, 2)))^{-1/2}, \quad (4)$$

where  $\text{cov}(\cdot)$  is the covariance matrix.  $\text{eye}(\text{size}(X_s, 2))$  and  $\text{eye}(\text{size}(X_t, 2))$  are identity matrices of the same size as  $X_s$  and  $X_t$ . We can regard this step as whitening the source domain; that is, the feature correlation is removed from the source domain.

With the aim of reducing the interdomain differences, the whitened source domain features need to be recolored next [38]. Meanwhile, the target domain features are kept unchanged.

$$\begin{aligned} A^s &= Q \cdot (\text{cov}(X_s) + \text{eye}(\text{size}(X_t, 2)))^{1/2}, \\ A^t &= X_t. \end{aligned} \quad (5)$$

**3.3. Intradomain Programming.** This step in cross-database micro-expression experiments aims to learn a transfer classifier. Before learning the nonparametric transfer classifier, we first need to understand the probability annotation matrix. Table 2 shows a probability annotation matrix  $M$ ,  $M \in \mathbb{R}^{C \times n_t}$ , and the elements of it satisfy the condition  $0 \leq M_{ci} \leq 1$ .  $c$  denotes the class of the label while  $c \in \{1, 2, \dots, C\}$ .  $M_{ci}$  is expressed as the annotation

probability that  $x_i^t$  is of class  $c$ . Similar to the principle of softmax classifier, the class corresponding to the highest probability value for  $x_i^t$  is the class to which it belongs. For example, the highest probability value of 0.5 for  $x_2^t$  implies that it belongs to  $C_3$ .

If  $D_{ci}$  represents the Euclidean distance from  $x_i^t$  to the center of class  $c$  of the source domain, and  $l_c$  is used to represent the center of class  $c$ , then the cost function is as follows:

$$\mathcal{F} = \sum_i^{n_t} \sum_c^C D_{ci} M_{ci}, \quad (6)$$

$I(y_j^s = c)$  can be summarized as the indicator function, the output is 1 when the input is correct, and 0 represents the incorrect input.

$$l_c = \frac{1}{|\Omega_s^{(c)}|} \sum_j^{n_s} x_j^s \cdot I(y_j^s = c), \quad (7)$$

$$D_{ci} = \|x_i^t - l_c\|^2.$$

The fact that  $\mathcal{Y}_s = \mathcal{Y}_t$ , with any label containing a number of samples not less than 1. Under ideal conditions, if  $x_i^t$  does not belong to class  $c$ , then  $M_{ci}$  is 0, and vice versa is 1. So, we can denote it as

$$\sum_i^{n_t} M_{ci} \geq 1, \quad \forall c \in \{1, \dots, C\}. \quad (8)$$

TABLE 2: A probabilistic annotation matrix.

Categories	Sample				
	$x_1^t$	$x_2^t$	$x_3^t$	...	$x_{n_t}^t$
$C_1$	0.2	0.1	0.5	...	...
$C_2$	0.5	0.4	0.3	...	...
$C_3$	0.3	0.5	0.2	...	...

Since  $M_{ci}$  is the probability that  $x_i^t$  is the  $c$  th class label and  $c \in \{1, \dots, C\}$ ,  $M_{ci}$  is bounded by the following condition:

$$\begin{aligned} \min \quad & \mathcal{J} = \sum_i^{n_t} \sum_c^C D_{ci} M_{ci}, \\ \text{s.t.} \quad & \begin{cases} 0 \leq M_{ci} \leq 1, \\ \sum_c^C M_{ci} = 1, \end{cases} \quad \forall i \in \{1, \dots, n_t\}, \sum_i^{n_t} M_{ci} \geq 1, \quad \forall c \in \{1, \dots, C\}. \end{aligned} \quad (10)$$

Linear programming is the solution to the problem of maximizing or minimizing a linear objective function under the constraints of a linear equation or inequality and is often solved very efficiently by an open-source linear programming package called PuLP (<https://pypi.org/project/PuLP/1.1/>). We can use this method to obtain  $M$ .

Ultimately, the labels  $y_i^t$  of the target domain can be obtained by the softmax function.

$$y_i^t = \arg \max_u \frac{M_{ui}}{\sum_c^C M_{ci}} \quad \text{for } u \in \{1, \dots, C\}. \quad (11)$$

## 4. Experiments

**4.1. Micro-Expression Database.** For further confirmation of the reliability of the IDSDA in cross-database facial MER, plenty of CDMER experiments will be conducted in this chapter. We choose two widely used databases, CASMEII and SMIC. The SMIC dataset [39, 40] was created by the University of Oulu, Finland, using a 100 fps high-speed video camera to record the subject who was required to look at videos with large emotional swings while attempting to conceal her/his emotions, and the recorder observed the subjects' expressions without watching the videos. Under this elicitation mechanism, 164 video sequences were obtained from 16 individuals (10 men and 6 women), and the obtained micro-expressions belonged to 3 categories (Positive, Surprised, and Negative). In order to study the micro-expressions more deeply, 71 video sequences of micro-expressions were later recorded under normal vision (VIS) and near-infrared (NIR) environments using cameras with a frame rate of 25 fps. SMIC (VIS, HS, NIR) will be used as an independent dataset in this experiment.

CASMEII [41], which was created by the Institute of Psychology, Chinese Academy of Sciences, uses a similar elicitation mechanism to ensure the reliability of the data. The dataset consists of 247 video sequences of 26 individuals recorded at 200 fps frame rate, and seven micro-expression

$$\sum_c^C M_{ci} = 1, \quad \forall i \in \{1, \dots, n_t\}. \quad (9)$$

We need to minimize the cost function to optimize the model in combination with equation (6), and with the constraints of equations (9) and (8), the final learning goal becomes

categories (Happy, Surprised, Disgusted, Depressed, Sad, Scared, and Other) are included within the dataset. As we have seen, the micro-expression categories in the databases CASMEII and SMIC are quite different. In order to make the two categories consistent, the categories in CASME II were selected and relabeled. Specifically, the Happy sample of micro-expressions is tagged as Positive, and then, the Other category was removed. The labels of the Surprised sample remained unchanged, and finally, the Disgusted and Depressed samples were relabeled as Negative. The details are shown in Table 3.

**4.2. Experimental Setup.** SMIC (VIS, HS, NIR) and CASMEII were set as the four datasets for this cross-database. In other words, when CASMEII is the source (target) domain dataset, one of the datasets in the SMIC database is the target (source) domain dataset. We can obtain 6 sets of cross-database micro-expression experiments, namely, No.1 : C-H, No.2 : C-N, No.3 : C-V, No.4 : H-C, No.5 : N-C, and No.6 : V-C. It should be explained that C represents CASMEII, H corresponds to SMIC (HS), N represents NIR, and V is recorded as VIS in the SMIC database. For the four data sets, LBP-TOP [42] descriptor is selected to operate feature extraction. An  $8 \times 8$  grid was used to divide the micro-expression sequences, on three orthogonal planes, on the premise that the neighborhood point P of the LBP operator is 8, and the neighborhood radius  $R$  is set as 2. Then, all LBP-TOP histograms in each block are connected in series to form facial feature vectors.

With the purpose of proving that the IDSDA is superior, a comparison with other domain adaptive methods that have good performance in cross-database identification has been chosen. These methods include support vector machine (SVM) [43], CORAL [37], geodesic flow kernel (GFK) [44], and DRFS-T [15]. Parameter settings of all the above methods in the experiment are shown as follows:

- (1) For SVM, set  $C = 1$  (linear), and use it as a traditional method. The results of experiments conducted

TABLE 3: Sample composition of the reconstituted CASMEII and SMIC databases.

Category	Database			
	CASMEII	SMIC (HS)	SMIC (NIR)	SMIC (VIS)
Positive	32	51	23	23
Surprise	25	43	20	20
Negative	91	70	28	28
Total	148	164	71	71

directly with SVM will be used as a baseline method for comparison.

- (2) DRFS-T involves two important parameters, i.e.,  $\lambda$  and  $\mu$ . We search for the best values of these two parameters in  $\lambda \in [0.01, 0.1, 1, 10, 100, 1000]$  and  $\mu \in [0.01: 0.01: 0.09, 0.1: 0.1: 1, 2: 10]$ .
- (3) For CORAL, we use the common SVM classifier and set  $C = 1$ . To enrich the types of classifiers in the experiments, 1NN is used in GFK and the optimal dimensionality reduction is chosen in [20, 30].

**4.3. Analysis of Experimental Results.** Table 4 shows the accuracy of the experimental results of the domain adaptive methods covered in this paper. It is clear that the experimental results of the IDSDA method were superior to other methods in most of the experiments, and the results are significantly improved. For example, in No.2 and No.3, the maximum difference between the IDSDA method and other DA methods can reach 23.94% and 26.76%. In addition, an interesting phenomenon was found by observing the experimental results of all methods. Under the conditional setting that the source domain samples were from CASMEII, the experimental results of SMIC (VIS) as the target domain dataset were superior to those of SMIC (NIR) as the target dataset (No.2 and No.3). Meanwhile, No. 6 micro-expression recognition accuracy is higher than No.5. As shown in Section 4.1, SMIC (NIR) data set is shot by a near-infrared camera, and SMIC (VIS) data set is shot under normal vision. We found that the difference in image quality at both source and target domains may cause this phenomenon.

The exact value in No.5 and No.6 is lower than that in No.2 and No.3, with a minimum difference of 17.33%. In other words, the recognition accuracy of CASMEII as a source domain database is higher than that as a target domain database. The analysis may be caused by the imbalance of sample categories. As shown in Table 3, the proportions of the three micro-expression categories in the CASMEII database used in this experiment are greatly different, with negative emotions taking up the majority (91/148), while the difference in the recognition rate between No.1 and No.4 was only 2.5%, significantly less than 17.33%. Carefully observed, it can be found that SMIC (HS) dataset also accounts for a large proportion of negative emotions. We found that they have the same database composition, which further proves that similar data composition of source and target domains is helpful for DA recognition. This further proves that category imbalance

TABLE 4: Experimental results of cross-database MER on CASMEII and SMIC (HS, NIR, and VIS).

No.	Task	SVM	GFK	CORAL	DRFS-T	IDSDA
1	C-H	46.95	40.85	51.83	48.17	57.93
2	C-N	40.85	42.25	50.70	32.39	60.56
3	C-V	45.07	46.48	57.75	45.07	71.83
4	H-C	67.57	43.92	51.35	73.65	55.41
5	N-C	28.38	39.19	39.19	40.54	43.23
6	V-C	52.03	41.22	56.76	58.78	50.00
-	AVG	46.81	42.32	51.26	49.77	56.49

TABLE 5: Parameters and runtime of each of the cross-database micro-expression methods.

Methods	CORAL	GFK	DRFS-T	IDSDA
Parameter	C	$d$	$\lambda, \mu$	None
Time (s)	70.58	115.36	>1400	58.09
AVG (%)	58.06	42.32	49.77	56.49

may be the main factor affecting the recognition effect in CASMEII and SMIC databases.

In addition, we also recorded the average running time (training and testing) of the DA methods involved in the paper in each group of experiments, the average precise values of the six groups of experiments, and the required parameters of each method in Table 5. The results show that the IDSDA method outperforms the other methods in terms of average accuracy (56.49%) while greatly reducing the running time (58.09 s). In summary, this experiment demonstrates the superiority of the IDSDA method in terms of accuracy and efficiency.

## 5. Conclusion and Discussion

In this paper, we use the intradomain structure for CDMER. The intradomain alignment is first performed to learn the nonparametric transfer features, while the classifier is later learned through intradomain planning. It is simple to operate as it does not require model picking and hyperparametric tuning. Extensive experiments are carried out on CASMEII and SMIC databases, and by a comparative analysis, there is a clear indication that the IDSDA is significantly superior to other cutting-edge domain adaptive methods when it comes to the comprehensive performance and efficiency for the CDMER task.

Although the Intradomain Structure Domain Adaptation (IDSDA) method has achieved good results in CDMER, there are still some problems to be investigated. (1) According to the experimental results, database category imbalance is an important factor affecting cross-database MER. Next, we need to focus on how to reduce the impact of category imbalance. (2) The recognition effect is also influenced by the difference of image quality between source and target domain databases when the domain adaptive method is dealing with cross-database micro-expression recognition. We will be able to consider transforming the database used in the experiment to the same image quality in our future work. The proposed approaches in the paper can

combine other methods and tools [45–53] to study the image recognition and identification problems of different plants and can be applied to other studies [54–62] in natural sciences and social sciences.

## Data Availability

The data used to support 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

Y. Zhang and Y. Liu jointly designed the study. Y. Liu collected and analyzed the data. H. Wang reviewed and edited the manuscript.

## Acknowledgments

This work was supported by the Henan Scientific and Technological Research Project under Grant 212102210504.

## References

- [1] P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2003.
- [2] E. A. Haggard and K. S. Isaacs, "Micro-momentary Facial Expressions as Indicators of Ego Mechanisms in Psychotherapy," *Methods Of Research In Psychotherapy*, Springer, Boston, MA, USA, pp. 154–165, 1966.
- [3] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.
- [4] D. Qiao and Y. Zhang, "Big data-enabled analysis of DRGs-Based payment on stroke patients in Jiaozuo, China," *Journal of Healthcare Engineering*, vol. 2020, Article ID 6690019, 9 pages, 2020.
- [5] J. M. Gottman and R. W. Levenson, "A two-factor model for predicting when a couple will divorce: exploratory analyses using 14-year longitudinal data," *Family Process*, vol. 41, no. 1, pp. 83–96, 2002.
- [6] F. Salter, K. Grammer, and A. Rikowski, "Sex differences in negotiating with powerful males," *Human Nature*, vol. 16, no. 3, pp. 306–321, 2015.
- [7] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.
- [8] L. D. Pool and P. Qualter, "Improving emotional intelligence and emotional self-efficacy through a teaching intervention for university students," *Learning and Individual Differences*, vol. 22, no. 3, pp. 306–312, 2012.
- [9] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, WW Norton and Company, New York, NY, USA, Revised Edition, 2009.
- [10] W. S. Chu, F. De La Torre, and J. F. Cohn, "Selective transfer machine for personalized facial action unit detection," *IEEE Conference on Computer Vision and Recognition*, vol. 2013, pp. 3515–3522, 2013.
- [11] W. S. Chu, F. De La Torre, and J. F. Cohn, "Selective transfer machine for personalized facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 529–545, 2017.
- [12] K. Yan, W. Zheng, Z. Cui, and Y. Zong, "Cross-database facial expression recognition via unsupervised domain adaptive dictionary learning," *Neural Information Processing in Proceedings of the International Conference on Neural Information Processing*, vol. 9948, pp. 427–434, Springer, Kyoto, Japan, October 2016.
- [13] W. Zheng and X. Zhou, "Cross-pose color facial expression recognition using transductive transfer linear discriminant analysis," in *Proceeding of the IEEE International Conference on Image Processing*, pp. 1935–1939, Quebec City, QC, Canada, September 2015.
- [14] W. Zheng and X. Zhou, "Cross-domain color facial expression recognition using transductive transfer subspace learning," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 21–37, 2016.
- [15] Y. Zong, W. Zheng, X. Huang, J. Shi, Z. Cui, and G. Zhao, "Domain regeneration for cross-database micro-expression recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2484–2498, 2018.
- [16] L. Li, X. Zhou, Y. Zong et al., "Unsupervised cross-database micro-expression recognition using target-adapted least-squares regression," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 7, pp. 1417–1421, 2019.
- [17] F. Ding, X. Zhang, and L. Xu, "The innovation algorithms for multivariable state-space models," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 11, pp. 1601–1618, 2019.
- [18] L. Xu, W. Xiong, A. Alsaedi, and T. Hayat, "Hierarchical parameter estimation for the frequency response based on the dynamical window data," *International Journal of Control, Automation and Systems*, vol. 16, no. 4, pp. 1756–1764, 2018.
- [19] L. Xu and G. Song, "A recursive parameter estimation algorithm for modeling signals with multi-frequencies," *Circuits, Systems, and Signal Processing*, vol. 39, no. 8, pp. 4198–4224, 2020.
- [20] L. Xu, F. Ding, X. Lu, L. Wan, and J. Sheng, "Hierarchical multi-innovation generalised extended stochastic gradient methods for multivariable equation-error autoregressive moving average systems," *IET Control Theory & Applications*, vol. 14, no. 10, pp. 1276–1286, 2020.
- [21] L. Xu, F. Ding, L. Wan, and J. Sheng, "Separable multi-innovation stochastic gradient estimation algorithm for the nonlinear dynamic responses of systems," *International Journal of Adaptive Control and Signal Processing*, vol. 34, no. 7, pp. 937–954, 2020.
- [22] Y. Zhou and F. Ding, "Modeling nonlinear processes using the radial basis function-based state-dependent autoregressive models," *IEEE Signal Processing Letters*, vol. 27, pp. 1600–1604, 2020.
- [23] J. Pan, H. Ma, X. Zhang et al., "Recursive coupled projection algorithms for multivariable output-error-like systems with coloured noises," *IET Signal Processing*, vol. 14, no. 7, pp. 455–466, 2020.
- [24] H. Ma, J. Pan, F. Ding, L. Xu, and W. Ding, "Partially-coupled least squares based iterative parameter estimation for multivariable output-error-like autoregressive moving average systems," *IET Control Theory and Applications*, vol. 13, no. 18, pp. 3040–3051, 2019.
- [25] H. Ma, X. Zhang, Q. Liu et al., "Partially-coupled gradient-based iterative algorithms for multivariable output-error-like

- systems with autoregressive moving average noises,” *IET Control Theory & Applications*, vol. 14, no. 17, pp. 2613–2627, 2020.
- [26] F. Ding, L. Lv, J. Pan, X. Wan, and X.-B. Jin, “Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data,” *International Journal of Control, Automation and Systems*, vol. 18, no. 4, pp. 886–896, 2020.
- [27] F. Ding, L. Xu, D. Meng, X.-B. Jin, A. Alsaedi, and T. Hayat, “Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model,” *Journal of Computational and Applied Mathematics*, vol. 369, Article ID 112575, 2020.
- [28] L. Wan, F. Ding, X. Liu, and C. Chen, “A new iterative least squares parameter estimation approach for equation-error autoregressive systems,” *International Journal of Control, Automation and Systems*, vol. 18, no. 3, pp. 780–790, 2020.
- [29] L. Wan and F. Ding, “Decomposition- and gradient-based iterative identification algorithms for multivariable systems using the multi-innovation theory,” *Circuits, Systems, and Signal Processing*, vol. 38, no. 7, pp. 2971–2991, 2019.
- [30] M. Li and X. Liu, “The least squares based iterative algorithms for parameter estimation of a bilinear system with autoregressive noise using the data filtering technique,” *Signal Processing*, vol. 147, pp. 23–34, 2018.
- [31] M. Li and X. Liu, “Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique,” *International Journal of Control, Automation and Systems*, vol. 18, no. 6, pp. 1581–1592, 2020.
- [32] M. Li, X. Liu, and F. Ding, “The filtering-based maximum likelihood iterative estimation algorithms for a special class of nonlinear systems with autoregressive moving average noise using the hierarchical identification principle,” *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1189–1211, 2019.
- [33] M. Jing, J. Zhao, J. Li, L. Zhu, Y. Yang, and H. T. Shen, “Adaptive component embedding for domain adaptation,” *IEEE Transactions on Cybernetics*, vol. 99, pp. 1–14, 2020.
- [34] J. Li, E. Chen, Z. Ding, L. Zhu, K. Lu, and H. T. Shen, “Maximum density divergence for domain adaptation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, p. 1, 2020.
- [35] J. Li, K. Lu, Z. Huang, L. Zhu, and H. T. Shen, “Heterogeneous domain adaptation through progressive alignment,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1381–1391, May 2019.
- [36] J. Li, M. Jing, K. Lu, L. Zhu, and H. T. Shen, “Locality preserving Joint transfer for domain adaptation,” *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6103–6115, 2019.
- [37] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” *AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, pp. 2058–2065, 2016.
- [38] J. Wang, Y. Chen, H. Yu et al., “Easy Transfer Learning by Exploiting Intra-domain Structures,” in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 1210–1215, Shanghai, China, July 2019.
- [39] X. Li, T. Pfister, X. Huang et al., “A spontaneous micro-expression database: inducement, collection and baseline,” in *Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–6, Shanghai, China, April 2013.
- [40] T. Pfister, X. Li, G. Zhao et al., “Recognising Spontaneous Facial Micro-expressions,” in *Proceedings of the International Conference on Computer Vision*, pp. 1449–1456, Barcelona, Spain, November 2011.
- [41] W. J. Yan, Q. Wu, Y. J. Liu et al., “CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces,” in *Proceedings of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–7, Shanghai, China, April 2013.
- [42] G. Zhao and M. Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.
- [43] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [44] B. Gong, Y. Shi, F. Sha et al., “Geodesic Flow Kernel for Unsupervised Domain Adaptation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2066–2073, Providence, RI, USA, June 2012.
- [45] X. Zhang and F. Ding, “Adaptive parameter estimation for a general dynamical system with unknown states,” *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1351–1372, 2020.
- [46] X. Zhang, F. Ding, and L. Xu, “Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems,” *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1373–1393, 2020.
- [47] X. Zhang and F. Ding, “Recursive parameter estimation and its convergence for bilinear systems,” *IET Control Theory and Applications*, vol. 14, no. 5, pp. 677–688, 2020.
- [48] X. Zhang and F. Ding, “Hierarchical parameter and state estimation for bilinear systems,” *International Journal of Systems Science*, vol. 51, no. 2, pp. 275–290, 2020.
- [49] M. Chen, F. Ding, R. Lin, T. Y. Ng, Y. Zhang, and W. Wei, “Maximum likelihood least squares-based iterative methods for output-error bilinear-parameter models with colored noises,” *International Journal of Robust and Nonlinear Control*, vol. 30, no. 15, pp. 6262–6280, 2020.
- [50] T. Cui, F. Ding, X.-B. Jin, A. Alsaedi, and T. Hayat, “Joint multi-innovation recursive extended least squares parameter and state estimation for a class of state-space systems,” *International Journal of Control, Automation and Systems*, vol. 18, no. 6, pp. 1412–1424, 2020.
- [51] J. Y. Ni and Y. L. Zhang, “Parameter estimation algorithms of linear systems with time-delays based on the frequency responses and harmonic balances under the multi-frequency sinusoidal signal excitation,” *Signal Processing*, vol. 181, Article ID 107904, 2021.
- [52] S. Y. Liu, Y. L. Zhang, L. Xu et al., “Extended gradient-based iterative algorithm for bilinear state-space systems with moving average noises by using the filtering technique,” *International Journal of Control Automation and Systems*, vol. 19, 2021.
- [53] Y. Ji, C. Zhang, Z. Kang, and T. Yu, “Parameter estimation for block-oriented nonlinear systems using the key term separation,” *International Journal of Robust and Nonlinear Control*, vol. 30, no. 9, pp. 3727–3752, 2020.
- [54] Y. Ji, X. Jiang, and L. Wan, “Hierarchical least squares parameter estimation algorithm for two-input Hammerstein finite impulse response systems,” *Journal of the Franklin Institute*, vol. 357, no. 8, pp. 5019–5032, 2020.
- [55] Y. Ji and Z. Kang, “Three-stage forgetting factor stochastic gradient parameter estimation methods for a class of nonlinear systems,” *International Journal of Robust and Nonlinear Control*, vol. 31, no. 3, pp. 871–987, 2021.

