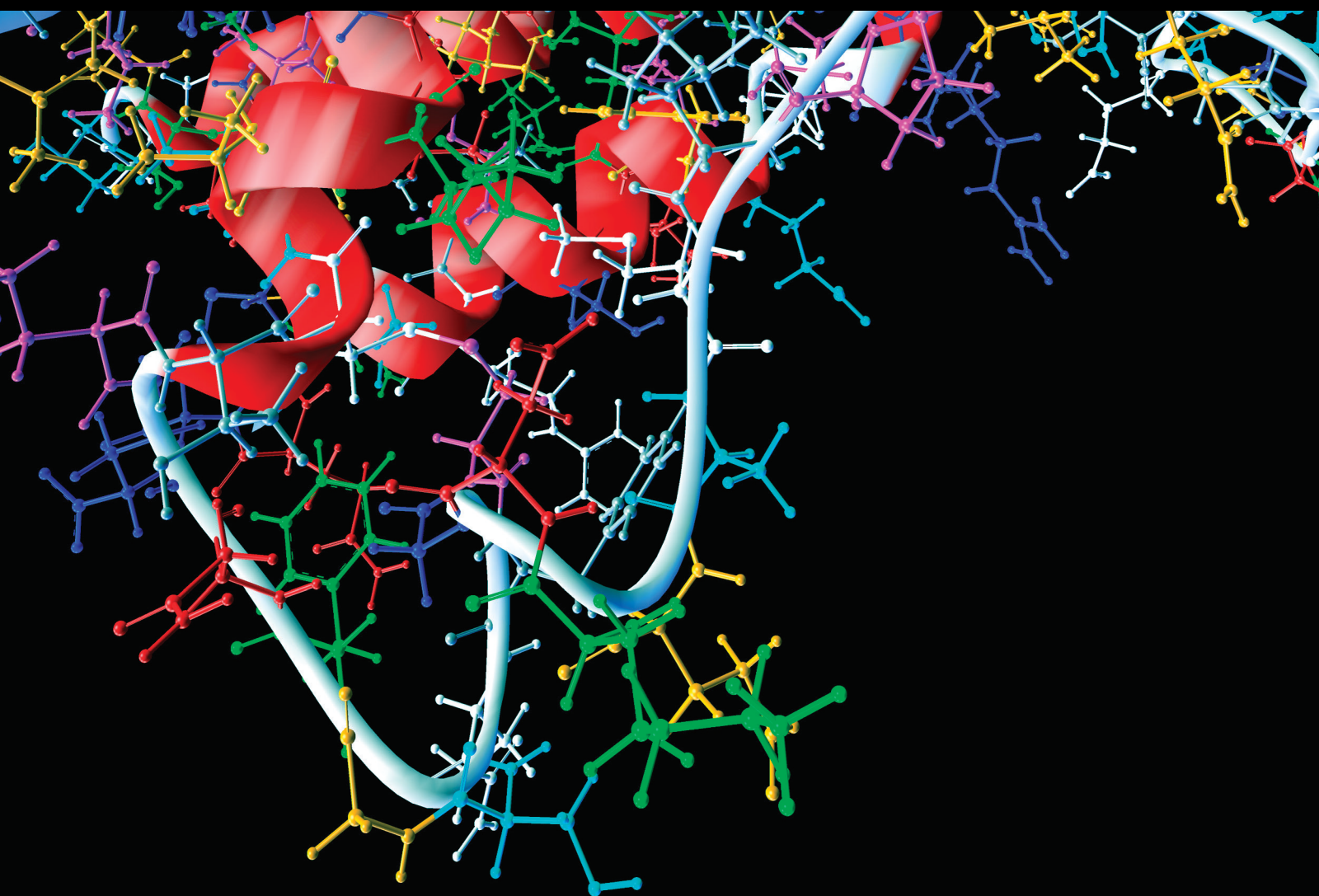



# Computational Intelligence Methods for Brain-Machine Interfacing or Brain-Computer Interfacing

Lead Guest Editor: Yi-Zhang Jiang

Guest Editors: Pritee Khanna, Kaijian Xia, Wen Si, and Chenxi Huang





---

# **Computational Intelligence Methods for Brain-Machine Interfacing or Brain-Computer Interfacing**






**Computational Intelligence Methods  
for Brain-Machine Interfacing or Brain-  
Computer Interfacing**

Lead Guest Editor: Yi-Zhang Jiang

Guest Editors: Pritee Khanna, Kaijian Xia, Wen Si,  
and Chenxi Huang



## Associate Editors

Ahmed Albahri, Iraq  
Konstantin Blyuss , United Kingdom  
Chuangyin Dang, Hong Kong  
Farai Nyabadza , South Africa  
Kathiravan Srinivasan , India