- [56] Y. Fan and X. Liu, "Two-stage auxiliary model gradient-based iterative algorithm for the input nonlinear controlled autoregressive system with variable-gain nonlinearity," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 14, pp. 5492–5509, 2020.
- [57] F. Ding, L. Xu, and Q. Zhu, "Performance analysis of the generalised projection identification for time-varying systems," *IET Control Theory & Applications*, vol. 10, no. 18, pp. 2506–2514, 2016.
- [58] F. Ding, X. Wang, L. Mao, and L. Xu, "Joint state and multi-innovation parameter estimation for time-delay linear systems and its convergence based on the Kalman filtering," *Digital Signal Processing*, vol. 62, pp. 211–223, 2017.
- [59] L. Xu, "The parameter estimation algorithms based on the dynamical response measurement data," *Advances in Mechanical Engineering*, vol. 9, no. 11, 2017.
- [60] F. Ding, F. Wang, L. Xu, T. Hayat, and A. Alsaedi, "Parameter estimation for pseudo-linear systems using the auxiliary model and the decomposition technique," *IET Control Theory and Applications*, vol. 11, no. 3, pp. 390–400, 2017.
- [61] X. Zhang, F. Ding, L. Xu, and E. Yang, "Highly computationally efficient state filter based on the delta operator," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 6, pp. 875–889, 2019.
- [62] X. Zhang, F. Ding, and E. Yang, "State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1157–1173, 2019.

## Research Article

# Predicting Mental Health Problems with Automatic Identification of Metaphors

Nan Shi,<sup>1</sup> Dongyu Zhang ,<sup>1</sup> Lulu Li,<sup>1</sup> and Shengjun Xu<sup>2</sup>

<sup>1</sup>School of Software, Dalian University of Technology, Dalian 116620, China

<sup>2</sup>Psychological Health Education and Counseling Center, Dalian University of Technology, Dalian 116620, China

Correspondence should be addressed to Dongyu Zhang; zhangdongyu@dlut.edu.cn

Received 19 February 2021; Revised 18 March 2021; Accepted 10 April 2021; Published 30 April 2021

Academic Editor: Han Wang

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

Mental health problems are prevalent and an important issue in medicine. However, clinical diagnosis of mental health problems is costly, time-consuming, and often significantly delayed, which highlights the need for novel methods to identify them. Previous psycholinguistic and psychiatry research has suggested that the use of metaphors in texts is linked to the mental health status of the authors. In this paper, we propose a method for automatically detecting metaphors in texts to predict various mental health problems, specifically anxiety, depression, inferiority, sensitivity, social phobias, and obsession. We perform experiments on a composition dataset collected from second-language students and on the eRisk2017 dataset collected from Social Media. The experimental results show that our approach can help predict mental health problems in authors of written texts, and our algorithm performs better than other state-of-the-art methods. In addition, we report that the use of metaphors even in nonnative languages can be indicative of various mental health problems.

## 1. Introduction

Mental health problems have become increasingly serious. They not only endanger people's physical and mental health, but also affect the development of the country and society. The WHO survey (<https://www.who.int/health-topics/mental-health>) shows that about 13% of people worldwide suffer from mental disorder, which cost the global economy one trillion dollars each year. Depression is one of the main causes of disability. Suicide is the second leading cause of death among children aged 15–29. About 20% of children and adolescents in the world suffer from mental illness, and the highly educated population also suffer from psychological distress which affects their academic performance [1–3]. However, clinical diagnosis of mental health problems is costly, time-consuming, and often significantly delayed, which highlights the need for novel methods to identify these conditions.

Metaphorical expressions are frequently used in human language [4–7]. They involve both linguistic expression and cognitive processes [8] and are an implicit way to convey emotions [9–11]. Human emotions and mental state, which

are important for mental health, are frequently communicated and expressed through metaphors. This suggests that the use of metaphorical expressions in texts may indicate mental and cognitive status and so help in mental health screening.

Psycholinguistic and psychiatry studies have indicated that the use of metaphors in texts is linked to the mental health illness of their authors [12–16]. For example, patients with schizophrenia may metaphorically use the phrase “time vessels” to refer to watches and “hand shoes” to refer to gloves. In other words, the use of metaphor in individuals with mental illness may differ from those without, which could offer new opportunities to identify mental illness using metaphors as a diagnostic indicator. Although it is not clear what causes these deviations in metaphor production, neuroscience research offers some clues. Scholars note that some mental illnesses such as schizophrenia relate to dysfunction of the amygdala, which processes and regulates emotion [17]. Other research suggests that metaphorical texts are associated more with activation of the amygdala than with areas relating to literal speech [18].

With the development of artificial intelligence and various data processing technologies [19–25], the efficiency of modern medical diagnosis is constantly improving. As an important part of artificial intelligence, natural language processing is widely applied in mental health related issues [26–28]. Shatte et al. [29] reviewed the application of machine learning in mental health: four main application areas, including detection and diagnosis [30, 31]; prognosis, treatment, and support; public health; research and clinical management. The most common mental health conditions addressed include depression, schizophrenia, and Alzheimer’s disease. Prior work has shown the feasibility of using NLP techniques with various features extracted from text messages such as linguistic, demographic, and behavioral features to predict mental illness such as depression [32], suicidality [33], and posttraumatic stress disorder [34]. However, few studies have involved the application of metaphor, a deep semantic feature, as a means of detecting and predicting mental health problems. Along with the rapid explosion of social media applications such as Twitter and Facebook, there seems to be a significant increase in metaphorical texts on a variety of topics, including products, services, public events, and tidbits of people’s lives. It seems to be an important and promising challenge to leverage metaphor features for supporting the identification and prediction of mental health problems.

In this paper, we propose the use of automatically detected metaphors in texts to predict various mental health problems including anxiety, depression, inferiority, sensitivity, social phobias, and obsession. We named our method Metaphor-Sentiment Model (MSM) and we performed experiments on a compositional dataset we created from second-language student essays and the eRisk2017 dataset collected from Social Media. Our contributions are as follows.

- (i) We propose a novel approach to identify several mental health problems by using linguistic metaphors in texts as features. To the best of our knowledge, we are the first to leverage metaphor features for supporting the identification and prediction of mental health problems.
- (ii) The experimental results show that our proposed approach can help predict the mental health of authors of written texts and our algorithm gives fairly good performance, compared to state-of-the-art methods.
- (iii) The work shows how semantic content, specifically usage of metaphors in writings produced by individuals, can help in detection of six mental health problems. This seems to be a new result where usage of metaphors even in nonnative languages can be used as indicative of various mental health problems.
- (iv) We contribute to a novel, scarce, and valuable dataset, which will be released publicly, consisting of second-language speakers’ essays and data on authors’ mental health problems obtained from a psychological survey.
- (v) Due to the scarcity of relevant work, exploring features that influence mental health using computational approaches can potentially help with early detection and treatment of mental health and related problems.

## 2. Related Work

*2.1. Mental Health in NLP.* NLP techniques have been applied to speculate on people’s mental health status, based on written texts, such as those on Facebook, Twitter, etc., and they can be used to obtain information on the user’s psychological state directly and efficiently [35]. In recent years, scholars have explored many different features of various datasets to explore the mental health status that lies behind a text. Nguyen et al. [36] used data from the foreign Live Journal Post website to collect 38 k posts from the mental illness community and 230 k posts from mentally healthy communities for mental illness prediction. They tried various approaches, including linguistic inquiry and word count (LIWC), which obtain features of language, social, effective, cognitive, perceptual, biological, relativistic, personal attention, and oral, emotional feature (also based on LIWC) and latest dirichlet allocation (LDA) topic models, ultimately achieving 93% accuracy. Franco-Penya and Sanchez [37] built a tree structure based on the n-grams feature and combined other features and support vector machine (SVM) learning methods to design classifiers to detect mental health status in CLPsych2016 [38]. Cohan et al. [39] comprehensively considered lexical features, contextual features, textual data features, and textual topical features on the same dataset, using SVM classifiers to complete detection tasks. Ramandrisoa et al. [40] tried a variety of lexical features in another evaluation task on the CLEF 2018 eRisk database [41], including bag of word models, specific category words, and special word combinations, and they converted text into vectors for classification. Weerasinghe et al. [42] investigated language patterns that differentiate individuals with mental illnesses from a control group, including bag-of-words, word clusters, part of speech n-gram features, and topic models to understand the machine learning model.

In addition to the use of text and other user characteristics, the rise of deep learning has provided new ways to detect mental illness through text. Benton et al. [43] modeled multiple scenarios to predict different suicide risk levels and built a multitasking learning framework (MTL) to meet the needs of different tasks. Trotszek et al. [44] first converted text into vectors and then completed classification task through a convolutional neural network to predict the mental health status of the user. Sekulic and Strube [45] applied a hierarchical attention network and analyzed phrases relevant with social media users’ mental status by inspecting the model’s word-level attention weights. Multimodal thinking is also applied in mental health research [46, 47]. They used a multimodal approach that consists of jointly analyzing text and visual and audio data and their relation to mental health more than text analysis.

**2.2. Datasets.** As discussed above, metaphorical expressions are associated with mental and cognitive status. Since a metaphor involves cognitive processes, it may be feasible to screen and monitor mental and affective status no matter the degree of fluency in the language. We thus assume that metaphor is an important textual feature for mental health detection among language users, including both native and second-language speakers. We collected data from two different sources to verify our assumptions and increase the reliability of our experiments in this study of the relationship between metaphor use and mental health status.

**2.3. Student Composition and Mental Health.** We collected English composition data from English-proficient Chinese college students who speak English as a second language. We also collected mental health data from these students using a psychological survey. First, we used online and offline campus advertisements to recruit 164 college freshman participants who passed the national college students' English level 4 test in China, which means they are native Chinese and fluent in English writing. Prior to participation, all participants provided a consent form indicating their willingness to take part in the study. Participants provided their personal information via a questionnaire and then wrote a composition with 500 English words or more within a two-hour period. The composition had two parts: described their previous life experience and then presented their future plans, including their ideal future lives, thoughts on life, targets for their future lives, and plans to overcome barriers. The content gave us a deep understanding of their psychological states [48], which is essential for the detection of mental health problems.

After writing their composition, students were required to complete a mental health questionnaire that assessed two levels of mental health problems. The first level involved is serious mental health problems, mainly serious psychoses such as hallucinations, suicidal behavior, and suicidal inclination. In our survey, only a few students had first-level problems. The second level involved common mental problems, such as anxiety, depression, inferiority, sensitivity, and social phobias. Mental problems were assessed on the basis of the standard score for screening indexes. Specifically, participants were assessed with mental health problems when their scores on certain indexes exceeded typical results. We excluded data from 8 students because we could not match their mental problems with fuzzy indexes from their mental health data. Effective mental health data for the remaining 156 students is presented in Table 1. Meanwhile, we extracted data from students without mental health problems to use as controls for analyzing differences in metaphor use with sentiment features in texts.

The process of data collection lasted four months and resulted in a total of 156 compositions with 130,044 words from 156 students (aged 18–23 yrs, mean = 19.06 yrs, SD = 0.19, males = 86, and females = 70), together with mental health data obtained from the psychological questionnaire. These data were kept secure and stored with no identifying factors, i.e., consent forms and questionnaires.

TABLE 1: Mental health data of students.

Problem	No. of students	Problem	No. of students
Anxiety	36	Sensitivity	49
Depression	36	Social phobia	44
Inferiority	29	Obsession	38
One problem	28	Two problems	21
Three problems	21	Four problems	10
Five problems	7	Six problems	7

$N$  problems mean the students with  $n$  mental problems at the same time.

**2.4. eRisk2017 Data.** The eRisk2017 task on early risk detection of depression [49] provides a dataset containing posting contents and comments from Reddit. The task identified 135 Reddit users with depression and 752 Reddit users without depression through their posts and comments. The word quantity for each Reddit user varies from 10 to 2,000. The dataset for each Reddit user contains individual identification, writing data, text title, writing type, and writing contents. Paper [50] details the construction of the eRisk data. They first selected Reddit from multiple social media and collected the post of depression diagnosis through specific search (such as I was diagnosed with depression). Posts are manually evaluated to identify users who are really diagnosed with depression. They collected patients' text records published on Reddit over a period of time. We combined the contents for each Reddit user in chronological order for the present study.

### 3. Methodology

Our work flow is shown in Figure 1. Metaphor is linked to the mental health problems as described above. We extracted metaphors from texts and designed metaphor feature sets to predict various mental health problems. Our method also considered sentiment features in the sample texts as this feature has been widely used in mental health research [14, 44, 51, 52]. We applied metaphor and sentiment features in our Metaphor-Sentiment Model (MSM) to predict mental health problem. The feature extraction algorithm is briefly summarized in algorithm 1, and more details will be introduced below.

**3.1. Metaphor Feature Extraction.** For metaphor-based features, we considered the following (Algorithm 1, Step 1):

- (i) The percentage of tokens tagged as metaphor by the automatic metaphor identification method
- (ii) The probability of a sentence containing metaphor

We also considered the sentiment of metaphor expressed in a sentence that is consistent with the sentence sentiment. First, SentiStrength (<http://sentistrength.wlv.ac.uk/>) was used to analyze the overall sentiment of a sentence. SentiStrength analysis yields two scores for sentiment strength: negative (scores  $-1$  to  $-5$ ) and positive (scores  $1$  to  $5$ ). The sum of the two values is the overall sentiment score for the sentence. A sentiment score of  $0$  is defined as neutral. Next,

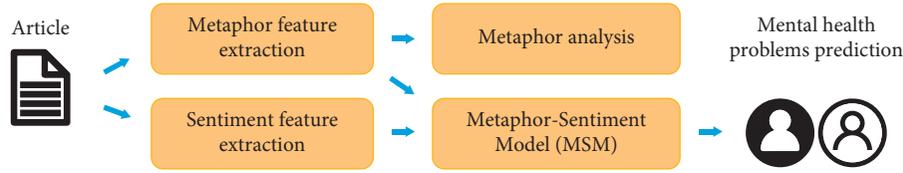


FIGURE 1: The figure of our work flow.

we determined sentiment of metaphor using three specific feature values (Algorithm 1, Step 3):

- (i) The number of metaphors with positive sentiment (positive sentiment score)
- (ii) The number of metaphors with negative sentiment (negative sentiment score)
- (iii) The average sentiment score for all metaphors

In our method, metaphors were identified automatically using a technique that has shown the best performance for token-level metaphor identification tasks to date [53]. The automatic metaphor identification system contains four steps: (1) it trains word embeddings on a Wikipedia dump based on Continuous Bag of Words (CBOW) and Skip-Gram models to obtain input and output vectors for every word; (2) it selects detected words to assess metaphoricity and separates the detected words from a given sentence; (3) it extracts all possible synonyms and direct hypernyms, including their inflections, of the detected word from WordNet, and it adds them to the candidate word set  $w$ , which contains all possible senses of the detected word; and (4) it selects the best fit word  $w^*$ , which represents the actual sense of the detected word in the a given sentence, from the candidate word set  $w$ , using the following formula:

$$w^* = \arg \max_k \text{SIM}(v_k, v_{\text{context}}), \quad (1)$$

where  $k \in w$ ,  $v_k$  is the input vector of the CBOW or Skip-Gram entry for a candidate word  $k$ , and  $v_{\text{context}}$  means the average of all input vectors for context words. The best fit word has the highest cosine similarity with the context words. Finally it computes the similarity value for the detected word and the best fit word using output vectors to measure the difference of sense between the detected word and the context. The detected word is labeled as metaphorical when the similarity value is less than the given threshold. In practical applications we detected every content word in the sentence. The detailed process is presented in Algorithm 2.

We trained and tested the identification algorithm on a metaphor dataset developed by Mohammad [10] that contains 210 metaphorical sentences whose detected words are annotated manually with at least 70% agreement. We selected the same number of literal sentences from thousands of literal sentences in the dataset. The best metaphor identification performance had a precision of 0.635, recall of 0.821, and F1 value of 0.716 with a threshold of 0.5, which matches the identification performance reported by Mao [53].

For evaluating the performance of the metaphor identification method with our dataset, we randomly selected ten compositions from each of the seven groups that correspond to six mental health problems and healthy control. In total, seventy compositions were analyzed. The metaphor identification performance using the student dataset had a precision of 0.632, recall of 0.935, and F1 value of 0.754.

Figure 2 shows examples of metaphors detected by the automatic metaphor identification method from the student composition dataset (a-c) and eRisk2017 dataset (d-f). The sentences match two words from different domains: for example, a source word tagged as metaphorical, such as *broken*, and a target word such as *dream*. However, this token-level metaphor identification algorithm produces some errors since it identifies a metaphor based on local information around the detected word and cannot effectively recognize fixed collocation. For example, in the phrase *I ultimately got up on my own*, the algorithm mistakenly tags the word *own* in *on my own* as metaphorical.

**3.2. Sentiment Feature Extraction.** The sentiment feature set included the average value of the five dimensions of all words; the proportion of positive sentences, negative sentences, and neutral sentences; the average emotional score of the sentences, and the emotional fluctuation value of each article, yielding ten specific features in total.

We used SentiStrength to obtain sentiment scores for sentences in the texts as above, in order to calculate the proportion of positive sentence, negative sentence, and neutral sentence; the average sentiment score of the sentences of article; the sentiment fluctuation score of each article (Algorithm 1, Step 2).

The average score of the sentences in each article was calculated to determine the emotional value of the article using the following formula:

$$E = \frac{\sum_{i=1}^n S_i}{n}, \quad (2)$$

where  $E$  represents the average sentiment value of the text,  $S_i$  represents the emotional score of the  $i$  th sentence, and  $n$  is the number of the sentences in the text. And the fluctuation score is obtained by subtracting the emotional scores of two consecutive sentences in the article and taking the absolute value. We used the average as its sentiment fluctuation value. It is determined by the following formula:

$$F = \frac{\sum_{i=2}^n |S_i - S_{i-1}|}{(n-1)}, \quad (3)$$

**Input:** The target text.

**Output:** Metaphor-Sentiment feature set.

- (1) Identifying the metaphoricity of each word in the text, count the frequency, and generate metaphorical statistical features
- (2) Using Sentistrength to obtain the score of positive and negative emotions, and generate the statistical characteristics and sentiment actuation value on sentence level
- (3) Determining the metaphorical words in the sentence by the sentiment information of the sentence to obtain the sentiment characteristics of the metaphorical words
- (4) Using SenticNet to get the word-level emotional scores of five dimensions, and calculating the average value to get the sentiment features of the text
- (5) Integrating the above characteristics, **return** Metaphor-Sentiment feature set

ALGORITHM 1: Framework of feature set generation for MSM.

**Input:** sentence; A dictionary that returns the corresponding word vector of key, *Word2vec*; A dictionary that returns the related word set of key, *WordNet*.

**Output:** A list of the corresponding metaphoricity labels of words in sentence, *labels*.

- (1) **function** TokenMetaphorIdentify (*detected word*; *sentence*)
- (2) *context* = []; *w\** = 0
- (3) *w* = *WordNet* [*detected word*]
- (4) **for each** *word* ∈ *sentence* and *word* ≠ *detected\_word* **do**
- (5) *context* = *context* ∪ *word*
- (6) **end for**
- (7) *v<sub>context</sub>* = *average* (*Word2vec* [*context*])
- (8) **for each** *k* ∈ *w* **do**
- (9) **if** *cosine*(*Word2vec*[*k*], *v<sub>context</sub>*) > *max\_cosine* **then**
- (10) *max\_cosine* = *cosine*(*Word2vec*[*k*], *v<sub>context</sub>*)
- (11) *w\** = *k*
- (12) **end if**
- (13) **end for**
- (14) *metaphor\_value* = *cosine* (*Word2vec* [*detected\_word*], *Word2vec* [*w\**])
- (15) **if** *metaphor\_value* < *threshold* **then**
- (16) **return** True
- (17) **else**
- (18) **return** False
- (19) **end if**
- (20) **end function**
- (21)
- (22) **function** MetaphorIdentify (*sentence*)
- (23) *labels* = []
- (24) **for each** *word* ∈ *sentence* **do**
- (25) *label* = TokenMetaphorIdentify (*word*; *sentence*)
- (26) *labels* = *labels* ∪ *label*
- (27) **end for**
- (28) **return** *labels*
- (29) **end function**

ALGORITHM 2: Metaphor identification algorithm.

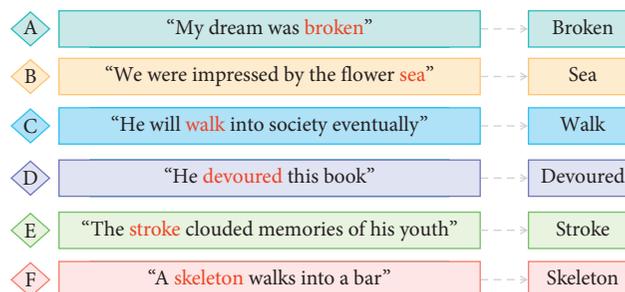


FIGURE 2: Metaphor examples identified by automatic metaphor identification method.

where  $F$  represents the emotional fluctuation value of the text.

Scores for the five dimensions (values of pleasantness, attention, sensitivity, aptitude, and polarity) were obtained using SenticNet (<http://www.sentic.net>) (Algorithm 1, Step 4). The averages of the five dimensions of all words were taken as an indicator of the article's emotions. Averages were calculated as this example for pleasantness values:

$$P = \frac{\sum_{i=1}^n W_i}{n}, \quad (4)$$

where  $P$  represents the average for pleasantness values and  $W_i$  represents the pleasantness value of the  $i$ th word. Averages for attention values, sensitivity values, aptitude values, and polarity values were computed similarly.

#### 4. Metaphor Analysis

We analyzed metaphor use for six mental problems and health control based on automatic identified result, including examples of identified metaphors and statistical analysis.

Table 2 shows examples of the most frequently used metaphors for each of seven mental health groups. In order to demonstrate characteristics of each group, we excluded the metaphorical words, which occurred with the most frequency in all mental health groups, such as *pay*, *top*, and *limit*. The same metaphor word was often used in a different way by those in the mentally healthy group compared to those in mental health problem groups, as illustrated in the following examples:

*Ex1. Teachers always try their best to meet the requirements of students.*

*Ex2. We always meet various difficulties on the way to study.*

The sentence in the first example was taken from a composition by a student in the healthy control group and expresses a positive sentiment, while the second example was taken from a composition by a student in depression group and expresses a negative sentiment.

We study the emotion of text and effect of metaphor in Student Composition data. The statistical information is shown in Table 3.

Avg. emotion denotes average score of emotion of all text and meta emotion is average emotional score of sentence with metaphor. People in sensitivity group have highest emotional score and that of obsession group is lowest. Meta emotion overall is 0.05 lower than avg. emotion, which shows that, in Student Composition Dataset, students are more likely to express negative emotions and describe sad things through metaphor, for example, sentences A and C in Figure 2.

*Ex3. My dream was broken*

*Ex4. He will walk into society eventually*

The former expresses the lost mood of broken dream, and the latter is used to show the helpless mood of growing up and entering the society. Both of them apply metaphor to express negative emotions.