## Academic Editors

Laith Abualigah , Jordan  
Yaser Ahangari Nanekaran , China  
Mubashir Ahmad, Pakistan  
Sultan Ahmad , Saudi Arabia  
Akif Akgul , Turkey  
Karthick Alagar, India  
Shadab Alam, Saudi Arabia  
Raul Alcaraz , Spain  
Emil Alexov, USA  
Enrique Baca-Garcia , Spain  
Sweta Bhattacharya , India  
Junguo Bian, USA  
Elia Biganzoli , Italy  
Antonio Boccaccio, Italy  
Hans A. Braun , Germany  
Zhicheng Cao, China  
Guy Carrault, France  
Sadaruddin Chachar , Pakistan  
Prem Chapagain , USA  
Huiling Chen , China  
Mengxin Chen , China  
Haruna Chiroma, Saudi Arabia  
Watcharaporn Cholanjiak , Thailand  
Maria N. D.S. Cordeiro , Portugal  
Cristiana Corsi , Italy  
Qi Dai , China  
Nagarajan Deivanayagam Pillai, India  
Didier Delignières , France  
Thomas Desaive , Belgium  
David Diller , USA  
Qamar Din, Pakistan  
Irina Doytchinova, Bulgaria  
Sheng Du , China  
D. Easwaramoorthy , India

Esmaeil Ebrahimie , Australia  
Issam El Naqa , USA  
Ilias Elmouki , Morocco  
Angelo Facchiano , Italy  
Luca Faes , Italy  
Maria E. Fantacci , Italy  
Giancarlo Ferrigno , Italy  
Marc Thilo Figge , Germany  
Giulia Fiscon , Italy  
Bapan Ghosh , India  
Igor I. Goryanin, Japan  
Marko Gosak , Slovenia  
Damien Hall, Australia  
Abdulsattar Hamad, Iraq  
Khalid Hattaf , Morocco  
Tingjun Hou , China  
Seiya Imoto , Japan  
Martti Juhola , Finland  
Rajesh Kaluri , India  
Karthick Kanagarathinam, India  
Rafik Karaman , Palestinian Authority  
Chandan Karmakar , Australia  
Kwang Gi Kim , Republic of Korea  
Andrzej Kloczkowski, USA  
Andrei Korobeinikov , China  
Sakthidasan Sankaran Krishnan, India  
Rajesh Kumar, India  
Kuruva Lakshmana , India  
Peng Li , USA  
Chung-Min Liao , Taiwan  
Pinyi Lu , USA  
Reinoud Maex, United Kingdom  
Valeri Makarov , Spain  
Juan Pablo Martínez , Spain  
Richard J. Maude, Thailand  
Zahid Mehmood , Pakistan  
John Mitchell , United Kingdom  
Fazal Ijaz Muhammad , Republic of Korea  
Vishal Nayak , USA  
Tongguang Ni, China  
Michele Nichelatti, Italy  
Kazuhisa Nishizawa , Japan  
Bing Niu , China