To better understand the characteristics of metaphor use for each mental problem, we labeled students as *sick* or *not sick* for every particular mental problem and analyzed metaphor features between the two groups. The histograms in Figure 3 show the situation of different metaphor features for each mental health problem. We found that the probability of a sentence containing metaphor was higher among students with inferiority or social phobia than students without these mental problems ( $t = 1.775$ ,  $p < 0.1$ ;  $t = 1.695$ ,  $p < 0.1$ ). Students with social phobia were more inclined to use metaphors with negative sentiments than students without social phobia ( $t = 1.978$ ,  $p < 0.05$ ). Additionally, students with obsession had significantly lower scores for average sentiment value of metaphor than students without obsession ( $t = -2.060$ ,  $p < 0.05$ ). The most distinguishing index of compositions by students with mental health problems was the probability of sentence with metaphor. Students with mental health problems had higher eigenvalues for this variable than those in the healthy group.

#### 5. Experiments

We compared the predictive performance of MSM and baseline on the eRisk2017 dataset [49] and on the second-language speaker essays dataset, and we evaluated the metaphor feature with common text features used in baseline. Each of the six mental health problems in the second-language speaker dataset was subjected to a separate bicategorization task. We planned to verify the effectiveness of metaphorical features in the detection of various mental health problems, and we used the same Metaphor-Sentiment feature set in each mental health problem prediction task. Different model parameters will be obtained for different mental problems to deal with metaphorical features.

We applied Synthetic Minority Oversampling Technique (SMOTE) to alleviate the imbalance between positive and negative samples on Student Composition dataset. SMOTE algorithm analyzes samples of minority and produces new samples to the dataset. The specific process: (1) randomly select a sample  $x$  from minority and calculate the Euclidean distance between it and other samples in this category; (2) randomly select a sample  $x_n$  from the  $k$  nearest neighbors of  $x$  calculated in the previous step; (3) according to the following formula, a new sample is constructed and added to the minority sample set; (4) repeat the above steps until the appropriate sample size is obtained.

$$x_{\text{new}} = x + \text{rand}(0, 1) * |x - x_n|. \quad (5)$$

*5.1. Baseline.* The prediction method proposed by [44] was chosen as a baseline since it showed the best performance in eRisk2017 and eRisk2018. They applied two methods to the eRisk2017 dataset to detect people suffering from depression. One method involved logistic regression using features extracted by four word frequency statistics tools—LIWC (<http://liwc.wpengin.com/>), NRC Emotion Lexicon (<http://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>), Opinion Lexicon (<http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>), and VADER Sentiment

TABLE 2: The examples of frequent metaphors in each state of mind.

Mental problem	Frequent metaphor	Example sentence
Anxiousness	Hit, present, join	The poor of property can't hit me, but a boring life can
Depression	Chase, clean, tough	Maybe there will be many difficulties in the way I chase my dream
Inferiority	Support, independent	All these support his spirit of "learning insatiably"
Sensitivity	Defeat, move, create	I know in this process some trouble will defeat me
Social phobia	Raise, affect, stop	It is really a burden for a poor family to raise a child
Obsession	Enter, guide, control	When you enter the society, you probably have problem in finding a job
Healthy control	Develop, pass, lead	I want to develop a wonderful game

TABLE 3: The statistical information of emotion in various mental states.

Mental state	Anxiousness	Depression	Inferiority	Sensitivity	Social Phobia	Obsession	Healthy control	Total
Avg. emotion	0.118	0.129	0.110	0.159	0.083	0.038	0.129	0.129
Meta emotion	0.047	0.100	0.066	0.118	0.033	0.036	0.079	0.079

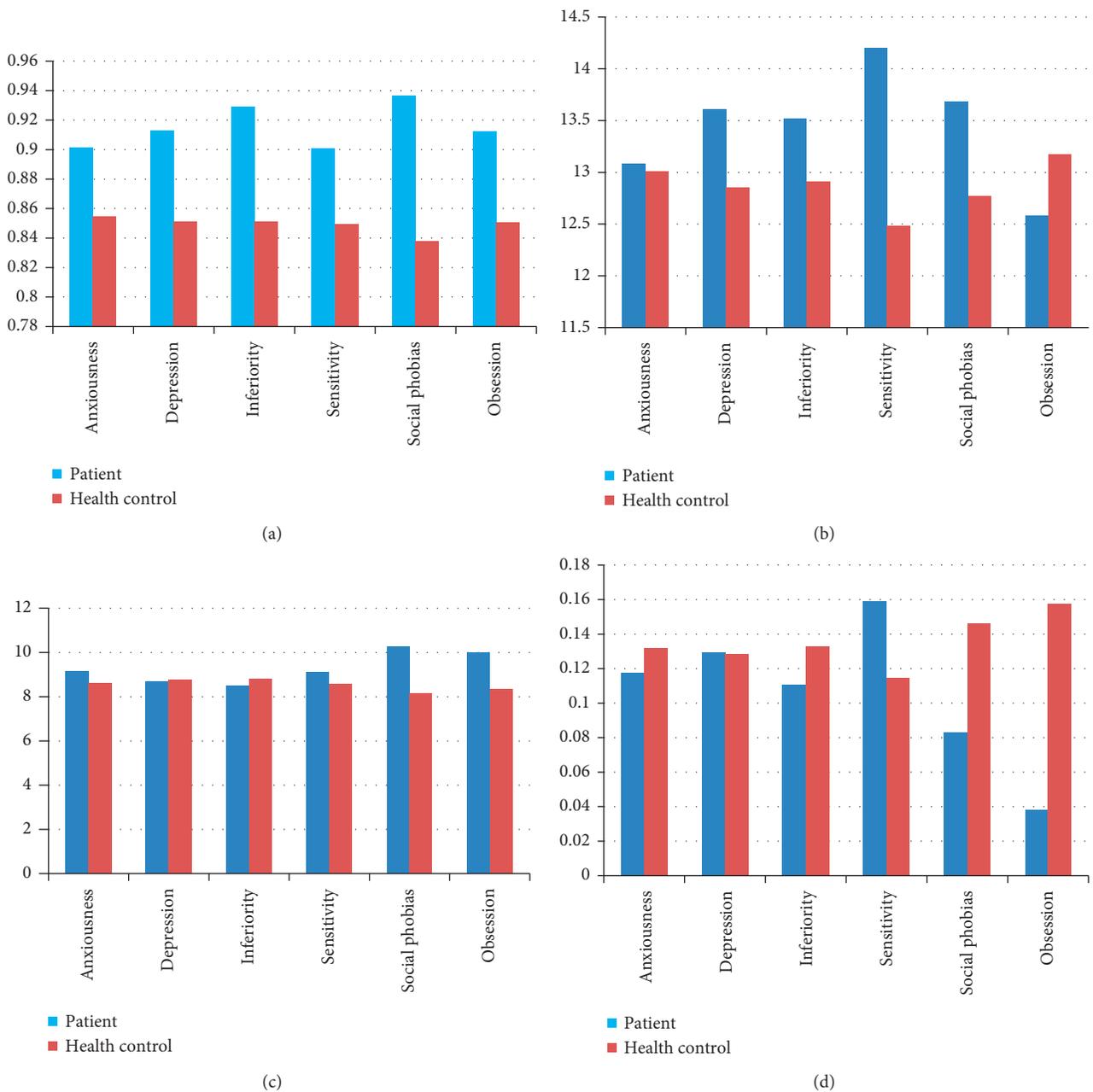


FIGURE 3: The characteristic of metaphor usage of each mental illness. (a) The probability of sentence with metaphor. (b) The number of positive metaphor. (c) The number of negative metaphor. (d) The average sentiment score of metaphor.

TABLE 4: Prediction performance on the eRisk 2017 dataset.

Method		Accuracy	F1-score
Trotzek et al.	CNN	0.88	0.59
	LR	0.88	0.69
MSM	Sentiment	0.81	0.61
	Metaphor	0.87	0.56
	ALL	0.89	0.70

All: sentiment + metaphor.

TABLE 5: Accuracy of baseline and MSM on six mental illnesses' prediction.

Method	Trotzek et al.			MSM	
	CNN	LR	All	Sent	Meta
Anxiousness	0.75	0.71	<b>0.82</b>	0.61	0.71
Depression	0.73	0.69	<b>0.75</b>	0.63	0.70
Inferiority	0.78	0.72	0.80	0.80	<b>0.85</b>
Sensitivity	0.58	0.62	<b>0.80</b>	0.53	0.78
Social phobia	0.64	0.65	<b>0.71</b>	0.60	0.70
Obsession	0.68	0.74	<b>0.78</b>	0.72	0.75
Average	0.69	0.69	<b>0.78</b>	0.65	0.75

Sent: sentiment-based feature set; meta: metaphor-based feature set; all: sent + meta.

TABLE 6: F1-score of baseline and MSM on six mental illnesses' prediction.

Method	Trotzek et al.			MSM	
	CNN	LR	All	Sent	Meta
Anxiousness	0.57	<b>0.65</b>	0.64	0.51	0.54
Depression	0.50	0.64	<b>0.67</b>	0.51	0.58
Inferiority	0.46	0.63	0.62	0.51	<b>0.71</b>
Sensitivity	0.46	0.59	<b>0.73</b>	0.44	0.70
Social phobia	0.47	0.61	0.58	0.50	<b>0.62</b>
Obsession	0.42	<b>0.69</b>	0.66	0.50	0.59
Average	0.48	0.64	<b>0.65</b>	0.50	0.62

Sent: sentiment-based feature set; meta: metaphor-based feature set; all: sent + meta.

Lexicon ([http://www.nltk.org/\\_modules/nltk/sentiment/vader.html](http://www.nltk.org/_modules/nltk/sentiment/vader.html)). These tools scan the input text to calculate the frequency of words in different categories, such as the normalized frequency of positive words—words used for expressing positive emotion. The output statistics of word frequency can be transferred into the classifier as text features. The other was a deep learning-based method that employed a convolutional neural network (CNN). We reproduced both methods and compared them with our method for the two datasets.

**5.2. Prediction Method.** We used a metaphor-based feature set and a sentiment-based feature set to build Metaphor-Sentiment Model (MSM) for predicting mental health status. We compared the performance of three common classifiers: logistic regression, SVM, and neural network. The neural network produced the best results as the relationship between features and mental health problems may be nonlinear. In order to prevent the neural network model from overfitting in training small-scale student dataset, we added L2 regularization,

dropout layer, and early-stop mechanism to the model. Meanwhile, the number of layers and hidden layer nodes in the network are determined by testing. 10-fold cross validation is applied in experiment to ensure the performance of model.

The neural network in this paper was built using Keras (<https://github.com/keras-team/keras>), which is a four-layer, fully connected neural network comprising one input layer, two hidden layers, and one output layer. The input layer was the vector that combined the metaphorical features and sentiment features extracted from the data. The two hidden layers had output dimensions of 100 and 50, respectively. The input layer and the two hidden layers used concatenated rectified linear units (CRELU) for the activation function. We added a dropout layer between two hidden layers with a dropout rate of 0.4 to avoid overfitting. The output layer used Softmax for the activation function, which yielded a generalization of logic functions and output vectors with two dimensions.

**5.3. Experiment Performance.** The eRisk2017 dataset has been divided into a training set and a test set [49]. We tested

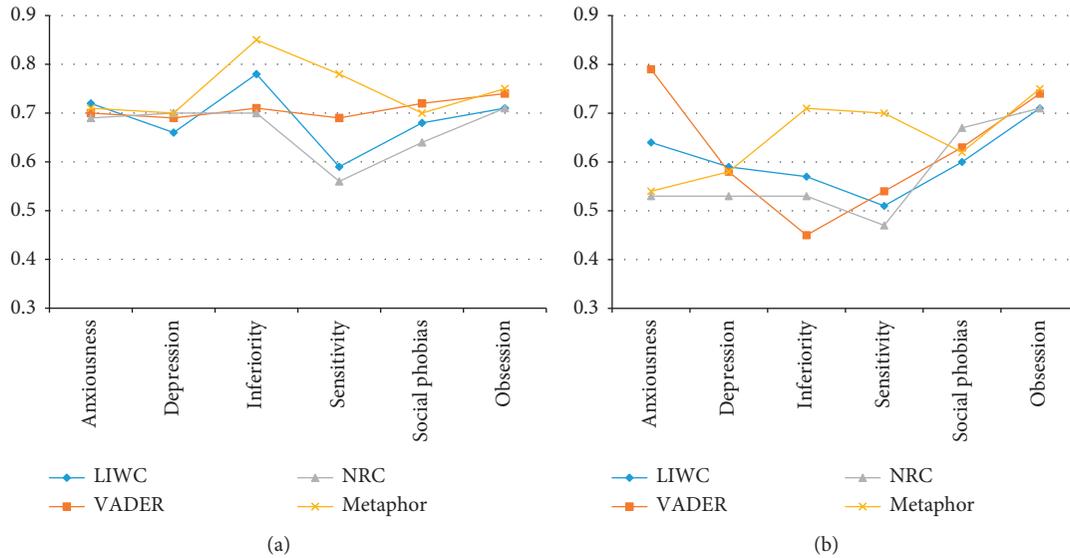


FIGURE 4: Accuracy and F1-score of every feature separately on six mental illness prediction tasks.

MSM using either a sentiment-based feature set, a metaphor-based feature set, or both and compared the results with those using the baseline method. The results are shown in Table 4. Our identification method outperformed the two baseline methods in terms of both accuracy and F1-score. In addition, the results indicate that metaphor-based feature sets are helpful for detecting depression. These results demonstrate the superiority of our prediction method compared with established methods such as those used for our baseline.

We used 10-fold cross validation to partition our composition dataset collected from second-language students to evaluate the prediction performance of MSM compared to the baseline methods. The results are shown in Tables 5 and 6.

Table 5 compares the accuracy of the two baseline methods with that of our method for the prediction of six mental health problems. The results show that MSM obtained the highest accuracy, with an average accuracy for all six mental health problems that was significantly higher than baseline (Fisher's exact test:  $p < 0.05$ ), especially with regard to the sensitivity prediction task (Fisher's exact test:  $p < 0.005$ ). The metaphor-based feature set played an important role in MSM and outperformed the sentiment-based feature set for all mental health group prediction tasks. It achieved the highest accuracy for predicting inferiority, which corresponds to the significant difference in metaphor use between students with inferiority and those without inferiority, as discussed above.

Considering the unbalanced samples, we also computed the F1-score for all mental health problem prediction tasks. The results are shown in Table 6. Overall, using all feature sets, our method showed the highest performance for prediction of the six mental health problems in terms of the average F1-score. The improvement in F1-score was significant with respect to students with sensitivity (Fisher's exact test:  $p < 0.05$ ). The logistic regression baseline method achieved

the same results as our method overall. The metaphor-based feature set from our method showed the highest F1-scores for predicting inferiority and social phobia.

To further assess the effectiveness of metaphor feature sets, we compared metaphor-based feature set and sentiment-based feature set with three common text features that are extracted by LIWC, NRC Emotion Lexicon, and VADER Sentiment Lexicon and used in the logistic regression baseline method. The line charts shown in Figure 4 present the accuracy and F1-score performance of each feature separately for prediction of the six mental health problems using the neural network classifier. The results show that metaphor feature sets are more effective at predicting inferiority and sensitivity than other textual features and equally effective at predicting other mental health problems.

## 6. Conclusions

To the best of our knowledge, we are the first to demonstrate the prediction of six mental problems—anxiety, depression, inferiority, sensitivity, social phobias, and obsession—using automatically detected metaphors in texts. We used metaphor-based feature sets and sentiment-based feature sets to predict these mental health problems using a compositional dataset produced by second-language students and the eRisk2017 dataset collected from Social Media. Our results show that the proposed method can predict the mental health status of authors of written texts, and our algorithm performs well compared to other state-of-the-art methods. We also analyzed differences in metaphor use among students with various mental health problems and evaluated the effectiveness of metaphor sets compared with other textual features in predicting mental health status from a compositional dataset of second-language students.

Our work demonstrates the value of metaphorical textual features for the prediction of mental health problems. The

experiment results remind us of the importance of metaphor, as a deep, complex, and cognitive feature for mental health identification, which often focuses on shallow linguistic features. Importantly, we show that metaphor is predictive even for nonnative speakers of the language. We also contribute to a novel, scarce, and valuable dataset, consisting of second-language speakers' essays and data on authors' mental health problems obtained from a psychological survey, which we will release publicly. We hope this paper will stimulate new ideas for the identification and prediction of mental health status through analysis of text and lead to improvement of automated methods for this purpose.

## 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 they have no conflicts of interest.

## Acknowledgments

This work was supported by Social Science Planning Fund Program, Niaoning Province, under grant L20BYY023.

## References

- [1] K. Levecque, F. Anseel, A. De Beuckelaer, J. Van der Heyden, and L. Gisle, "Work organization and mental health problems in PhD students," *Research Policy*, vol. 46, no. 4, pp. 868–879, 2017.
- [2] R. Bruffaerts, P. Mortier, G. Kiekens et al., "Mental health problems in college freshmen: prevalence and academic functioning," *Journal of Affective Disorders*, vol. 225, pp. 97–103, 2018.
- [3] T. M. Evans, L. Bira, J. B. Gastelum, L. T. Weiss, and N. L. Vanderford, "Evidence for a mental health crisis in graduate education," *Nature Biotechnology*, vol. 36, no. 3, pp. 282–284, 2018.
- [4] L. Cameron, *Metaphor in Educational Discourse*, A&C Black, London, UK, 2003.
- [5] E. Shutova and S. Teufel, "Metaphor corpus annotated for source-target domain mappings," in *Proceedings of the International Conference on Language Resources and Evaluation, LREC*, vol. 2, Citeseer, Valletta, Malta, May 2010.
- [6] D. Zhang, H. Lin, X. Liu, H. Zhang, and S. Zhang, "Combining the attention network and semantic representation for Chinese verb metaphor identification," *IEEE Access*, vol. 7, pp. 137103–137110, 2019.
- [7] P. H. Thibodeau, T. Matlock, and S. J. Flusberg, "The role of metaphor in communication and thought," *Language and Linguistics Compass*, vol. 13, no. 5, Article ID e12327, 2019.
- [8] G. Lakoff and M. Johnson, "Conceptual metaphor in everyday language," *The Journal of Philosophy*, vol. 77, no. 8, pp. 453–486, 1980.
- [9] Z. Kovecses, "Anger: its language, conceptualization, and physiology in the light of cross-cultural evidence," *Language and the Cognitive Construal of the World*, Cambridge University, Cambridge, UK, pp. 181–196, 1995.
- [10] S. Mohammad, E. Shutova, and P. Turney, "Metaphor as a medium for emotion: an empirical study," in *Proceedings of the Fifth Joint Conference on Lexical And Computational Semantics*, pp. 23–33, Berlin, Germany, August 2016.
- [11] V. Dankers, M. Rei, M. Lewis, and E. Shutova, "Modelling the interplay of metaphor and emotion through multitask learning," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing And the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2218–2229, Association of Computational Linguistics, Hong Kong, China, November 2019.
- [12] R. M. Billow, J. Rossman, N. Lewis, D. Goldman, and C. Raps, "Observing expressive and deviant language in schizophrenia," *Metaphor and Symbol*, vol. 12, no. 3, pp. 205–216, 1997.
- [13] B. Elvevåg, K. Helsen, M. De Hert, K. Sweers, and G. Storms, "Metaphor interpretation and use: a window into semantics in schizophrenia," *Schizophrenia Research*, vol. 133, no. 1, pp. 205–211, 2011.
- [14] E. D. Gutierrez, G. A. Cecchi, C. Corcoran, and P. Corlett, "Using automated metaphor identification to aid in detection and prediction of first episode schizophrenia," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2923–2930, Copenhagen, Denmark, September 2017.
- [15] J. Llewellyn-Beardsley, S. Rennick-Egglestone, F. Callard et al., "Characteristics of mental health recovery narratives: systematic review and narrative synthesis," *PLoS One*, vol. 14, pp. 1–31, 2019.
- [16] D. Magaña, "Cultural competence and metaphor in mental healthcare interactions: a linguistic perspective," *Patient Education and Counseling*, vol. 102, no. 12, pp. 2192–2198, 2019.
- [17] R. Rasetti, V. S. Mattay, L. M. Wiedholz et al., "Evidence that altered amygdala activity in schizophrenia is related to clinical state and not genetic risk," *American Journal of Psychiatry*, vol. 166, no. 2, pp. 216–225, 2009.
- [18] F. M. M. Citron and A. E. Goldberg, "Metaphorical sentences are more emotionally engaging than their literal counterparts," *Journal of Cognitive Neuroscience*, vol. 26, no. 11, pp. 2585–2595, 2014.
- [19] X. Li, M. Zhao, M. Zeng et al., "Hardware Impaired Ambient Backscatter Noma Systems: reliability and security," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2723–2736, 2021.
- [20] L. Sun, L. Wan, and X. Wang, "Learning-based resource allocation strategy for industrial iot in uav-enabled mec systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7–, pp. 5031–5040, 2020.
- [21] L. Sun, J. Wang, and B. Lin, "Task allocation strategy for mec-enabled iiots via bayesian network based evolutionary computation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5–, pp. 3441–3449, 2020.
- [22] H. Wang, L. Xu, Z. Yan, and T. A. Gulliver, "Low complexity mimo-fbmc sparse channel parameter estimation for industrial big data communications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5–, pp. 3422–3430, 2020.
- [23] L. Wan, L. Sun, K. Liu, X. Wang, Q. Lin, and T. Zhu, "Autonomous vehicle source enumeration exploiting non-cooperative UAV in software defined internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2020.
- [24] L. Wan, Y. Sun, L. Sun, Z. Ning, and J. J. P. C. Rodrigues, "Deep learning based autonomous vehicle super resolution doa estimation for safety driving," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

- [25] L. Xu, H. Wang, and T. A. Gulliver, "Outage probability performance analysis and prediction for mobile iov networks based on ics-bp neural network," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3524–3533, 2020.
- [26] M. Conway and D. O'Connor, "Social media, big data, and mental health: current advances and ethical implications," *Current Opinion in Psychology*, vol. 9, pp. 77–82, 2016.
- [27] S. Graham, C. Depp, E. E. Lee et al., "Artificial intelligence for mental health and mental illnesses: an overview," *Current Psychiatry Reports*, vol. 21, no. 11, p. 116, 2019.
- [28] S. D'Alfonso, "Ai in mental health," *Current Opinion in Psychology*, vol. 36, pp. 112–117, 2020.
- [29] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," *Psychological Medicine*, vol. 49, no. 09, pp. 1426–1448, 2019.
- [30] S. Mulyana, S. Hartati, R. Wardoyo, and Subandi, "A processing model using natural language processing (nlp) for narrative text of medical record for producing symptoms of mental disorders," in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pp. 1–6, Semarang, Indonesia, October 2019.
- [31] M. Levis, C. Leonard Westgate, J. Gui, B. V. Watts, and B. Shiner, "Natural language processing of clinical mental health notes may add predictive value to existing suicide risk models," *Psychological Medicine*, pp. 1–10, 2020.
- [32] H. A. Schwartz, J. Eichstaedt, M. Kern et al., "Towards assessing changes in degree of depression through facebook," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 118–125, University of Cambridge, Cambridge, UK, October 2014.
- [33] C. Homan, R. Johar, T. Liu, M. Lytle, V. Silenzio, and C. O. Alm, "Toward macro-insights for suicide prevention: analyzing fine-grained distress at scale," in *Proceedings of the Workshop on Computational Linguistics And Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 107–117, Baltimore, MA, USA, June 2014.
- [34] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying mental health signals in twitter," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 51–60, Baltimore, MA, USA, June 2014.
- [35] R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, "Natural language processing in mental health applications using non-clinical texts," *Natural Language Engineering*, vol. 23, no. 5, pp. 649–685, 2017.
- [36] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 217–226, 2014.
- [37] H.-H. Franco-Penya and L. M. Sanchez, "Text-based experiments for predicting mental health emergencies in online web forum posts," in *Proceedings of the Third Workshop on Computational Linguistics And Clinical Psychology*, pp. 193–197, San Diego, CA, USA, January 2016.
- [38] D. N. Milne, G. Pink, B. Hachez, and R. A. Calvo, "Clpsych 2016 shared task: triaging content in online peer-support forums," in *Proceedings of the Third Workshop on Computational Linguistics And Clinical Psychology*, pp. 118–127, San Diego, CA, USA, June 2016.
- [39] A. Cohan, S. Young, and N. Goharian, "Triaging mental health forum posts," in *Proceedings of the Third Workshop on Computational Linguistics And Clinical Psychology*, pp. 143–147, San Diego, CA, USA, June 2016.
- [40] F. Ramiandrisoa, J. Mothe, F. Benamara, and V. Moriceau, "IRIT at eRisk 2018," in *Proceedings of the 9th Conference And Labs of the Evaluation Forum, Living Labs (CLEF 2018)*, pp. 1–12, Avignon, France, September 2018.
- [41] D. E. Losada, F. Crestani, and J. Parapar, "Overview of erisk: early risk prediction on the internet," in *International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 343–361, Springer, Berlin, Germany, 2018.
- [42] J. Weerasinghe, K. Morales, and R. Greenstadt, "'Because... I was told... So much': linguistic indicators of mental health status on twitter," *Proceedings on Privacy Enhancing Technologies*, vol. 2019, no. 4, pp. 152–171, 2019.
- [43] A. Benton, M. Mitchell, and D. Hovy, "Multitask learning for mental health conditions with limited social media data," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 152–162, Valencia, Spain, April 2017.
- [44] M. Troztek, S. Koitka, and C. M. Friedrich, "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences," 2018, <http://arxiv.org/abs/1804.07000>.
- [45] I. Sekulic and M. Strube, "Adapting deep learning methods for mental health prediction on social media," in *Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019)*, Hong Kong, China, November 2019.
- [46] L. Tavabi, "Multimodal machine learning for interactive mental health therapy," in *2019 International Conference on Multimodal Interaction, ICMI'19*, pp. 453–456, Association for Computing Machinery, New York, NY, USA, October 2019.
- [47] Z. Xu, V. P'erez-Rosas, and R. Mihalcea, "Inferring social media users' mental health status from multimodal information," in *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 6292–6299, European Language Resources Association, Marseille, France, May 2020.
- [48] J. B. Hirsh and J. B. Peterson, "Personality and language use in selfnarratives," *Journal of Research in Personality*, vol. 43, no. 3, pp. 524–527.
- [49] D. E. Losada, F. Crestani, and J. Parapar, "Erisk 2017: clef lab on early risk prediction on the internet: experimental foundations," in *Proceedings of the International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 346–360, Springer, Berlin, Germany, August 2017.
- [50] D. E. Losada and F. Crestani, "A test collection for research on depression and language use," in *Proceedings of the International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 28–39, Springer, Berlin, Germany, August 2016.
- [51] J. P. Pestian, P. Matykiewicz, M. Linn-Gust et al., "Sentiment analysis of suicide notes: a shared task," *Biomedical Informatics Insights*, vol. 5, pp. 1–6, 2012.
- [52] S. Gohil, S. Vuik, and A. Darzi, "Sentiment analysis of health care tweets: review of the methods used," *JMIR Public Health and Surveillance*, vol. 4, no. 2, p. e43, 2018.
- [53] R. Mao, C. Lin, and F. Guerin, "Word embedding and wordnet based metaphor identification and interpretation," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1222–1231, Melbourne, Australia, July 2018.

## Review Article

# Analysis on the Development Status of Coal Mine Dust Disaster Prevention Technology in China

Hui Zhang,<sup>1</sup> Weihai Han,<sup>2</sup> Youlin Xu ,<sup>1</sup> and Zhenfeng Wang<sup>2</sup>

<sup>1</sup>*Institute of Mining Engineering, Guizhou Institute of Technology, Guiyang 550003, China*

<sup>2</sup>*School of Energy Science and Engineering, Henan Polytechnic University, Jiaozuo 454003, China*

Correspondence should be addressed to Youlin Xu; xuyoulin2006@163.com

Received 19 February 2021; Revised 29 March 2021; Accepted 8 April 2021; Published 20 April 2021

Academic Editor: Han Wang

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

In recent years, with the increasing level of mechanization, automation, and intelligence in mine mining, dust pollution in the working environment of coal mines has become increasingly serious. Coal mine dust prevention is an important work related to the life, health, and production safety of miners, and it is also one of the technical problems of mine safety in production. With the continuous revision and improvement of China's Occupational Disease Prevention and Control Law, coal mining enterprises have generally strengthened the publicity, prevention, and control of occupational diseases among employees, and the control of coal mine dust has been used as a means for enterprises to improve the production environment and strengthen the occupational health of employees. Key work-based studies have shown that China's coal mines have formed a theoretical system and technology system of dust prevention and control. In the future, China's coal mines will start from intelligent dust prevention, achieve high-precision dust sensing-transmission-assessment and early warning, and develop a combined dust collector that integrates the functions of atomization dust removal, miniaturization, dry and wet mixing, and large air suction capacity. The combined dust collector realizes the efficient ventilation control and dust removal of the fine dust in wide-area complex spaces such as fully mechanized mining face and fully mechanized mining face. At the same time, breakthroughs have been achieved in low-permeability coal seams, such as strong hydraulic permeability-enhancing technology, intelligent dust-proof robots, and chemical dust suppression. This article introduces the basic concepts, generation, distribution, and hazards of coal mine dust and analyzes the characteristics, applicable conditions, and use effects of various dust control measures such as ventilation dust removal and wet dust removal. Moreover, this article also proposes specific prevention and control measures for related occupational diseases and discusses the development trend of dust prevention and control technology in the hope of providing guidance and reference for coal mine dust prevention and control.

## 1. Introduction

Coal is China's main energy source, accounting for more than 60% of the primary energy. The status of this subject energy will be difficult to change for a long time in the future. More than 90% of coal production in China comes from underground mining. Roadways are a necessary passage for coal mines. According to statistics, the total length of roadways newly excavated in coal mines in China each year reaches 12,000 km [1, 2]. At present, China's coal mine production safety situation continues to improve and the number of coal mine accidents dropped significantly, but the coal industry has still a relatively high prevalence of

pneumoconiosis because, for a very long time, enterprises attached importance to production safety, ignoring occupational health work. As a result, relevant scientific research investment is low, technological progress cannot match the rapid increase in coal production, and the working environment of coal mines cannot be effectively improved; especially, the dust concentration in coal mine workplaces has been too high for a long time and the number of new cases of pneumoconiosis has increased significantly [3]. For nearly 10 years, with the increasing level of intelligence, mechanization, and automation of mines, the probability of accidents and disasters caused by factors such as gas, coal dust, and fire has also increased. In particular, dust pollution in

coal mines has severely affected the safety of enterprises and threatened the occupational health of workers [4]. According to statistics, the proportion of floating dust generated by various production links in coal mines is as follows: The coal mining face accounts for 50%, the tunneling face accounts for 35%, the shotcrete operation point accounts for 10%, and the coal loading and transportation and unloading links account for 5% [5], of which the dust production in coal mining, tunneling and bolting, and shotcreting operations accounted for more than 95% of the mine's total dust production. According to on-site measurements, without any dust prevention measures on the fully mechanized mining face, when the shearer cutting and frameshifting work together, the time-weighted total dust mass concentration in the main work area of personnel can reach 500~850 mg/m<sup>3</sup>, even if dust-proof measures are taken, the harsh working environment of the working face cannot be changed [6]. Literature [7] stated that in recent years, anchor spraying operations have mainly adopted dry spraying and wet spraying processes. Although the operation is simple, the free silica content of the dust on the job site is as high as 80%, and the time-weighted dust mass concentration exceeds 200 mg/m<sup>3</sup>. The above-mentioned dust concentration also greatly exceeds the upper limit set by the state [8]; the number of miners who have pneumoconiosis risk is extremely high. At present, most scholars only conduct research on a single coal mine dust prevention technology, and there are few studies on comprehensive coal mine dust prevention theory and technology systems. Therefore, in order to reflect the latest research results, promote the development of coal mine dust prevention and control technology in China, and improve mine safety and occupational health, this article analyzes the research status of coal mine dust prevention in China, summarizes the achievements of dust prevention theory and technology, and proposes future development directions.

## 2. Pneumoconiosis Status of the Coal Industry

The main occupational hazards in coal mines are dust, high temperature, noise, and toxic and harmful gases. Among them, the proportion of pneumoconiosis caused by dust is the highest. Coal mine pneumoconiosis accounts for more than 70% of coal mine occupational diseases, and it accounts for more than 55% of China's pneumoconiosis. Because dust can cause harm to people's health, the "Coal Mine Safety Regulations" stipulates the maximum allowable concentration of total dust and respirable dust in the workplace. The "Regulations on the Prevention of Occupational Hazards in Coal Mine Workplaces" also specifically regulate the concentration of respirable dust in the workplace. Specific criteria are listed in Tables 1 and 2:

Therefore, coal mine dust is the main occupational hazard factor in coal mines. Over the past decade, though China's coal mine production safety situation continues to improve, the coal mine pneumoconiosis reported cases was a higher trend. As can be seen from Figure 1, China coal mine accidents (excluding occupational diseases) in terms of the number of deaths continued to decline since 2005 from more

than 6,000 people to 316 people in 2019. On the contrary, the reported cases of coal mine pneumoconiosis are gradually increasing; especially since 2010, they have been rising sharply, with more than 13,000 cases each year, with an average of more than 14 thousand cases, which is about 2.7 times the average pneumoconiosis embodiment from 2005 to 2009 [9].

In addition to management reasons, the severe status of coal mine pneumoconiosis is also caused by coal mine enterprises and workers' insufficient awareness of the severity of dust hazards and insufficient attention to occupational health. Therefore, the promotion of new technologies is severely constrained, which has led to the coal industry's pneumoconiosis be in serious condition.

## 3. The Harm of Coal Mine Dust to Human Health

The hazards of coal mine dust mainly include four aspects:

- (1) The spontaneous combustion nature of coal dust and the explosive nature of coal dust: coal dust explosion hazard in coal mines is one of the most serious disasters. Compared with gas explosions, coal dust explosions have greater intensity and disaster range, more destructiveness, and more serious consequences [10]. The disaster of coal dust explosion in China is very serious. According to statistics, from the beginning of 2000 to the beginning of 2019, 16 coal dust accidents occurred nationwide, resulting in more than 500 deaths.
- (2) Coal dust causes pneumoconiosis. A portion of the dust generated in the production process of coal mines is eliminated through measures such as spray dust reduction or ventilation and dust removal, and a portion of the dust particles with smaller particle size is flying and suspended in the production space. For workers, due to the long contact with respirable dust, respirable dust will slowly be deposited in the human lung, for example, small bronchi or alveolar. Prolonged exposure to respiratory dust will produce a series of physiological and pathological changes, leading to lung tissue fibrosis disease [11], called coal worker's pneumoconiosis. In the coal industry, pneumoconiosis is a common occupational disease. Compared with "obvious" coal mine disasters such as gas explosions, pneumoconiosis is more lethal, can damage more groups, and has higher potential harm. According to statistics, the number of deaths due to pneumoconiosis each year is 6 times the number of miners killed in mining disasters and other industrial accidents [12]. According to statistics, since 2010, China has reported an average of 28,000 new cases of occupational diseases each year. As of the end of 2018, a total of 975,000 cases of occupational diseases have been reported, of which 873,000 cases were occupational pneumoconiosis cases, accounting for about 90% of the total number of reported cases of occupational diseases. From 2008 to 2018, the total

TABLE 1: Dust concentration in workplace air standard.

The content of free SiO <sub>2</sub> in dust (%)	Maximum allowable concentration (mg/m <sup>3</sup> )	
	Total dust	Breathable dust
<10	10	3.5
10~50	2	1
50~80	2	0.5
≥80	2	0.3

TABLE 2: Judgment standard for management limit of dust exposure concentration in coal mine workplace.

Type of dust	The content of free SiO <sub>2</sub> in dust (%)	Maximum allowable concentration of respirable dust (mg/m <sup>3</sup> )
Coal dust	≤5	5.0
	5~10	2.5
Rock dust	10~30	1.0
	30~50	0.5
Cement dust	≥50	0.2

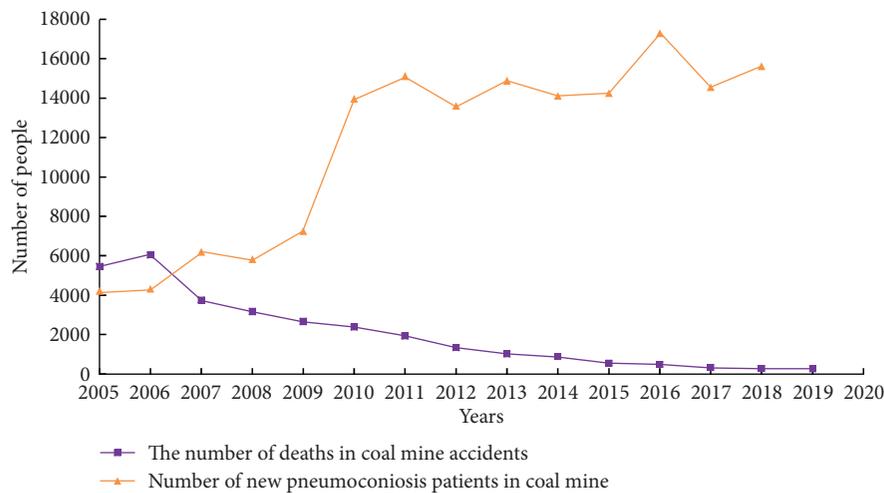


FIGURE 1: Comparison of the number of new pneumoconiosis and accident deaths in coal mines across the country in recent years.

number of new cases of pneumoconiosis nationwide was 247,611, of which 125,418 were newly added to coal mine pneumoconiosis, accounting for about 50.65% [13].

- (3) The visibility of the workplace is reduced and work-related injury accidents are increased.
- (4) The wear of the rotating parts of the equipment is accelerated and the working accuracy and life of the equipment are reduced.

#### 4. The Law of Coal Dust Dispersion Pollution in China

**4.1. Coal Dust Pollution at Fully Mechanized Mining Face.** Dust pollution in fully mechanized mining has always been the focus of mine dust prevention and control. The research methods for dust pollution in fully mechanized mining are mainly divided into experiments and numerical simulations. Guo et al. [14] established a mathematical model of dust movement based on the theory of gas-solid two-phase flow. According to the specific properties and measured data of a

fully mechanized mining face, the law of dust movement generated by a fully mechanized mining face was clarified. By constructing a two-stage mathematical model, Perret et al. [15] simulated the dust distribution in the mine space and studied the diffusion law of coal dust. Relevant scholars have also used CFD (computational fluid dynamics) numerical calculation methods to study the law of dust migration in fully mechanized mining faces. Seaman et al. [16] used the LES large eddy method to simulate the gas motion process and used the Euler-Lagrange method to numerically simulated the spatial distribution of dust particles. Wang et al. [17] used CFD methods to calculate and simulate the airflow and dust flow characteristics in the mine and, based on the results, proposed two possible dust control schemes (Figures 2 and 3).

Sun et al. [18] used numerical simulation to study the influence of wind flow turbulence on dust pollution and, at the same time, proposed a partial spray closed dust control technology, which improved the dust removal rate (Figure 4). Tan et al. [19] compared the measured dust mass concentration distribution on-site with the simulated coal cutting dust movement law on fully mechanized mining face

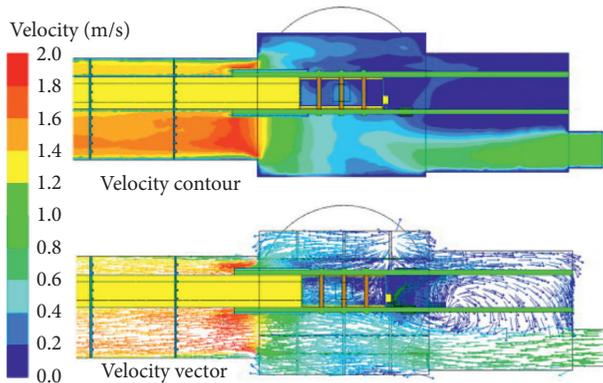


FIGURE 2: Distribution of airflow velocity at the height of breathing zone.

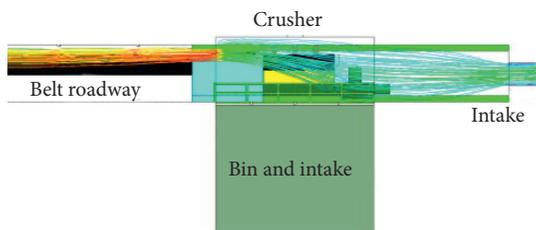


FIGURE 3: Inlet air streamline diagram.

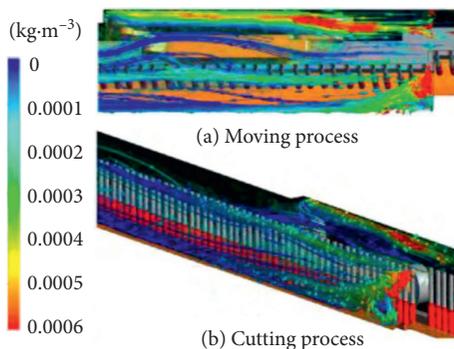


FIGURE 4: Dust concentration movement trace diagram.

and obtained several important factors that affect the dust quality concentration on fully mechanized mining face, mainly the wind speed of the working face, the speed of the shearer drum, the speed of the scraper conveyor chain, and the coal wall surface conditions. Yao et al. [20] conducted numerical simulations on the airflow movement and coal dust distribution law of fully mechanized mining face under different ventilation conditions and obtained the dust reduction mechanism of upward ventilation and downward ventilation and the optimal dust exhaust wind speed of large inclination fully mechanized mining face.

In summary, domestic and foreign scholars have conducted research on the law of dust pollution on fully mechanized coal mining faces through experiments and numerical simulations. According to the characteristics of different dust sources on a fully mechanized mining face, a mathematical model of airflow-dust particle DPM coupled

flow suitable for a fully mechanized mining face is established, and the Euler-Lagrange method is used to describe the turbulent diffusion of airborne dust. The CFD numerical simulation of the dispersion process of the coal dust generated by the shearer cutting on the fully mechanized working face, the coal dust generated by the moving frame, and the coal dust carried by the ventilation is carried out. The results obtained have laid a scientific theoretical foundation for dust prevention and control in a fully mechanized mining face.

#### 4.2. Coal Dust Pollution at Comprehensive Excavation Face.

There are many factors affecting the dust dispersion in the comprehensive mechanized tunneling face, such as the changeable position of dust production, and the complicated technical process, and the researchers have conducted in-depth and detailed research. Wen et al. [21] used a combination of experiment and numerical simulation to study the movement of wind and dust on the fully mechanized excavation face. The research results show that there is a corresponding relationship between the airflow streamline and the dust trace, and the test and simulation results are consistent with the field measurement. Wu et al. [22] analyzed the dust generation mechanism of the fully mechanized excavation face and used the Fluent software to conduct a numerical simulation study on the dust distribution and migration law of the working floor to obtain the overall dust migration and distribution law.

## 5. The Theory of Coal Dust Wetting

Coal seam water injection and spray dust reduction are the main measures to prevent and control dust on the mining face, water injection into the coal seam to be mined through boreholes, or fog field generated by high-pressure atomization and can moisten the coal (dust), thereby reducing the floating coal produced during the mining process. Therefore, the wettability of pulverized coal is an important factor affecting the dust reduction effect. For this reason, in order to better understand the wettability of coal dust, it is first necessary to analyze and study the solid-liquid relationship between coal dust and water. In this regard, after a detailed analysis of the surface wettability of ultrafine coal powder, Abhishek Kumar et al. [23] found that with the degree of deterioration, the surface hydrophobicity of ultrafine coal powder after pulverization will continue to increase. Zhao et al. [24] analyzed the characteristics of coal surface free energy and the microscopic mechanism of coal adsorption of water based on the structural characteristics of coal macromolecules and surface and concluded that coal adsorption of water molecules is multilayer adsorption. The first layer of water absorption is mainly due to the hydrogen bonding between coal and water molecules, while the adsorption of other water molecular layers is caused by long-range forces between molecules. Sergei et al. [25] analyzed the movement process of water in the coal body when water was injected into the coal seam, applied the theory of interface chemistry to analyze the wetting process of water on the coal surface,

summarized the conditions under which the coal body can wet by itself, and the wetting mechanism of water on the coal surface is preliminarily explained.

Kollipara et al. [26] studied their physical properties, mineralogical properties, and wetting properties for different dust samples. The fixed time wettability (trying to simulate the wettability around the mining environment) and absolute time wettability (assessing internal wetting rate) are used to evaluate the wettability of dust, and it is concluded that the fixed wettability of coal dust is 57% to 99%, the wettability of most mines is above 90%, and the wettability in the middle of the coal seam is the worst. The contact time between dust particles and water droplets is an important factor in improving the wettability of coal dust. The results show that it takes more time for dust with larger particles to be completely wetted. Arkhipov et al. [27] proposed a new method for estimating the wettability of fine coal particles on water droplets, which improved the accuracy of particle wettability estimation. The combination of wet dust removal and chemical dust suppression, especially anionic surfactants, has the most significant impact on coal wettability. Li et al. [28] and others systematically analyzed the physical properties of coal dust and its wetting behavior. The effects of different surfactants on the wettability of coal dust were studied and compared with deionized water. The research results show that the finer the particle size of the coal, the more complex the microstructure of the coal, and the worse the wettability of the coal. Among the three different coal dust samples, the coal with a higher volatile content has poorer wetting performance because the volatiles are easier to release and a gas film is more likely to form around the particles.

## 6. Theory of Dust Mist Condensation and Wet Dust Removal

Due to the particularity of coal mine production conditions, in addition to coal seam water injection prewetting coal body technology, dust mist-condensation and wet dust suppression technology is currently an effective measure generally adopted in coal mining face in China. Zhou and Wang [29] believed that the most obvious factor that affects the trapping of dust particles by fog droplets is the droplet diameter. The small diameter of the droplets makes it easier to capture dust particles. However, during the actual dust reduction operation, if the diameter of the formed droplets is too small, it will cause the droplets to evaporate too quickly and the retention time is too short, which will affect the overall dust collection efficiency. Lin et al. [30] studied the gas-liquid ratio of single-phase nozzles and two-phase nozzles, as well as the variation law between axial distance and radial distance. Zhou et al. [31] combined experiments and theoretical analysis to study the spray parameters of the dust suppression system and obtained the relationship between pressure and other parameters. In order to correctly evaluate the interaction between water droplets, dust, and flowing airflow, Swanson and Langefeld [32] established a wind tunnel to simulate underground operations under different environmental conditions, evaluated the control conditions

and requirements of each dust source, and developed an optimized dust source control and dust removal system. Tessum and Raynor [33] used a pneumatic particle size analyzer to measure the particle size and concentration of the charge-separated particles and studied the dust collection performance of different types of surfactants on different coal dust particle sizes and charges. The above studies have promoted the development of dust-mist-condensation and wet-type dust reduction technology. However, the current spray dust reduction measures in fully mechanized mining face still have problems such as the lack of scientificity in nozzle selection and layout and poor spray dust reduction effects. In order to further improve the efficiency of spray dust reduction, the current more advanced theories are mainly to improve the dust mist-condensation mechanism from a microscopic point of view by measuring the droplet size, velocity, and concentration distribution of different types of nozzles [34], formed a system of nozzle atomization characteristics and microspray dust reduction mechanism, and developed a new type of high-efficiency single-water and Fengshui atomizing nozzles, which can be made of copper, stainless steel, ceramics and other forms (Figure 5).

## 7. Ventilation and Dust Removal in China

Ventilation and dust removal are to use the mine ventilation system to dilute and discharge the dust in the mine air to prevent excessive dust accumulation and concentration exceeding the limit. Doing a good job of ventilation is an important part of achieving a good dust-proof effect. Ventilation and dust removal methods and measures for tunneling working face are as follows:

- (1) The most commonly used method is to use the local ventilator ventilation method. This kind of ventilation method has three types: press-in type, extraction type, and mixed type, and mixed type is generally used.
- (2) Technology is used to control the dust-laden airflow to diffuse outwards, such as wall air ducts and high-pressure wind shielding.
- (3) The long-pressure short-extraction dust removal system is matched with the roadheader and dust collector to comprehensively reduce dust.
- (4) Take wind-guiding measures, set up wind barriers, or use airflow sprayers to guide the wind to isolate dust.

Ventilation and dust removal methods and measures for fully mechanized mining faces are as follows:

- (1) Select the best ventilation parameters (the best dust exhaust wind speed is 1.5~4 m/s) to ensure the effect of ventilation and dust removal.
- (2) Change the ventilation system or the direction of airflow at the working face (W-type and E-type ventilation systems are the best, and adopting the direction of coal flow; that is, downward ventilation can greatly reduce the concentration of working dust).



FIGURE 5: Atomization effect of 4 MPa pressure nozzle.

- (3) Install simple ventilation facilities to reduce air leakage and ensure air supply. The main methods include setting goaf air curtains, sidewalk air curtains, and coal shearer dust curtains.

## 8. Wet Dust Removal in China

**8.1. Wet Rock Drilling.** Wet rock drilling technology is that when a rock drill (or electric coal drill, bolter, drill truck, etc.) is drilling, it sends pressure water to the bottom of the borehole to moisturize, flush, and discharge the dust generated by the rotation of the drill bit. This technology can effectively reduce the amount of dust produced, and the dust reduction rate can reach about 90%.

**8.2. Dust Reduction by Spray.** Dust reduction by spray is the spraying of water stream into fine water droplets under the rotation and impact of sprayer (also called nozzle). At present, the commonly used spray dust reduction technologies are as follows:

- (1) The mining unit sprays mainly include internal and external spray dust reduction, radial fog screen dust reduction, high-pressure spray negative pressure dust reduction, and dust-containing airflow control dust reduction.
- (2) Interrack spraying includes nozzles installed at the front beam of the support, the front and back of the racks, and the coal caving port to achieve simultaneous spraying and dust reduction when raising, lowering, and moving the rack.

**8.3. Sprinkle Water to Reduce Dust.** Sprinkle water to reduce dust is to wet the deposited coal dust with water so that the coal dust can be condensed into larger particles that can adhere to the surface of the coal and rock, ensuring that it is not easy to fly during shipment. General sprinkling has a poor dust reduction (low pressure) effect and large water consumption. At present, a new technology of high-pressure sprinkling (water pressure greater than 910 kPa) has emerged, which makes sprinkling and dust reduction measures more perfect.

**8.4. Water Cannon Mud.** Water cannon mud is a plastic bag filled with water to replace part of the cannon mud in the blasthole. During blasting, water vaporization and dust condensation achieve the dust reduction effect, which can reduce the dust concentration by 20% to 50%.

**8.5. Coal Seam Water Injection.** Coal seam water injection is to inject pressurized water into the borehole under the action of the water injection system to prewet the native coal dust in the coal body and at the same time make the entire coal body surrounded by water, thereby playing the role of dust reduction and dust suppression. Coal seam water injection is active and effective in reducing dust and has a continuous dust prevention effect. At present, long-hole water injection is widely used internationally.

## 9. Purifying Airflow

Purifying airflow is to install dust-catching facilities in the working face or roadway to build a dust barrier to purify the dust-laden air (airflow). At present, purified water curtain, catching dust net, and wet dust removal device are commonly used.

- (1) Purified water curtain: the full-face water curtain formed by the water flow from multiple nozzles installed on the top of the roadway and the working face has the effect of purifying the inlet airflow and reducing the dust concentration of the return airflow. At present, the purified water curtain has specific forms such as microhole spray, nozzle spray, fine water spray, and Feng Shui linkage spray. There are various control methods such as manual, photoelectric, touch, and mechanical transmission. The dust reduction rate of the water curtain can reach 70% to 95%.
- (2) Catching dust net: install a net in the roadway that can pass airflow and collect dust. It is generally used in conjunction with water curtains and sprays.
- (3) Wet dust removal device: the device removes dust by a collision between dust particles and liquid droplets. At present, China has wet dust collectors such as SCF series dust collectors, KGC series tunneling machine dust collectors, TC series tunneling machine dust collectors, and MAD series airflow purifiers.

## 10. Airtight Dust Extraction

Airtight dust extraction is to use a dust cover to close the local dust production point, and draw the dust-containing air through the pipeline to the dust collector or dust collector to filter, collect, purify and discharge. It is suitable for the situation where water shortage or water cannot be used and the dust source needs to be sealed, common dust-catching aperture, and hole bottom catching dust in two ways. This kind of dry dust-catching method has no water mist in the workplace, the air is dry, and the compressed air consumption is large.

## 11. Personal Protection

Personal protection refers to wearing dust-proof equipment to filter and purify dusty air, so that workers can breathe clean air. Individual dust-proof appliances mainly include dust-proof caps, dust-proof masks, and dust-proof

respirators. With the advancement of science and technology, advanced protective equipment, such as self-priming filter dust masks, power supply air filter dust masks (air supply masks and airflow helmets), and compressed air respirators, has been developed.

## 12. Prevention and Control Measures of Coal Mine Dust and Related Occupational Diseases

Through the analysis of the causes of coal mine dust-related occupational diseases, long-term, large-scale exposure and inhalation of fine dust particles are the root causes of the disease. For coal mine dust control and related occupational disease prevention, measures should be taken from the following aspects:

- (1) Improve the awareness of occupational disease prevention and control of enterprise managers and increase relevant investment. First of all, coal mine managers should raise their awareness of occupational disease prevention and control, clarify the main responsibility of enterprises in the prevention and control of occupational diseases, and establish a people-oriented concept. Based on the principle of observing regulations and caring for employees, we earnestly implement the "Occupational Disease Prevention and Control Law" and take concrete and effective measures to reduce the generation of mine dust and reduce the harm to personnel. In terms of organization and management, a special mine dust prevention management department should be established, equipped with full-time dust prevention technical personnel, and implement the "three simultaneous" occupational disease protection facilities. Increase investment in mine dust prevention and control, set up special funds, timely update mine dust removal equipment, purchase dust masks, face masks, and other labor protection products for employees, and supervise employees to wear and use them correctly.
- (2) Improve the production process to reduce dust generation. For coal shearers, roadheaders, crushers, loading and unloading machinery, self-moving hydraulic supports, and other equipment, an automatic spray device and a dust cover can be installed, and a magnetized water device can be added to enhance the water mist absorption and dust reduction capabilities. In the process of rock tunnel excavation, wet drilling and wet rock drilling should be promoted on the blasting working surface, and the dust in the blasthole should be moistened and washed with pressure water. At the same time, automatic water curtain purification measures should be adopted to achieve the purpose of dust reduction. When the rock formation conditions are not suitable for the wet drilling method, the dry drilling method can be used with a dry dust catcher to catch dust.

- (3) Improve mine ventilation and purify air flow. The "Coal Mine Safety Regulations" stipulate that the minimum wind speed during the excavation of rock tunnels shall not be less than 0.15 m/s. At this wind speed, the mineral dust below 5 microns can be suspended and mixed evenly with the air and discharged with the wind flow. In order to better dilute and discharge the coal dust, the wind speed can be appropriately increased in the operation sites with large dust production and high coal dust concentration, but if the wind speed is too high, the dust content in the air will increase. Therefore, the "Coal Mine Safety Regulations" stipulates that the maximum allowable wind speed at the coal mining face is 4 m/s.
- (4) Strengthen education and training to improve employees' awareness of occupational disease prevention. Most coal mine enterprises lack the awareness of self-protection due to the poor overall quality of their employees and lack of awareness of the dangers of coal mine dust. They often do not use labor protection equipment and dust-proof facilities as required at work. In this regard, the education and training of coal mine employees should be strengthened to make them fully aware of the hazards of coal mine dust to human health so as to increase their awareness of occupational disease prevention and actively use dust removal equipment as required.

## 13. Prospects of Research on Coal Mine Dust Prevention in China

In recent years, with the increasing level of mechanization, automation, and intelligent mining in China's coal mines, dust pollution in the operating environment has become more and more serious, which has greatly affected mine safety production and seriously threatened the occupational health of miners. Scholars at home and abroad have actively explored mine dust prevention and control and have achieved many innovative results. However, different mining areas in China's coal mines have great differences in coal seam conditions, mining methods, technical equipment levels, and comprehensive management levels. There is still a lot of room for improvement in the development of coal mine dust prevention theory and technology. Based on this, in the next period of time, the focus of coal mine dust prevention research will be mainly in the following aspects:

- (1) Develop high-precision, wide measuring range of respirable dust sensor to achieve continuous online monitoring of the mass concentration of respirable dust and to the direction of miniaturization, long-distance transmission, large-area coverage, and continuous monitoring.
- (2) Through the research on the "Set of Coal Mine Dust Monitoring Technology," the mass concentration, toxicity, and other parameters of respirable dust are collected and combined with the accumulated dust

- exposure time, dust physicochemical characteristics, and other indicators, the existing research results of pneumoconiosis are used to establish an early warning indicator system and discrimination model for pneumoconiosis. On this basis, according to the pneumoconiosis danger circle theory and the pneumoconiosis excess risk classification method, prepare to establish a coal mine dust hazard warning system to realize dust hazard warning, and promote the optimization of dust prevention and control measures, scientifically guide the monitoring and management of dust prevention and pneumoconiosis.
- (3) Carry out research on high-pressure hydraulic anti-reflection technology for low-permeability, hydrophobic, and difficult-to-inject coal seams, artificially create cracks or fissures in the coal seam in advance, and by adding a wetting agent, improve the wettability of the hydrophobic hard-to-water-injection coal seam, thereby improving the water injection effect of the hard-water-injection coal seam. Study the coupling technology of gas drainage and coal seam water injection so as to effectively suppress the dust intensity during coal seam mining.
  - (4) Study the key technologies and equipment for the treatment of respirable dust hazards in coal mines and effectively reduce the mass concentration of respirable dust. Research and promote the application of monorail hoisting technology in fully mechanized excavation face to realize the synchronous movement of dust removal system and roadheader. Study the ventilation and dust removal system for rock tunnel excavation.
  - (5) Research and development of coal mine dust monitoring equipment remote online monitoring system, improve the management capacity of the downhole equipment.
- (2) Carry out research on high-pressure hydraulic anti-reflection technology for low-permeability, hydrophobic, and difficult-to-inject coal seams, artificially create cracks or fissures in the coal seam in advance, and by adding a wetting agent, improve the wettability of the hydrophobic hard-to-water-injection coal seam, thereby improving the water injection effect of the hard-water-injection coal seam. Study the coupling technology of gas drainage and coal seam water injection so as to effectively suppress the dust intensity during coal seam mining.
  - (3) Study the key technologies and equipment for the treatment of respirable dust hazards in coal mines and effectively reduce the mass concentration of respirable dust. Research and promote the application of monorail hoisting technology in fully mechanized excavation face to realize the synchronous movement of dust removal system and roadheader. Study the ventilation and dust removal system for rock tunnel excavation.
  - (4) Research and development of coal mine dust monitoring equipment remote online monitoring system improve the management capacity of the downhole equipment.

## 14. Conclusions

- (1) Through the research on the “Set of Coal Mine Dust Monitoring Technology,” the mass concentration, toxicity, and other parameters of respirable dust are collected and combined with the accumulated dust exposure time, dust physicochemical characteristics, and other indicators, the existing research results of pneumoconiosis are used to establish an early warning indicator system and discrimination model for pneumoconiosis. On this basis, according to the pneumoconiosis danger circle theory and the pneumoconiosis excess risk classification method, prepare to establish a coal mine dust hazard warning system to realize dust hazard warning, and promote the optimization of dust prevention and control measures, scientifically guide the monitoring and management of dust prevention and pneumoconiosis.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (51764010 and 51874109) and the Science and Technology Project for Outing and Young Talents of Guizhou (Talents of Science Platform in Guizhou [2019] 5674).

## References

- [1] H. Kang, “Support technologies for deep and complex roadways in underground coal mines: a review,” *International Journal of Coal Science & Technology*, vol. 1, no. 3, pp. 261–277, 2014.
- [2] H. Kang, “Sixty years development and prospects of rock bolting technology for underground coal mine roadways in China,” *Journal of China University of Mining and Technology*, vol. 45, no. 6, pp. 1071–1081, 2016.
- [3] F. Wu, J. Chen, and J. Wang, “Development of dust sampler based flow closed-loop control,” *Acta Metrologica Sinica*, vol. 38, no. 1, pp. 87–90, 2017.
- [4] K. J. Candra, S. A. Pulung, and M. A. Sadashiv, “Dust dispersion and management in underground mining faces,” *International Journal of Mining Science and Technology*, vol. 24, no. 1, pp. 39–44, 2014.
- [5] R. Tong, M. Cheng, X. Yang, Y. Yang, and M. Shi, “Exposure levels and health damage assessment of dust in a coal mine of

- Shanxi Province, China,” *Process Safety and Environmental Protection*, vol. 128, pp. 184–192, 2019.
- [6] S. Li, B. Xie, S. Hu et al., “Removal of dust produced in the roadway of coal mine using a mining dust filtration system,” *Advanced Powder Technology*, vol. 30, no. 5, pp. 911–919, 2019.
  - [7] G. Zhou, W. Yin, and B. Feng, “Numerical simulation on the distribution characteristics of dust-droplet field during support movement in a fully-mechanized mining face and related engineering applications,” *Journal of China Coal Society*, vol. 43, no. 12, pp. 3425–3435, 2018.
  - [8] Y. Niu, L. Zhang, and B. Shi, “Experimental study on the explosion-propagation law of coal dust with different moisture contents induced by methane explosion,” *Powder Technology*, vol. 361, pp. 507–511, 2020.
  - [9] M. M. Biliaiev, V. V. Biliaieva, V. A. Kozachyna, O. V. Berlov, M. O. Oladipo, and P. S. Kirichenko, “Reducing of coal dust release from train wagon with barrier,” *IOP Conference Series: Materials Science and Engineering*, vol. 985, p. 12018, 2020.
  - [10] M. Tousif, A. Harish, S. Muthu Kumaran, and V. Raghavan, “Numerical study of interaction of coal dust with premixed fuel-lean methane-air flames,” *Advanced Powder Technology*, vol. 31, no. 9, pp. 3833–3844, 2020.
  - [11] L. Mao, L. Peng, and H. Wang, “Consensus of Chinese experts on pneumoconiosis treatment,” *Environment and Occupational Medicine*, vol. 35, no. 8, pp. 677–689, 2018.
  - [12] Y. Lin, Y. Ge, W. Shi et al., “Review of current research on occupational disease prevention in China,” *Chinese Health Law*, vol. 3, no. 4, pp. 20–22, 2010.
  - [13] L. Dong, X. Tong, X. Li, J. Zhou, S. Wang, and B. Liu, “Some developments and new insights of environmental problems and deep mining strategy for cleaner production in mines,” *Journal of Cleaner Production*, vol. 210, no. 10, pp. 1562–1578, 2019.
  - [14] Q. Guo, W. Ren, and J. Shi, “Foam for coal dust suppression during underground coal mine tunneling,” *Tunnelling and Underground Space Technology*, vol. 89, pp. 170–178, 2019.
  - [15] J. L. Perret, S. Miles, F. Brims et al., “Respiratory surveillance for coal mine dust and artificial stone exposed workers in Australia and New Zealand: a position statement from the Thoracic Society of Australia and New Zealand,” *Respirology*, vol. 25, no. 11, pp. 1193–1202, 2020.
  - [16] C. E. Seaman, M. R. Shahan, T. W. Beck, and S. E. Mischler, “Design of a water curtain to reduce accumulations of float coal dust in longwall returns,” *International Journal of Mining Science and Technology*, vol. 30, no. 4, pp. 443–447, 2020.
  - [17] J. Wang, W. Zhao, J. Zhao, B. Liu, and L. Song, “Effect of water quality on wettability of coking coal dust,” *IOP Conference Series: Earth and Environmental Science*, vol. 545, p. 12032, 2020.
  - [18] B. Sun, W. Cheng, J. Wang, and H. Wang, “Effects of turbulent airflow from coal cutting on pollution characteristics of coal dust in fully-mechanized mining face: a case study,” *Journal of Cleaner Production*, vol. 201, pp. 308–324, 2018.
  - [19] C. Tan, Z. Jiang, J. Chen et al., “Numerical simulation of influencing factors of dust migration in fully mechanized coal mining,” *Journal of University of Science and Technology Beijing*, vol. 36, no. 6, pp. 716–721, 2014.
  - [20] X. Yao, G. Lu, K. Xu et al., “Ventilation and dust reduction system of large angle fully mechanized caving face based on fluent,” *Journal of Northeast University: Natural Science Edition*, vol. 35, no. 10, pp. 1497–1501, 2014.
  - [21] N. Wen, W. Wei, M. Xiao, Y. Liu, H. Peng, and Q. Liu, “The effects of ventilation parameters on the migration behaviors of head- on dusts in the heading face,” *Tunnelling and Underground Space Technology*, vol. 70, pp. 400–408, 2017.
  - [22] M. Wu, H. Xiangming, Z. Qian, L. Wei, Z. Yanyun, and H. Zhenglong, “Study on preparation and properties of environmentally-friendly dust suppressant with semi-interpenetrating network structure,” *Journal of Cleaner Production*, vol. 259, p. 120870, 2020.
  - [23] T. Abhishek Kumar, S. Prasad, and S. Ray, “Laboratory investigation on the impact of coal dust deposition on PVPanel performance,” *IOP Conference Series: Materials Science and Engineering*, vol. 993, p. 12088, 2020.
  - [24] Q. Zhao, J. Liu, C. Huang et al., “Characteristics of coal dust deflagration under the atmosphere of methane and their inhibition by coal ash,” *Fuel*, vol. 291, pp. 120–121, 2021.
  - [25] A. Sergei and O. Konstantin, “Mathematical modeling of coal dust screening by means of sieve analysis and coal dust combustion based on new methods of piece-linear function approximation,” *Applied Sciences*, vol. 11, no. 4, p. 1609, 2021.
  - [26] V. K. Kollipara, Y. P. Chugh, and K. Mondal, “Physical, mineralogical and wetting characteristics of dusts from interior basin coal mines,” *International Journal of Coal Geology*, vol. 127, pp. 75–87, 2014.
  - [27] V. A. Arkhipov, D. Y. Paleev, Y. F. Patrakov, and A. S. Usanina, “Coal dust wettability estimation,” *Journal of Mining Science*, vol. 50, no. 3, pp. 587–594, 2014.
  - [28] Q. Li, B. Lin, S. Zhao, and H. Dai, “Surface physical properties and its effects on the wetting behaviors of respirable coal mine dust,” *Powder Technology*, vol. 233, pp. 137–145, 2013.
  - [29] J. Zhou and H. Wang, “Improvement and optimization of spray dust control system for fully mechanized coal face shearer,” *Mechanical and Electrical Engineering*, vol. 32, no. 2, pp. 211–214, 2015.
  - [30] Y. Lin, Y. Ge, W. Shi et al., “Experimental study and numerical simulation on spray characteristics of different nozzle,” *Chinese Health Law*, vol. 35, no. 12, pp. 20–22, 2010.
  - [31] G. Zhou, N. Wen, W. Cheng et al., “Influence regulations analysis of high-pressure atomization dust-settling to dust particle’s microscopic parameters in fully mechanized caving coal face,” *Journal of China Coal Society*, vol. 39, no. 10, pp. 2053–2059, 2014.
  - [32] L. Swanson and O. Langefeld, “Fundamental research in water spray systems for dust control,” *Mining Technology*, vol. 124, no. 2, pp. 78–82, 2015.
  - [33] M. W. Tessum and P. C. Raynor, “Effects of spray surfactant and particle charge on respirable coal dust capture,” *Safety and Health at Work*, vol. 8, no. 3, 2017.
  - [34] G. Zhou, W. Cheng, G. Wang, and X.-F. Cui, “Experiment research of the coupling relationship between dust field and droplet field about fully mechanized and roof caving work-face,” *Journal of China Coal Society*, vol. 35, no. 10, pp. 1660–1664, 2010.

## Research Article

# Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning

Shujun Zhang <sup>1</sup>, Shuhao Xu,<sup>1</sup> Liwei Tan,<sup>1</sup> Hongyan Wang <sup>2</sup> and Jianli Meng<sup>3</sup>

<sup>1</sup>College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China

<sup>2</sup>Qingdao Municipal Hospital Qingdao, Qingdao 266071, China

<sup>3</sup>Qilu Hospital of Shandong University, Qingdao 266035, China

Correspondence should be addressed to Shujun Zhang; zhangsj@qust.edu.cn and Hongyan Wang; whywrs@126.com

Received 8 February 2021; Revised 24 March 2021; Accepted 29 March 2021; Published 10 April 2021

Academic Editor: Han Wang

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

Stroke is a kind of cerebrovascular disease that heavily damages people's life and health. The quantitative analysis of brain MRI images plays an important role in the diagnosis and treatment of stroke. Deep neural networks with massive data learning ability supply a powerful tool for lesion detection. In order to study the property of the stroke lesions and complete intelligent automatic detection, we collaborated with two authoritative hospitals and collected 5,668 brain MRI images of 300 ischemic stroke patients. All the lesion regions in the images were accurately labeled by professional doctors to ensure the authority and effectiveness of the data. Three categories of deep learning object detection networks including Faster R-CNN, YOLOV3, and SSD are applied to implement automatic lesion detection with the best precision of 89.77%. Meanwhile, statistical analysis of the locations, shapes of the lesions, and possible related diseases is conducted with valid conclusions. The research contributes to the intelligent assisted diagnosis and prevention and treatment of ischemic stroke.

## 1. Introduction

As the second most deadly disease in the world, stroke has always been one of the major causes which damage human beings' life and health. It is characterized by high morbidity, disability rate, mortality rate, and recurrence rate and brings a heavy burden to the society and the families of patients. At present, the incidence of stroke shows a rapid increase in low-income groups and younger groups [1]. With the increasing importance of medical images in clinical diagnosis, MRI has become an important basis for the diagnosis and treatment of stroke, especially for ischemic stroke, which is hardly distinguished from CT scans compared with hemorrhagic stroke.

Most hospitals all over the world still rely on doctors for medical image diagnosis, which faces up the following problems: (1) It costs much time and effort for radiologists and doctors to examine the MRI images, and patients usually need to wait for over 24 hours to get the imaging conclusion. This will probably delay the valuable best treatment hours of

stroke and reduce the recovery opportunity of patients. (2) The workload is too heavy. When the doctors get tired, sort of misdiagnosis or missed diagnosis will occur and influence patients' treatment. Therefore, automatic and intelligent computer-aided diagnosis of MRI has become a priority.

The traditional machine learning (ML) method for object detection or analysis requires image denoising, segmentation, manual setting of features, combined with classifiers for recognition. The steps are cumbersome, time-consuming, and laborious, with low accuracy and poor diagnostic effect. Deep learning (DL) is a new research direction in the field of machine learning, which is closer to the original purpose of machine learning-artificial intelligence. It can simulate the hierarchical structure of the human brain, learn the inherent principles and presentation levels of data samples, automatically extract eigenvalues, and have a strong learning ability. In recent years, the practicality of deep learning has greatly improved with the help of high-performance GPU servers and massive datasets. Deep learning technology represented by a convolutional neural

network (CNN) automatically extracts characteristic values from a large number of samples to obtain more advanced abstract features for classification, detection, and segmentation, making it possible for intelligent MRI interpretation.

In order to further study automatic diagnosis and prevention of ischemic stroke, we cooperated with two local Grade III A hospitals and collected 5,668 brain MRI images and their clinical imaging reports from 300 cases, with all the lesion areas accurately labeled by professional neurologists. Three kinds of object detection networks, Faster R-CNN, YOLOv3, and SSD, were designed and implemented to carry out automatic lesion detection on MRI images. Statistical analysis of lesion locations, shapes of lesions, and distributions of suspected diseases are carried out and reliable conclusions are given. The research can supply effective data and methods for automatic diagnosis and prevention of ischemic stroke, which is beneficial to people's life and health.

## 2. Related Work

*2.1. Disease Recognition and Classification.* The medical image lesion detection and auxiliary diagnosis system based on deep learning can extract the advanced features of the lesion in the medical image, and the combination with clinical practice will greatly reduce the workload of doctors. In the computer-based lesion detection methods, the characteristics of body parts or organs are calculated and examined through supervised learning methods or classical image processing techniques (such as filtering and mathematical morphology). Among them, the training data samples used in the machine learning method based on supervised learning need to be provided by professional physicians with comprehensive pathological images and manual annotation.

In the past few decades, deep learning technology has developed rapidly and found wide applications in the medical field. In particular, CNN has been applied to lesion detection, with an accuracy improvement of 13–34% [2]. Sirinukunwattana et al. [3] used a spatially constrained convolutional neural network (SC-CNN) to detect and classify colorectal adenocarcinoma cells. It adopts the neighboring ensemble predictor (NEP) method for classification and recognition, which can achieve higher accuracy compared with the traditional feature classification methods. Dou et al. [4] used 3D CNN to automatically detect cerebral hemorrhage (CMBS) from MR images, which can extract more representative advanced features. Compared with 2D CNN and traditional manual feature extraction, 3D CNN detection accuracy can be as high as 93.16%. Different types of CNN architectures have made great success in diagnosing various diseases. Therefore, the rapid development of deep learning has brought big prospects in the field of medicine. Another important application of deep learning in medical images is lesion recognition. Deep learning can effectively mine useful information from the training data and improve the accuracy and speed of medical diagnosis. Kooi et al. [5] used CNN to identify malignant breast lesions.

*2.2. Lesion Detection Networks.* Lesions detection in medical images belongs to a special kind of object detection task [2]. Before the formal involvement of deep learning in object detection, the traditional detection methods are region selection, feature extraction, classification, and regression, which have two difficult problems to solve. Firstly, the strategy of regional selection has a poor effect and high time complexity. The other is the poor robustness of feature extraction by hand. With the fast development of deep learning, object detection algorithms are mainly divided into two factions, one-stage and two-stage.

The two-stage method, represented by the RNN system, solves the problem in two steps: (1) generating region proposals and using CNN to extract features and (2) putting them in the classifier to classify and correct the position. This series of algorithms are inseparable from region proposal. R-CNN system makes full use of the value of feature maps, in which Faster R-CNN [6] and Mask R-CNN [7] are the good representatives. Compared with the previous network, Faster R-CNN not only improves the detection accuracy but also improves the detection speed. It truly realizes the end-to-end target detection framework and shows excellent performance. Cai et al. [8] used one-stage lesion detection to detect different lesions in CT images. Yap et al. [9] pretrained on the ImageNet dataset and used the prior information of natural images for breast tumor detection. Therefore, this paper first chooses Faster R-CNN as the lesion detection network in brain MRI images of ischemic stroke.

The one-stage method is represented by YOLO and SSD. Regression is performed directly on the predicted target object. Among them, YOLOv3 [10] introduces the FPN structure, and its detection layer is integrated by three-level feature layers. YOLOv3 focuses on solving the problem of small object detection and achieves better performance. Liu et al. [11] adopted simplified VGGNet and multiscale output to SSD, making the network more robust for the processing of target scale. SSD was also used for breast cancer lesion detection and significantly got higher performance than other similar algorithms [12]. Therefore, YOLOv3 and SSD are also selected as two MRI lesion detection networks for ischemic stroke in this paper.

In addition, utilizing multimodal images is also quite common. Ben-Cohen et al. [13] used PET images to help lesion detection in CT scans of the liver. Zhang et al. [14] developed a strategy to detect breast masses from digital tomosynthesis by fine-tuning the model pretrained on mammography datasets. Zhao et al. [15] also used multimodal data for liver tumor detection.

*2.3. Current Problems with Medical Data in Deep Learning.* The current medical image data in deep learning mainly has the following three problems:

- (1) Available medical image datasets are in great demand. The reason why deep learning has such strong expressive ability is that many useful features are extracted from massive data. If the dataset is relatively small, the model learning is not sufficient, the recognition accuracy might be very low.

- (2) The resolution and dimension of medical images need to be improved. At present, the medical images are mainly 2D gray images, which is difficult to distinguish between pulmonary vascular cross-sectional and pulmonary nodules. In addition, medical images are affected by the differences between devices and patients, which brings difficulties to deep learning.
- (3) The authoritative labeling of medical images is lacking. Only if the images are marked by professional doctors, can the dataset be used. However, doctors are busy with clinical work. They need to spend extra time and effort to build valuable datasets.

For the above reasons, we are making effort to build a special ischemic stroke MRI dataset.

### 3. Data Collection and Statistical Analysis

**3.1. Data Collection and Preprocessing.** In order to systematically and deeply study the pathological changes of ischemic stroke, our research team cooperated with two local Grade III A hospitals including Qilu Hospital of Shandong University (Qingdao) and Qingdao Municipal Hospital to collect the brain MRI images of 300 ischemic stroke patients and the corresponding clinical diagnosis reports. The database contained a total of 5,668 DWI sequence images. Furthermore, all lesion images are accurately marked by experienced neurologists.

Currently, there are two main storage formats for object detection labels: VOC format and COCO format. This dataset uses the former format. Neurologists use the LabelImg tool to label the lesions in the images and mark the Bounding Box of the lesions. The annotation is stored in an xml file, which includes not only the width, height, and channels of the images but also the coordinates and category of the bounding boxes. The entire database contains four files, as shown in Table 1.

**3.2. Data Analysis and Visualization.** The images and reports in the database are randomly selected from the clinical data of ischemic stroke in the two hospitals from 2017 to 2019, which can objectively reflect the comprehensive characteristics of ischemic stroke. The images and reports were anonymized before being used by the researchers. The collected MRI data and reports are statistically analyzed and visualized by Excel and Python for further exploring the internal useful information about ischemic stroke. The process of data analysis is divided into tabulation and statistical analysis. The purpose of tabulation is to summarize important medical data of patients for easy browsing, query, and storage. Statistical analysis on the MRI images and clinical imaging reports includes the following five aspects: (1) gender, (2) age, (3) possible related diseases, (4) lesion location, and (5) lesion shape.

**3.2.1. Gender.** Figure 1 shows that the proportion of males and females of the patients is about 9/8, reflecting that males

are more likely to suffer from ischemic stroke and should take more precautions to keep their health. The gender characteristic of stroke in China is that the morbidity of men is higher than women [1]. Men are more likely to suffer from stroke than women [16]. A similar ratio has also appeared in world-wide research of stroke [17]. The above references come to the same conclusion.

**3.2.2. Age.** The second index is the age distribution of the 300 patients. Their average age is 66.527, the minimum is 19, the maximum is 92, and the standard deviation of age is 12.107. In addition, we also calculated the number of cases of different age groups, as shown in Figure 2. Obviously, more than 80% of the patients are older than 50 and the age section between 60 and 70 has the biggest risk. Therefore, the elderly need to be more cautious about ischemic stroke. Their family members should keep a good eye on them and help with their cerebrovascular health.

**3.2.3. Possible Related Diseases.** We conducted data analysis on clinical diagnosis or Conclusion and counted the five most common diseases of stroke patients. The result is shown in Figure 3. Since the database is about ischemic stroke, also known as cerebral infarction, the diagnosis result of cerebral infarction is 100%. The second-highest stroke symptom is senile degeneration of the brain, accounting for approximately 74%. Senile degeneration of the brain is a kind of degenerative changes in the brain. Brain atrophy as a type of senile degeneration of the brain is the most common chronic disease in middle age and older age. The high proportion of senile degeneration of the brain indicates that the elderly are more likely to suffer from stroke. The third place is sinusitis. Among all the 300 stroke patients, 169 patients have sinusitis, which is approximately 56%. Sinusitis is a chronic inflammation of the sinuses. According to the statistical results, we can see that the incidence of the disease is very high. Ethmoiditis is also a kind of sinusitis.

**3.2.4. Lesion Location.** We also analyzed the vocabulary in the MRI description, namely, Finding, and studied the high incidence location and the shape of the lesion in stroke patients. The results of this data analysis will improve the reliability of the prevention, diagnosis, and treatment of ischemic stroke. Finding word segmentation is performed through the Jieba library in natural language processing, and the position and shape word frequency of the lesion are calculated. The Jieba library is an excellent third-party Chinese word segmentation library in Python and is suitable for word frequency statistics. The location distribution of the lesion is displayed as a histogram, indicating high-risk areas, as shown in Figure 4.

From Figure 4, we can see, the most vulnerable pathological areas are basal ganglia, frontal lobe, radiation crown, pons, occipital lobe, parietal lobe, temporal lobe, and thalamus. Among them, the number of lesions in the basal ganglia was the largest, with 61 cases, followed by the frontal lobe with 46 cases. There were 40 cases of radiation crown. In

TABLE 1: The composition of the entire database.

File name	Description
DicomImages	Brain MRI images of ischemic stroke patients (dicom format)
Annotations	xml annotation files used for object detection
ImageSets	Txt files of train set, test set, and validation set
Reports	The clinical diagnosis report of each patient, including finding and conclusion

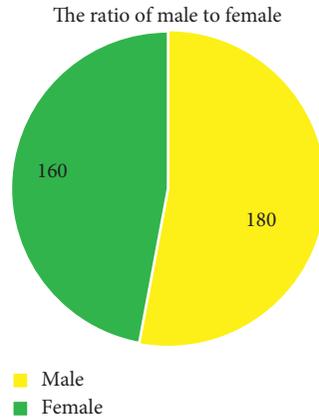


FIGURE 1: The ratio of male to female.

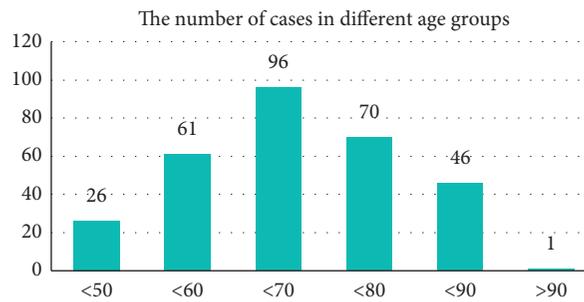


FIGURE 2: Number of cases in different age groups.

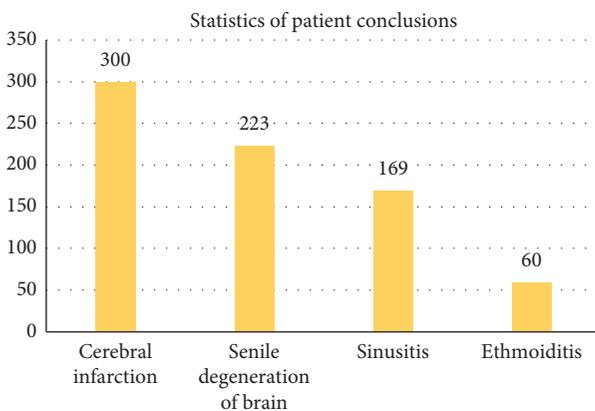


FIGURE 3: Statistics of possible diseases from the imaging reports.

addition, more than 20 cases were pons, occipital lobe, parietal lobe, and temporal lobe. The greater the number of lesions in a brain area, the higher the risk. From this point of

view, the basal ganglia, frontal lobe, radial coronal area, pons, occipital lobe, parietal lobe, and temporal lobe are the most common sites for cerebral infarction. The thalamus and cerebellar hemispheres also have the possibility of stroke. The locations of the brain in the high incidence areas are shown in Figure 5.

3.2.5. *Lesion Shape.* The fifth index is the statistics of the shape of the lesion area. Among them, patchy accounts for about 50% of the total lesion, as shown in Figure 6. Because the shape of the lesion in the diagnosis report is affected by the radiologists' subjective factors, they may offer different shape descriptions for the same lesion.

#### 4. Experimental Results and Analysis

Based on the collected data, automatic lesion detection is implemented using three categories of object detection networks. The experiment is deployed on Ubuntu16. 04 system. The server is equipped with NVIDIA GTX TITAN X

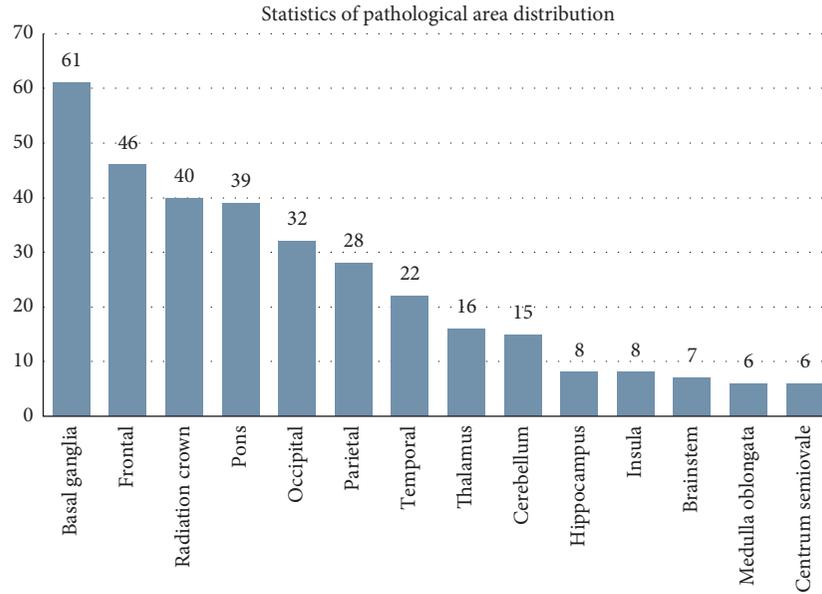


FIGURE 4: Statistics of pathological area distribution.

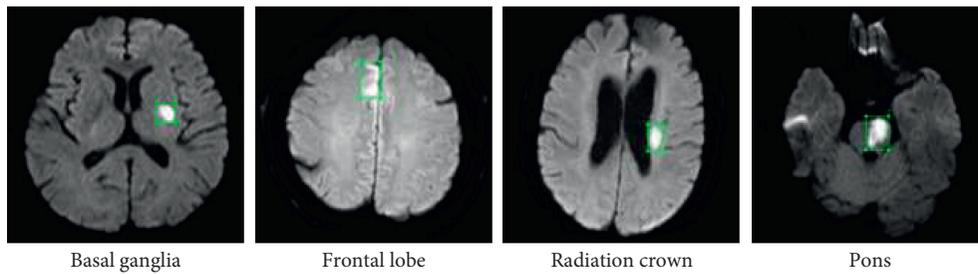


FIGURE 5: Example of areas with a high incidence of lesions.

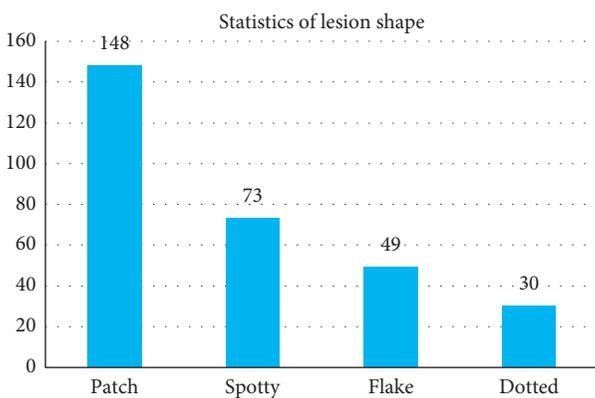


FIGURE 6: Statistics of lesion shape.

and the CPU is Intel Xeon E5-2620 v4. The deep learning framework is PyTorch.

**4.1. Lesion Detection Network Design.** In this paper, three kinds of better-performing target detection networks (Faster R-CNN, YOLOv3, and SSD) are applied to automatically

detect the lesions of ischemic stroke on the collected data. Faster R-CNN may use VGG-16 or ResNet-101 for feature extraction. Therefore, four object detection networks are experimented overall.

The Faster R-CNN algorithm uses a two-stage detection architecture. First, the Region Proposal Network (RPN) is used to generate the Region of Interest (ROI), and then the generated ROI is classified and regressed. The feature extraction network uses the weights of the VGG-16 or ResNet-101 models trained on ImageNet for initialization. The Faster R-CNN (VGG-16) network structure designed and implemented in this paper is shown in Figure 7:

YOLOv3 uses the idea of regression to directly complete object detection using a one-stage network. YOLOv3 brings forth the network structure of DarkNet-53, which draws on the residual idea of ResNet to make the model easier to converge. At the same time, it utilizes multilayer feature maps and does not use a pooling layer. In order to ensure the efficiency of object detection, we used Tiny-DarkNet. YOLOv3 makes multiscale prediction similarly to FPN, replacing the Softmax function with the Logistic function to process the category's prediction score. Logistic classifiers are independent of each other and can achieve multicategory prediction.

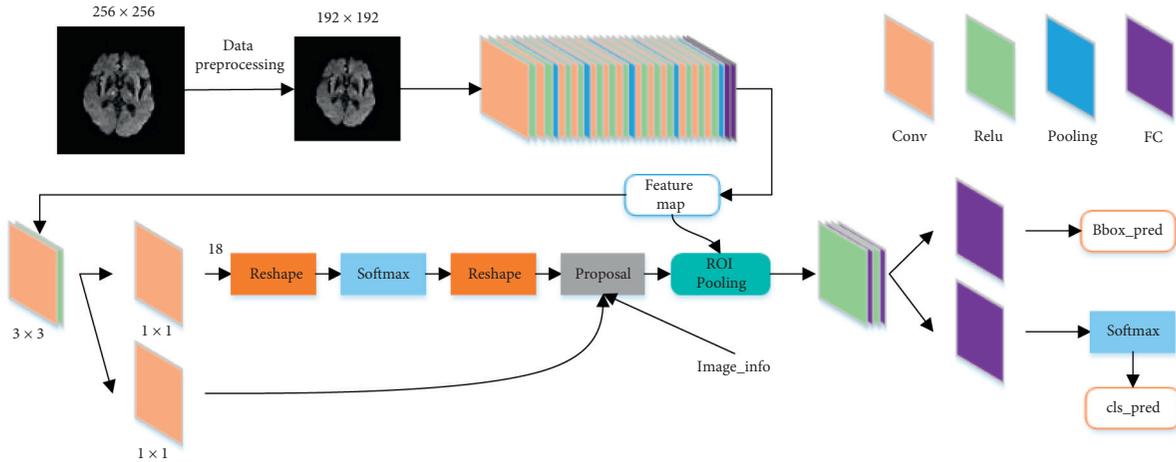


FIGURE 7: Network structure diagram of Faster R-CNN (VGG-16).

Based on the ideas of Faster R-CNN and YOLOv3, SSD utilizes a fixed frame for region generation based on a one-stage network and adopts multilayer feature information. The SSD algorithm also uses VGGNet for feature extraction. On this basis, several convolutional layers are added, and then a  $3 \times 3$  convolution kernel is called to predict on 6 feature layers with different sizes and depths to obtain the classification. The predicted value of regression of the preselection box is then calculated, and the final result is got with the network loss. The difference of SSD from YOLOv3 is that the features of SSD are predicted separately from shallow to deep layers, without depth fusion.

The training parameters of the four object detection networks used in our experiment are shown in Table 2.

#### 4.2. Comparison and Analysis of the Detection Networks.

The entire database is divided into a train set and a test set according to the 4:1 ratio. Since not all images contain lesions, the amount of image data used for object detection is less than 5,668. Totally, there are 1,137 images with lesion labeling. The train set includes 910 images and the test set includes 227 images. Table 3 shows the quantized experimental results.

From Table 3, it can be seen that the detection accuracy (mAP) obtained by Faster R-CNN (VGG-16), Faster R-CNN (ResNet-101), and YOLOv3 networks is similar, while SSD performs the best. This reflects SSD has the best learning ability and adaptability to a variety of lesions of different sizes, types, shapes, and gray levels. And the robustness of the database annotation is also proved.

In addition, we further studied the changes in the accuracy of YOLOv3 and SSD as the number of iterations increases. When the number of YOLOv3 training iterations was 2000, the obtained mAP was 60.1%; when it increased to 20,000, mAP also increased to 74.9%. When the number of the SSD training iterations was 2,000, the mAP was 70.8%. Similarly, when it increases to 20,000, the entire network tends to stabilize, and the maximum mAP is 89.77%. It shows that, as the number of iterations increases, the

detection accuracy of the network also increases. And the SSD increment is larger.

4.3. Visualized Analysis of Lesion Detection Results. The visualized lesion detection results obtained by the above networks on the test set are shown in Figure 8. The first row is the original manual labeling ground truth, and the second to fifth rows are lesion detection results from Faster R-CNN (VGG-16), Faster R-CNN (ResNet-101), SSD, and YOLOv3, respectively. Figures 8(a)–8(g) show different brain MRI image layers. Figure 8 shows that the four detection networks can all perform well as to single clear lesion and large-area lesions, in which SSD is still the best one. For example, in column (g), the left bright line, although similar to a lesion part observed from the gray scale with eyes, is not wrong mistaken by the networks. This bright line is actually an MRI machine scanning trace and easy to cause confusion even for professional doctors. The four networks have obtained accurate detection results, which show the effectiveness of the object detection algorithms on our database.

Figure 9 shows the detection results of the four networks as to the more difficult lesions in the database. Faster R-CNN (VGG-16) and Faster R-CNN (ResNet-101) in Figures 9(a)–9(d) detect some normal regions as lesions, leading to redundant inspections. Faster R-CNN generates multiscale ROIs through RPN, which achieve better results for small target detection and, on the other hand, increase the risk of redundant inspections. In column (e), the two Faster R-CNN networks generate multiple small lesion regions compared with one entire area in GT. This is because some areas have multiple adjacent lesions and it is difficult to separate these lesions. When data labeling is performed, the areas are marked as one lesion. The YOLOv3 network has fewer parameters; thus, the test speed of YOLOv3 is the fastest, reaching 30FPS. By combining the regression idea from YOLO and the anchor mechanism from Faster R-CNN, SSD regresses multiscale regional features and achieves more accurate experimental results for lesions in the

TABLE 2: Four networks training parameter settings.

Parameters	Faster R-CNN (VGG-16)	Faster R-CNN (ResNet-101)	YOLOv3	SSD
Initial learning rate	0.001	0.001	0.0001	0.0001
Learning rate strategy	Step	Step	Multistep	Multistep
Batch size	16	16	32	16
Optimizer	Adam	Adam	SGD	SGD

TABLE 3: Comparison of four network performance.

Networks	mAP_50 (%)	FPS
Faster R-CNN (VGG-16)	76.04	7.5
Faster R-CNN (ResNet-101)	76.5	8.3
YOLOv3	74.9	30
SSD	89.77	27.5

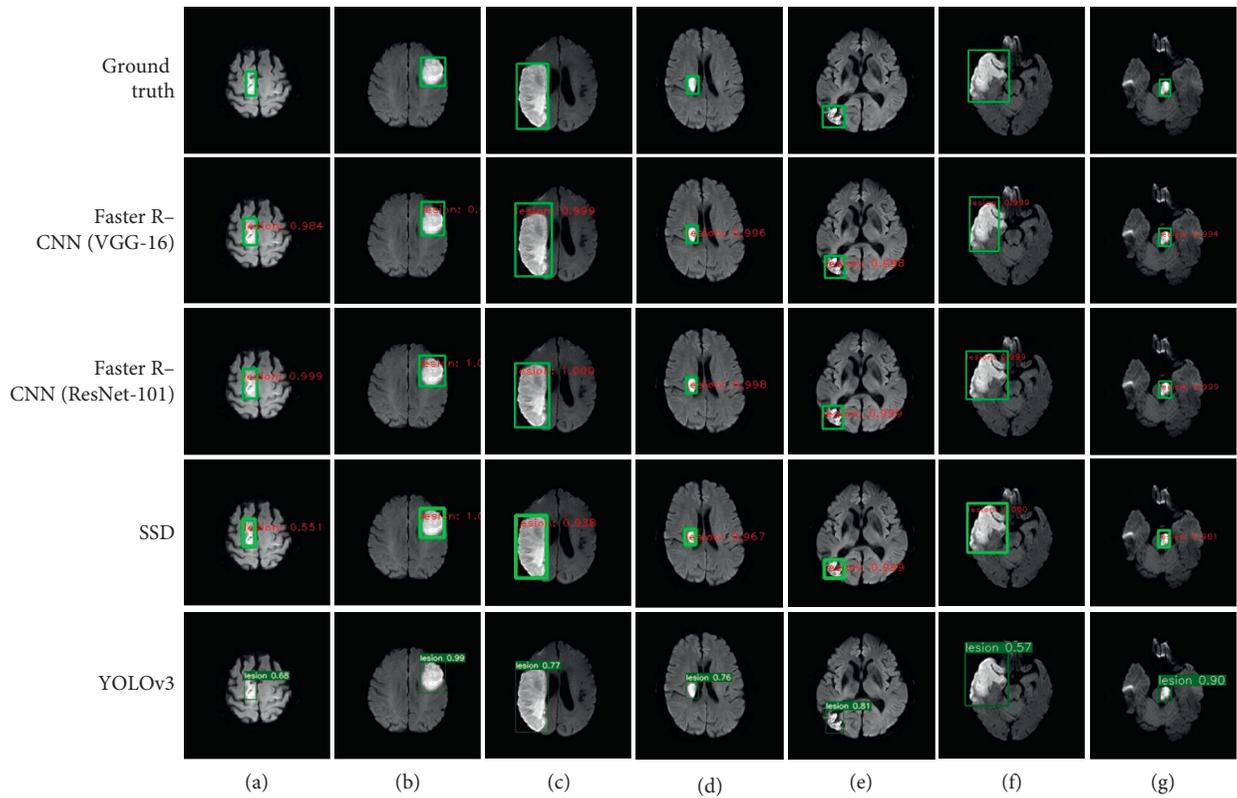


FIGURE 8: Lesion detection results of each brain MRI image layer.

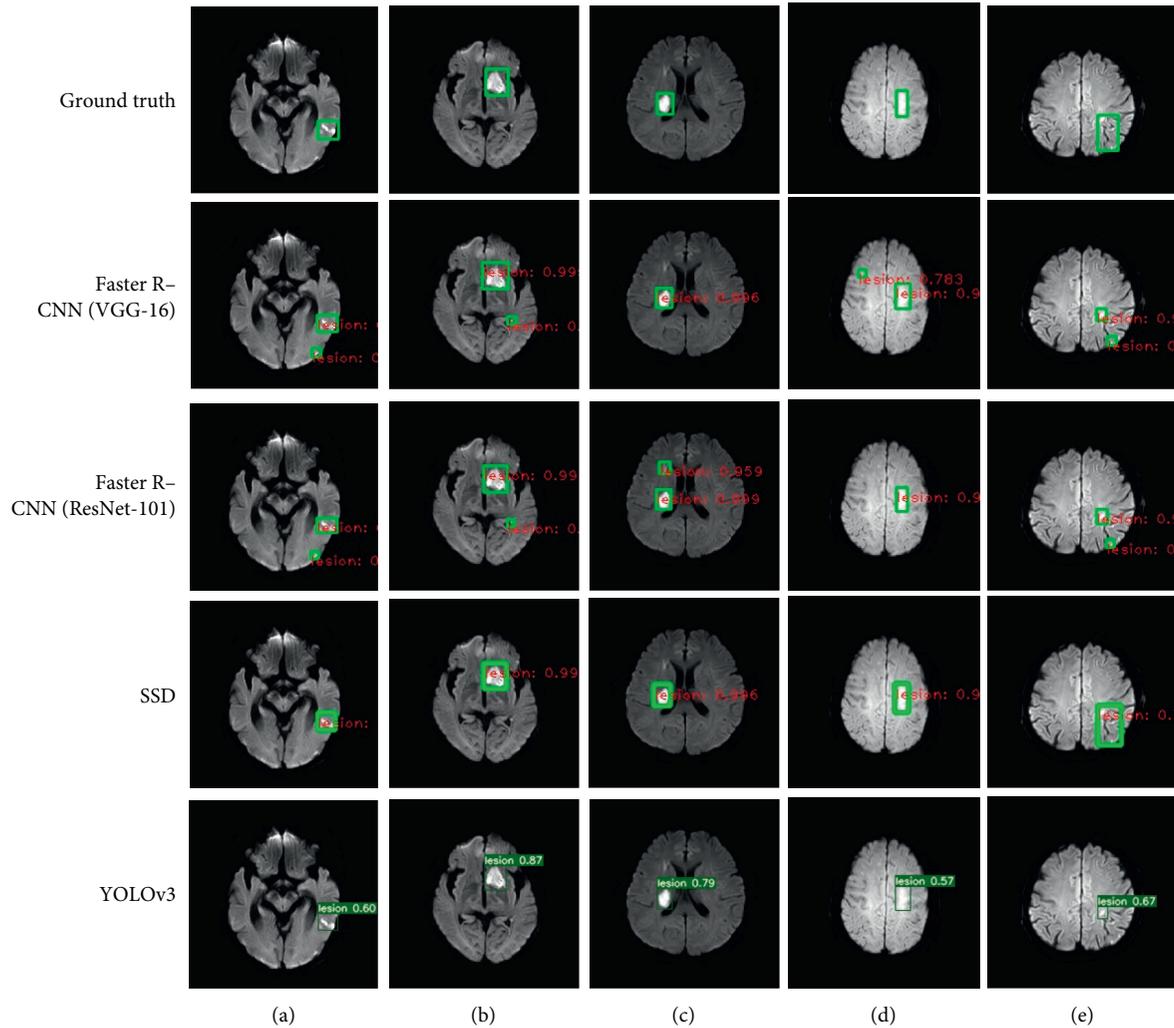


FIGURE 9: Comparison of detection results.

image. Therefore, SSD is more accurate for the detection of ischemic stroke lesions, which also verifies the effectiveness of our database.

## 5. Conclusion

In order to further study the disease of ischemic stroke, we collected image data from two local Grade III A hospitals and completed professional labeling for afterward open research. Statistical analysis is carried out aimed at five indexes and valuable findings and suggestions are given for ischemic stroke precaution and treatment. Deep learning methods including Faster R-CNN, SSD, and YOLOv3 networks are conducted for automatic lesion detection with a precision of 89.77%.

In future, we will continue to collect more data on ischemic stroke and hope to offer a better open research platform. Meanwhile, we may also divide the dataset according to different attributes. The research can improve the intelligence level of computer-aided diagnosis of stroke and promote the development of the theory and practice of artificial intelligence in the medical field.

## Data Availability

The database used to support the findings of this study is still being collected and is expected to be released when the data size reaches 1000 cases.

## Disclosure

Shujun Zhang and Hongyan Wang are co-corresponding authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61702295 and 61672305) and the Key Research and Development Plan Project of Shandong Province (2017GGX10127).

## References

- [1] Report on stroke prevention and treatment in China Writing Group, "Brief report on stroke prevention and treatment in China, 2019," *Chinese Journal of Cerebrovascular Diseases*, vol. 17, no. 5, pp. 272–281, 2020.
- [2] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Medical Image Analysis*, vol. 2021, Article ID 101985, 2021.
- [3] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *Institute of Electrical and Electronics Engineers Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.
- [4] Q. Dou, H. Chen, L. Yu et al., "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *Institute of Electrical and Electronics Engineers Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.
- [5] T. Kooi, G. Litjens, B. Van Ginneken et al., "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, 2017.
- [6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.
- [7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, Venice, Italy, October 2017.
- [8] G. Cai, J. Chen, Z. Wu et al., "One stage lesion detection based on 3D context convolutional neural networks," *Computers and Electrical Engineering*, vol. 79, Article ID 106449, 2019.
- [9] M. H. Yap, G. Pons, J. Marti et al., "Automated breast ultrasound lesions detection using convolutional neural networks," *Institute of Electrical and Electronics Engineers Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2017.
- [10] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, <https://arxiv.org/abs/1804.02767>.
- [11] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: Single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, Springer, Amsterdam, Netherlands, October 2016.
- [12] Z. Chen and S. Huang, "Application of object detection algorithm in imaging diagnosis of breast Cancer," *Modern Computer*, vol. 20, pp. 28–31+37, 2019.
- [13] A. Ben-Cohen, E. Klang, S. P. Raskin et al., "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection," *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 186–194, 2019.
- [14] J. Zhang, E. H. Cain, A. Saha, Z. Zhu, and M. A. Mazurowski, "Breast mass detection in mammography and tomosynthesis via fully convolutional network-based heatmap regression," *Medical Imaging 2018: Computer-Aided Diagnosis. International Society for Optics and Photonics*, vol. 10575, Article ID 1057525, 2018.
- [15] J. Zhao, D. Li, Z. Kassam et al., "Tripartite-GAN: synthesizing liver contrast-enhanced MRI to improve tumor detection," *Medical Image Analysis*, vol. 63, Article ID 101667, 2020.
- [16] L. Wang, Z. Chen, M. Li et al., "Epidemiological investigation and risk factors analysis of cerebral apoplexy in six cities and provinces," *Chinese Medical Record*, vol. 18, no. 3, pp. 97–100, 2017.
- [17] L. Chen, P. Bentley, and D. Rueckert, "Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks," *Neuro Image: Clinical*, vol. 15, pp. 633–643, 2017.

## Research Article

# An Improved Double Channel Long Short-Term Memory Model for Medical Text Classification

Shengbin Liang,<sup>1,2</sup> Xinan Chen ,<sup>2</sup> Jixin Ma ,<sup>3</sup> Wencai Du ,<sup>2</sup> and Huawei Ma <sup>2,4</sup>

<sup>1</sup>School of Software, Henan University, Kaifeng, China

<sup>2</sup>Institute of Data Science, City University of Macau, Taipa, Macau, China

<sup>3</sup>School of Computing and Mathematical Science, The University of Greenwich, London, UK

<sup>4</sup>School of Information Technology, Beijing Institute of Technology, Zhuhai, China

Correspondence should be addressed to Wencai Du; georgedu@cityu.mo and Huawei Ma; mkmhw@bitzh.edu.cn

Received 4 December 2020; Revised 18 January 2021; Accepted 6 February 2021; Published 23 February 2021

Academic Editor: Xingwang Li

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

There are a large number of symptom consultation texts in medical and healthcare Internet communities, and Chinese health segmentation is more complex, which leads to the low accuracy of the existing algorithms for medical text classification. The deep learning model has advantages in extracting abstract features of text effectively. However, for a large number of samples of complex text data, especially for words with ambiguous meanings in the field of Chinese medical diagnosis, the word-level neural network model is insufficient. Therefore, in order to solve the triage and precise treatment of patients, we present an improved Double Channel (DC) mechanism as a significant enhancement to Long Short-Term Memory (LSTM). In this DC mechanism, two channels are used to receive word-level and char-level embedding, respectively, at the same time. Hybrid attention is proposed to combine the current time output with the current time unit state and then using attention to calculate the weight. By calculating the probability distribution of each timestep input data weight, the weight score is obtained, and then weighted summation is performed. At last, the data input by each timestep is subjected to trade-off learning to improve the generalization ability of the model learning. Moreover, we conduct an extensive performance evaluation on two different datasets: cMedQA and Sentiment140. The experimental results show that the DC-LSTM model proposed in this paper has significantly superior accuracy and ROC compared with the basic CNN-LSTM model.

## 1. Introduction

People consult medical experts online in healthcare communities and ask for treatment plans through symptom description or seek the recommended hospital and department. Using a deep learning algorithm to classify disease symptom text can optimize the allocation of medical resources and improve the efficiency of medical treatment. Text categorization is a classic problem in the field of Natural Language Processing (NLP). For an effective medical diagnosis, we proposed the idea of using an improved LSTM model to implement medical consultation text classification. The commonly used classification methods include Naïve Bayes [1], Support Vector Machine (SVM) [2], and Decision

Trees [3]. These classic machine learning classification algorithms have achieved significant results in text classification tasks. With the research and development of neural networks and deep learning, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) have been found to exhibit excellent performance in text classification tasks. At the same time, some models of Chinese question and answer in the medical field have been proposed. Jain and Dodiya [4] have proposed a rule-based framework for the medical question-answering system. Yin et al. [5] designed an algorithm for clustering and similarity evaluation of similar questions and answers for the problem of low efficiency of online healthcare consultation. Feng et al. [6] used CNNs to learn the representation of question and answer

combination and further used it to calculate different questions and candidates. Zhang et al. [7] proposed an end-to-end word embedding multiscale CNN model for question and answer matching in the medical field.

Our primary contribution is a new Double Channel LSTM model, called DC-LSTM, and we add a hybrid attention mechanism to LSTM, which can selectively learn long sequences and make deep neural networks in each batch of training. This proposed model can learn different forms of features, enhance model learning and expression skills, and prevent overfitting. The experimental results show that the DC-LSTM model can significantly improve the accuracy compared with other CNN or RNN models. Particularly in the medical diagnosis classification, using this model can help people quickly choose the right outpatient department for medical treatment and improve the efficiency of outpatient service.

## 2. Related Work

CNN and RNN are two typical deep neural network models. CNN is typically applied in image processing [8] and speech recognition [9], while RNN is usually applied in machine translation and text sequence problem [10]. At the same time, some improved models have been derived. The LSTM proposed by Hochreiter and Schmidhuber in 1997 [11], which is based on the RNN-derived network model, is suitable for processing and predicting important tasks with relatively long intervals and delays in time series.

**2.1. CNN Model.** CNN is a common deep learning network architecture. Hubel and Wiesel are inspired by the natural visual cognitive mechanism of biology. With the improvement of data volume and computing power, CNN has become a research hotspot in recent years. CNN consists of convolution, activation, and pooling layer. Many researchers have proposed some improved CNN models in text classification, such as fastText [12], textCNN [13], and Bi-LSTM [14,15], which are very effective, but they still have some problems such as generality and difficulty in extracting specific context features.

**2.2. Convolutional Layer.** Convolution is the most basic operation in CNN, each convolutional layer is composed of several convolution units, and the parameters of each convolution unit are optimized by a backpropagation algorithm. The purpose of convolution operation is to extract different features of input. More layers of network can extract more complex features from low-level features iteratively. The convolution kernel can scan the input features according to a certain law and multiply the input features by matrix elements in the receptive field and then sum all the results and add the bias. The formula used is as shown in

$$Z^{l+1}(i, j) = [Z^l \otimes w^l](i, j) + b = \sum_{k=1}^{K_l} \sum_{x=1}^f \sum_{y=1}^f Z_k^l \cdot [(s_0 i + x, s_0 j + y) w_k^{l+1}(x, y)] + b, \quad (1)$$

$$(i, j) \in \{0, 1, \dots, L_{l+1}\},$$

$$L_{l+1} = \frac{L_l + 2p - f}{s_0} + 1,$$

where  $b$  is the amount of deviation;  $z^l$  and  $z^{l+1}$  represent the input and output of the  $l + 1$ th convolution layer;  $L^{l+1}$  is the size of  $Z^{l+1}$ ;  $Z(i, j)$  is the corresponding feature matrix;  $f$  is the size of convolution Kernel;  $s_0$  is the step size of convolution; and  $p$  is the padding number.

**2.3. Pooling Layer.** After the feature extraction of the convolutional layer, the output feature matrix is passed to the pooling layer for further feature extraction and information filtering. The pooling layer contains a preset pooling function, which can use the features of its neighboring regions in the feature matrix. The statistics are replaced, and the definition of pooling is as shown in

$$A_k^l(i, j) = \left[ \sum_{x=1}^f \sum_{y=1}^f A_k^l(s_0 i + x, s_0 j + y)^p \right]^{1/p}, \quad (2)$$

where  $s_0$  is the size of the pooling step and  $p$  is the parameter that has been customized specifically. When  $p = 1$ , this is called average pooling, and when  $k \rightarrow \infty$ , this is called maximum pooling.

**2.4. LSTM Model.** LSTM is a special RNN structure that can learn long-term dependencies. RNN can propagate historical information through chained neural network architecture. When processing sequential data, it looks at the current input  $x_t$  and the previous output of the hidden state  $h_{t-1}$  for each time step. However, as the gap between the two-time steps becomes larger, the traditional RNN that can learn the long-term dependency characteristics becomes more difficult. The LSTM proposal addresses this long-term dependency problem and has achieved significant good performance in the statistical machine translation task of Chen et al. [15], making LSTM a successful model. The structure of the LSTM model is shown in Figure 1.

The LSTM architecture provides a series of repeating modules for each time step in a standard RNN. These modules are called cells. At each time step, the output of the module is controlled by a set of gates in  $R_d$  as a function of the old hidden state  $h_{t-1}$  and an input of the current time step  $x_t$  described as follows: forget the gate  $f_t$ , input the gate  $i_t$ , and output the gate  $o_t$ . These gates together determine how to update the current memory unit  $c_t$  and the current hidden state  $h_t$ . We use  $d$  to represent the memory dimension in LSTM, and all vectors in this architecture share the same

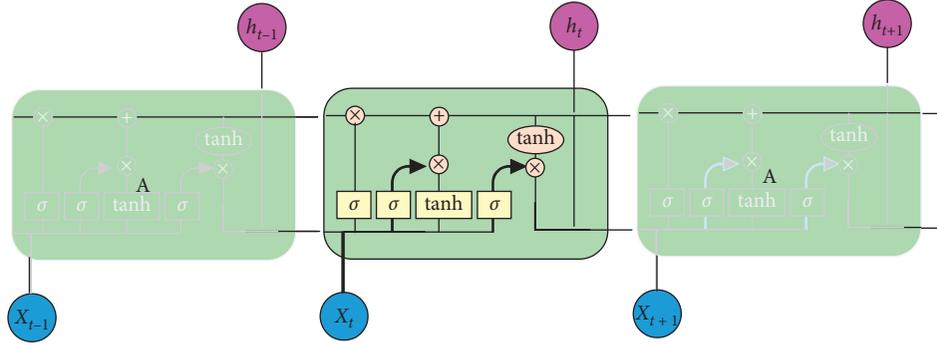


FIGURE 1: Structure of LSTM.

dimension. The LSTM conversion functions are defined as follows:

$$\begin{aligned}
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\
 q_t &= \tanh(W_q \cdot [h_{t-1}, x_t] + b_q), \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\
 c_t &= f_t \cdot c_{t-1} + i_t \cdot q_t, \\
 h_t &= o_t \cdot \tanh(c_t),
 \end{aligned} \tag{3}$$

where  $\sigma$  is a logical sigmoid function whose output is  $[0, 1]$ ,  $\tanh$  represents a hyperbolic tangent function, and its output is  $[-1, 1]$ . LSTM is specifically designed to learn time-series data for long-term dependencies, so we chose LSTM on the convolutional layer to learn this dependency in higher-level feature sequences.

**2.5. Word Embedding.** Glove is also known as Global Vectors for word representation [16]. It is a word representation tool that is count-based and uses overall statistics. A vector of real numbers is obtained that captures some semantic properties between words, such as similarity and analogy. We can calculate the semantic similarity between two words by computing the vector, such as Euclidean distance or cosine similarity.

Word2Vec is a three-layer neural network, which consists of the input layer, the hidden layer, and the softmax layer. The training process is to train the central words and context words by constructing a fake supervised task called Fake Task. The middle-hidden layer weight is used as a trained word vector. According to the size of the input and output data, it has two methods: Skip-Gram method and CBOw method. The former method uses the central word as input and uses the context word of the central word as the label to be predicted for training. The latter method is just the opposite, using the context word as input. As the input data, the context word is trained with the central word as the output to be predicted. After the training is completed, the output information is discarded, and the weight of the middle layer is used as the trained word vector. The Word2Vec model solves the computational bottleneck in the NNLM model. It can easily process tens of millions of text

data and can use variable-length sequences as input. With this advantage, the neural network model can model more complex contexts, and Word vectors can contain richer semantic information.

### 3. Methodology

Traditional CNN and RNN networks often only use word-level embedding, and the semantic features are limited. These traditional models have a very limited capability, especially for words that need to use context to determine semantics. Therefore, it is necessary to expand channels and use multilevel embedding to improve input characteristic diversity. At the same time, the relative importance of each word in the text is different for the modality expressed. Some words contribute more to the modality, some words have less contribution to the modality, and the emotions expressed by each word are also prioritized. Therefore, in order to solve the problem of not being able to selectively learn the emotional characteristics of each word, we can add hybrid attention after LSTM and make trade-offs for different words with different emotions, improve the learning ability of LSTM model, and improve the special characteristics of neural network learning. This creates the ability to simultaneously improve generalization and prevent overfitting from occurring. The model is divided into three parts of CNN-LSTM, Double Channel, and hybrid attention, of which Double Channel is the most important structure.

**3.1. CNN-LSTM.** In our work, we use both CNN and LSTM. The improved CNN structure in our model is similar to ConvNets proposed by Zhang et al. [17] in 2015. The ConvNets consists of nine layers deep with six convolutional layers and three fully connected layers. The structure of the ConvNets is shown in Figure 2.

On the other hand, RNNs have been widely exploited to deal with variable-length sequence input. However, when the length of the input sequence becomes longer, CNNs may suffer from a gradient problem of disappearing or exploding, which will make it more difficult to learn information from a longer time context. LSTM is one of the popular variations of RNN which is proposed to solve this problem. Its network solves this problem by introducing a gate structure in each LSTM unit. The forgetting gate decides what information to

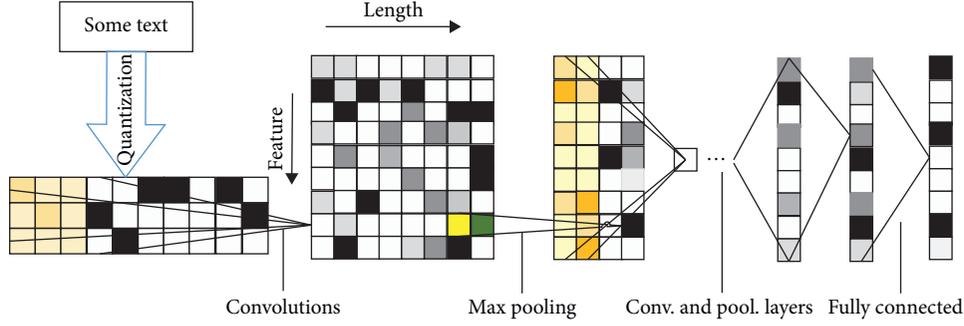


FIGURE 2: Structure of ConvNets model.

be discarded from a cell state and how many new inputs are determined by the input gate. The information is added to the cell state, and the output gate determines what value to be output based on the current state of the cell.

We introduce the Double Channel mechanism in the CNN-LSTM model and input multiple levels of embeddings at the same time to acquire multiple levels of features, in order to solve the problem that the word-level and character-level features cannot be extracted at the same time.

In this model, according to the embedding granularity that is used, the structure is divided into Char-Channel and Word-Channel. The model structure is the same in each channel, which is divided into two parts: CNN and LSTM neural network. In the CNN part, the convolution result  $c$  is calculated for the input sequence  $X$  and the convolution kernel  $K$ ,

$$c = \text{conv}(X, K) + b. \quad (4)$$

For the above LSTM calculation process, in order to simplify the formulation, it is unified as LSTM ( $x$ ). For the convolution neural network and long-term and short-term memory neural network, there are two kinds of structures that can be used: series and parallel. At present, the most commonly used structure is series structure, but there is the information loss phenomenon with this, because information compression and loss will occur in the convolution process. The long-term and short-term memory neural network also receives compressed information and loses most of the time-series characteristics with the series structure. It is unable to give full play to the advantages of LSTM. To solve this problem, the parallel structure is chosen over the series structure and the results are mosaic. At the same time, the structure in each channel is recorded as

$$\text{channel}(x) = [\text{conv}(x) \oplus \text{LSTM}(x)]. \quad (5)$$

From equation (5), we can get the basic description of Char-Channel and Word-Channel. Their input is corpus  $x$ , and the output is  $C_{\text{out}}$  and  $W_{\text{out}}$ , respectively,

$$C_{\text{out}} = \text{channel}_{\text{emb}=V_e}(x), \quad (6)$$

$$W_{\text{out}} = \text{channel}_{\text{emb}=V_g}(x). \quad (7)$$

$V_e$  and  $V_g$  in equations (6) and (7) are the word-level embedding vectors trained and the char-level embedding

vector. We concatenate the results of the two channel outputs as a hidden layer output,

$$h = [c_{\text{out}} \oplus w_{\text{out}}]. \quad (8)$$

Then, the result of the hidden layer is sent to the fully connected layer, and then the softmax layer is used for classification output,

$$\hat{y} = \text{softmax}(\text{dense}(h)). \quad (9)$$

The Double Channel structure is as shown in Figure 3.

**3.2. Hybrid Attention.** The weight score  $\omega$  is an important component of the dynamic adaptive weight, and its calculation method is as shown in equations (10) and (11):

$$e_i = v_a^T \tanh(W_a h_i + b), \quad (10)$$

$$h_i = \langle h'_t : c_t \rangle, \quad (11)$$

where  $h'_t$  is the LSTM output at time  $t$ ,  $c_t$  is the status in LSTM at time  $t$ ,  $h$  is the hidden layer output,  $w_a$  is the random initialization weight matrix,  $v_a$  is the random initialization vector, and  $b$  is the random initialized bias. Next, the score  $\omega$  is calculated as shown in

$$\omega = \frac{\exp(e_i)}{\sum_{k=1}^{T_x} \exp(e_i k)}, \quad (12)$$

where  $x$  is the length of the sequence.

The output vector  $c_i$  weighted by the dynamic adaptive weight, as shown in

$$c_i = \sum_{j=1}^{T_x} \omega \cdot h_j. \quad (13)$$

**3.3. DC-LSTM Overview.** In order to better obtain semantic representation and extract text features, we have designed DC-LSTM using the convolution layer through first embedding the input text sequence, then obtaining the vector representation of these sequences, and finally convolving the sequence using the convolution layer. This model can extract word-level semantic features and the pool, reduce the input data and the output size, and also reduce the risk of

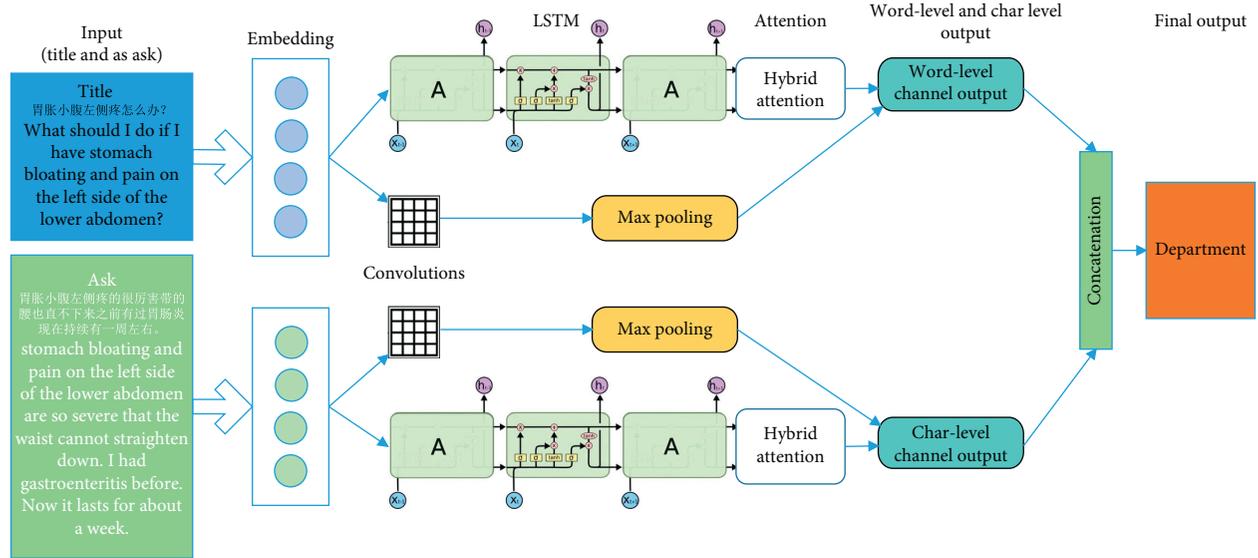


FIGURE 3: The DC-LSTM structure.

overfitting. The data processed by the convolution layer is sent to the LSTM layer, and the LSTM can analyze the timing characteristics in the data. This algorithm can extract some information of context semantics, ignore secondary information, and ensure the accuracy of classification tasks. The DC-LSTM model is shown in Figure 4.

#### 4. Experiments

In order to verify the reliability of the model, a complete contrast experiment was designed. The two datasets of cMedQA and Sentiment140 were used to compare the DC-LSTM model and the basic CNN-LSTM model proposed in this paper was used.

**4.1. Datasets.** To better validate the model's effects, the cMedQA medical diagnosis dataset and the Sentiment140 Twitter dataset are used for verification experiments. The cMedQA dataset is a Chinese text dataset with 792,099 medical consultations which include Andriatria, Internal Medicine, OAGD, Oncology, Pediatric, and Surgical department. The distribution of cMedQA is shown in Figure 5. The question-answering pairs have been preprocessed and classified into different categories. Each pair of QA is encoded as a series.

Meanwhile, to verify the model's good generality, we also select another dataset unrelated to medicine: Sentiment140. Sentiment140 dataset is a tweeter sentiment analysis dataset created and organized by three computer science students from Stanford University, Alec Go, Richa Bhayani, and Lei Huang, with 1.6 million training data and 498 test data. These data are divided into negative, neutral, and positive categories according to emotional polarity. The detailed description of these two datasets is shown in Table 1.

Some samples in cMedQA dataset are as shown in Table 2.

As we can see from Table 2, the dataset includes three main features, namely, department, title, and ask. Title and ask indicate the consultant's symptoms, while department is the answer which indicates the department of treatment. Title points out the core demands of consultants, while ask further describes the content of demands, which puts forward higher requirements for the ability of the model to explain the context.

**4.2. Evaluation.** The evaluation indicators use the accuracy rate, precision rate, recall rate, and  $F_1$ -score to measure the performance of the model.

Define TP is True Positive, FP is False Positive, TN is True Negative, FN is False Negative, and then

$$\begin{aligned} \text{precision} &= \frac{TP}{TP + FP}, \\ \text{recall} &= \frac{TP}{TP + FN}, \\ \text{accuracy} &= \frac{TP + TN}{TP + FP + TN + FN}. \end{aligned} \quad (14)$$

Accuracy refers to the proportion of correctly predicted samples to all samples, precision refers to the proportion of samples that are positively positive, recall refers to the proportion of all positive samples that are correctly predicted, and  $F_1$ -score refers to the harmonic average of precision and recall.

**4.3. Hyperparameter.** It is well known that the quality of hyperparameters will directly affect the training effect of the model, so it is important to choose a series of optimal hyperparameters. The settings of the hyperparameters are shown in Table 3.

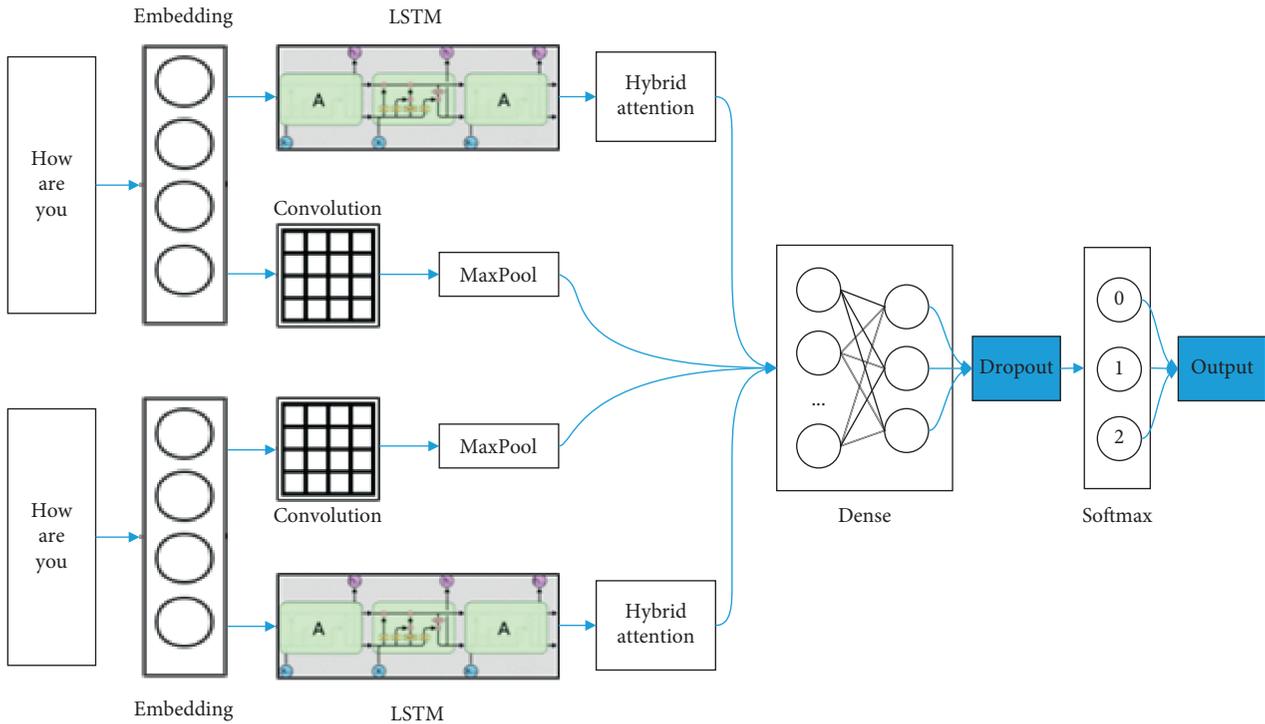


FIGURE 4: DC-LSTM model.

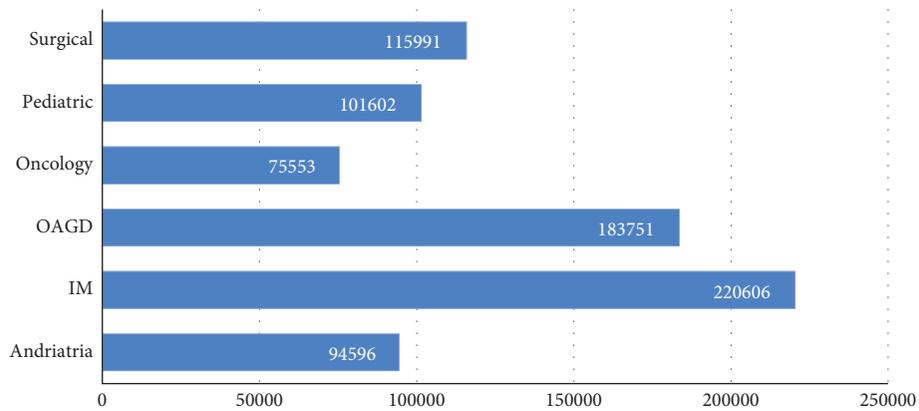


FIGURE 5: Data statistics of cMedQA.

TABLE 1: Overview of datasets.

Datasets	Number of training sets	Number of test sets	Categories	Features
cMedQA	792099	2500	6	3
Sentiment140	1600000	498	3	6

**4.4. Experiment Results.** This section compares experiments and uses our proposed improved model, DC-LSTM, to compare it with CNN, LSTM, CNN-LSTM, and GRU models. The environment used in this paper is based on Tensorflow [18] as the background of Keras as the development verification environment, CUDA [19] as the GPU acceleration environment, and cuDNN [20] as the numerical

computing environment of the deep neural network. The experimental results are shown in Table 4.

It can be seen from Table 4 that the DC-LSTM model proposed in this paper is superior to other models such as CNN, LSTM, CNN-LSTM, and GRU in terms of accuracy, precision, recall, and  $F_1$ -score on cMedQA dataset and Sentiment140 dataset. The model has good generalization

TABLE 2: Some samples in cMedQA.

Department(Answer)	Title	Ask
心血管科 Cardiovascular Department	高血压患者能吃党参吗? Can hypertensive patients take <i>Codonopsis pilosa</i> ?	我有高血压这两天女婿来的时候给我拿了些党参泡水喝, 您好高血压 可以吃党参吗? Hello, I am a hypertensive patient; my son-in-law gave me some <i>Codonopsis pilosa</i> as a gift, can I make tea with <i>Codonopsis pilosa</i> ?
消化科 Digestive System Department	哪家医院能治胃反流 Which hospital can treat gastric reflux?	烧心, 打隔, 咳嗽低烧, 以有4年多 Heartburn, interval, cough, and low fever, more than 4 years

TABLE 3: Hyperparameter.

Items	Values
Word vector	Baidu Baike word + char 300d
Batch size	256
Filter size	3/5/7
Feature map number	150
Activation function	ReLU
LSTM output	128
Full connection layer output	200
Learning rate	Adadelta
Dropout	0.25
Loss	Binary crossentropy
Optimizer	Adam

TABLE 4: The results of the experiment.

Model	Dataset	Accuracy	Precision	Recall	$F_1$ -score
CNN	cMedQA	0.9570	0.8739	0.8803	0.8760
	Sentiment140	0.8831	0.8189	0.8188	0.8188
LSTM	cMedQA	0.9605	0.8857	0.8851	0.8852
	Sentiment140	0.9083	0.8255	0.8267	0.8256
CNN-LSTM	cMedQA	0.9615	0.8892	0.8898	0.8877
	Sentiment140	0.9010	0.8272	0.8271	0.8271
GRU	cMedQA	0.9611	0.8896	0.8795	0.8816
	Sentiment140	0.9044	0.8533	0.8532	0.8532
DC-LSTM	cMedQA	0.9729	0.9184	0.9183	0.9192
	Sentiment140	0.9112	0.8703	0.8703	0.8703

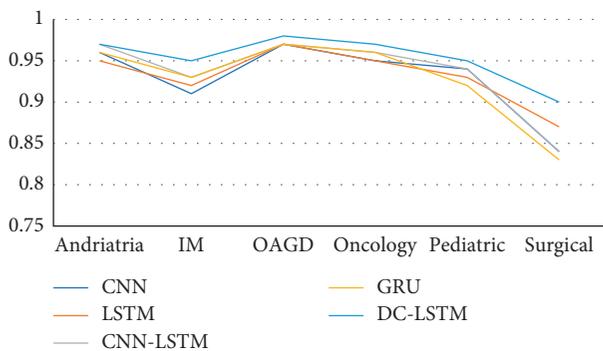


FIGURE 6: AUC value comparison of different models in cMedQA outpatient data.

ability, and the model not only performs well in medical field but also performs better in other datasets. Figure 6 further shows that the AUC value of DC-LSTM model on various

outpatient data of cMedQA is higher than 0.9, which is also significantly higher than other models. In general, the improved model DC-LSTM has been improved in many evaluation indexes. This is due to the introduction of a multichannel mechanism, which can make full use of the attention mechanism's ability to calculate text weight and also make use of the powerful temporal feature learning ability of LSTM. In the channel, learning the semantic information carried by word vector and character vector can learn more features and more fine-grained features.

## 5. Conclusions

In this paper, we find that the basic CNN-LSTM model cannot perform differential learning when dealing with complex long-sequence data. After analyzing the possible causes, we propose an improved method, called DC-LSTM, which incorporates multiplication by weights ( $w$ ) according to each time step of the sequence. The upper bias calculates

the weight score, calculates the probability distribution of the weight score, and adds the hybrid attention according to the probability distribution. Experiments results have shown that DC-LSTM can effectively distinguish the emotional level of different words in sentences and assign different learning weights to different words, so that it can learn the sentiment features of each word in a differentiated way.

## Data Availability

The cMedQA data used to support the findings of this study have been deposited in the GitHub repository (<https://github.com/liangbin/Chinese-medical-dialogue-data>). The Sentiment140 data used to support the findings of this study have been deposited in the Kaggle website (<https://www.kaggle.com/kazanova/sentiment140>).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was financially supported by the National Key Research Plan, Ministry of Science and Technology of China (YFE0101000), the Foundation of Macau (nos. MF1908 and MF1809), and the Key Scientific Research Projects of Colleges and Universities in Henan Province (no. 20A520008).

## References

- [1] E. Abdulhay, V. S. Balaji, and K. Narasimhan, "Automated diagnosis of epilepsy from EEG signals using ensemble learning approach," *Pattern Recognition Letters*, vol. 139, pp. 174–181, 2020.
- [2] G. McDonald, N. García-Pedrajas, C. Macdonald, and I. Ounis, "A study of SVM kernel functions for sensitivity classification ensembles with POS sequences," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 1097–1100, Tokyo, Japan, August 2017.
- [3] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, 2014.
- [4] S. Jain and T. Dodiya, "Rule based architecture for medical question answering system," in *Proceedings of the Second International Conference on Soft Computing for Problem Solving (Soc ProS 2012)*, pp. 1225–1233, Springer, Berlin, Germany, 2012.
- [5] Y. Yin, Y. Zhang, X. Liu, Y. Zhang, C. Xing, and H. Chen, "Health QA: a Chinese QA summary system for smart health," in *Proceedings of the 2nd International Conference on Smart Health (CSH 2014)*, pp. 51–62, Beijing, China, July 2014.
- [6] M. W. Feng, B. Xiang, M. R. Glass, L. Wang, and B. Zhou, "Applying deep learning to answer selection: a study and an open task," in *Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 813–820, Scottsdale, AZ, USA, December 2015.
- [7] S. Zhang, X. Zhang, H. Wang, J. Cheng, P. Li, and Z. Ding, "Chinese medical question answer matching using end-to-end character-level multi-scale CNNs," *Applied Science*, vol. 7, no. 8, pp. 1–17, 2017.
- [8] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater image processing and object detection based on deep CNN method," *Journal of Sensors*, vol. 2020, Article ID 6707328, 20 pages, 2020.
- [9] M. Seo and M. Kim, "Fusing visual attention CNN and bag of visual words for cross-corpus speech emotion recognition," *Sensors*, vol. 20, no. 19, p. 5559, 2020.
- [10] S. Lai, L. Xu, and K. Liu, "Recurrent convolutional neural networks for text classification," *P*, in *proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2015)*, pp. 2267–2273, Austin, TX, USA, January 2015.
- [11] S. Hochreiter and J. Schmidhuber, "LSTM can solve hard long-time lag problems," in *Advances in Neural Information Processing Systems*, pp. 473–479, MIT Press, Cambridge, MA, USA, 1997.
- [12] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 427–431, Valencia, Spain, April 2017.
- [13] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, Doha, Qatar, October 2014.
- [14] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.
- [15] C. W. Chen, S. P. Tseng, T. W. Tang, and J. F. Wang, "Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital," *Information*, vol. 11, no. 2, 2020.
- [16] J. Pennington, R. Socher, and C. Manning, "Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Doha, Qatar, October 2014.
- [17] X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," in *Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, 2015.
- [18] Y. Yu, M. Abadi, and P. Barham, "Dynamic control flow in Large-Scale machine learning," in *Proceedings of the 13th EuroSys Conference (EuroSys)*, Porto, Portugal, April 2018.
- [19] J. Nickolls, I. Buck, and M. Garland, "Scalable parallel programming," in *Proceedings of the IEEE Hot Chips 20 Symposium (HCS)*, pp. 40–53, Stanford, CA, USA, August 2016.
- [20] S. Chetlur, C. Woolley, and P. Vandermersch, "cuDNN: efficient primitives for deep learning," *Computer Science*, vol. 1, no. 10, pp. 4-5, 2014.