Hyuntae Park , Japan  
Jovana Paunovic , Serbia  
Manuel F. G. Penedo , Spain  
Riccardo Pernice , Italy  
Kemal Polat , Turkey  
Alberto Policriti, Italy  
Giuseppe Pontrelli , Italy  
Jesús Poza , Spain  
Maciej Przybyłek , Poland  
Bhanwar Lal Puniya , USA  
Mihai V. Putz , Romania  
Suresh Rasappan, Oman  
Jose Joaquin Rieta , Spain  
Fathalla Rihan , United Arab Emirates  
Sidheswar Routray, India  
Sudipta Roy , India  
Jan Rychtar , USA  
Mario Sansone , Italy  
Murat Sari , Turkey  
Shahzad Sarwar, Saudi Arabia  
Kamal Shah, Saudi Arabia  
Bhisham Sharma , India  
Simon A. Sherman, USA  
Mingsong Shi, China  
Mohammed Shuaib , Malaysia  
Prabhishek Singh , India  
Neelakandan Subramani, India  
Junwei Sun, China  
Yung-Shin Sun , Taiwan  
Min Tang , China  
Hongxun Tao, China  
Alireza Tavakkoli , USA  
João M. Tavares , Portugal  
Jlenia Toppi , Italy  
Anna Tsantili-Kakoulidou , Greece  
Markos G. Tsipouras, North Macedonia  
Po-Hsiang Tsui , Taiwan  
Sathishkumar V E , Republic of Korea  
Durai Raj Vincent P M , India  
Gajendra Kumar Vishwakarma, India  
Liangjiang Wang, USA  
Ruisheng Wang , USA  
Zhouchao Wei, China  
Gabriel Wittum, Germany  
Xiang Wu, China

KI Yanover , Israel  
Xiaojun Yao , China  
Kaan Yetilmezsoy, Turkey  
Hiro Yoshida, USA  
Yuhai Zhao , China


# Contents

## **Retracted: Tap Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy**

Computational and Mathematical Methods in Medicine


Retraction (1 page), Article ID 9898025, Volume 2023 (2023)

## **Research on Ultrasonic Image Recognition Based on Optimization Immune Algorithm**

Xueqiang Zeng and Sufen Chen 



Research Article (11 pages), Article ID 5868949, Volume 2021 (2021)

## **[Retracted] TAP Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy**

Changliang Wang, Yongxue Gu, Xiaojuan Men, Peng Sun, and Meili Chen 


Research Article (9 pages), Article ID 8072126, Volume 2021 (2021)

## **Task Transfer Learning for EEG Classification in Motor Imagery-Based BCI System**

Xuanci Zheng, Jie Li , Hongfei Ji , Lili Duan, Maozhen Li, Zilong Pang, Jie Zhuang, Lu Rongrong, and Gao Tianhao

Research Article (11 pages), Article ID 6056383, Volume 2020 (2020)

## **DRG-Oriented Mathematical Calculation Model and Method of Integrated Medical Service Cost**

Xiaowei Sun and Yi Zhu 





Research Article (9 pages), Article ID 3189676, Volume 2020 (2020)

## **Intelligent Rehabilitation Assistance Tools for Distal Radius Fracture: A Systematic Review Based on Literatures and Mobile Application Stores**

Yalan Chen, Yijun Yu, Xin Lin, Zhenwei Han, Zhe Feng, Xinyi Hua, Dongliang Chen, Xiaotao Xu, Yuanpeng Zhang, and Guheng Wang 

Review Article (9 pages), Article ID 7613569, Volume 2020 (2020)

## **EEG-Based Epilepsy Recognition via Multiple Kernel Learning**

Yufeng Yao , Yan Ding , Shan Zhong , and Zhiming Cui 


Research Article (9 pages), Article ID 7980249, Volume 2020 (2020)

## **Automatic Detection of Coronary Metallic Stent Struts Based on YOLOv3 and R-FCN**

Xiaolu Jiang , Yanqiu Zeng , Shixiao Xiao , Shaojie He, Caizhi Ye, Yu Qi, Jiangsheng Zhao, Dezhi Wei , Muhua Hu , and Fei Chen 

Research Article (10 pages), Article ID 1793517, Volume 2020 (2020)

## **Instance Transfer Subject-Dependent Strategy for Motor Imagery Signal Classification Using Deep Convolutional Neural Networks**

Kai Zhang, Guanghua Xu , Longtin Chen, Peiyuan Tian, ChengCheng Han, Sicong Zhang, and Nan Duan


Research Article (10 pages), Article ID 1683013, Volume 2020 (2020)

### **Design and Development of Virtual Medical System Interface Based on VR-AR Hybrid Technology**

Xu Cong and Tingting Li 

Research Article (7 pages), Article ID 7108147, Volume 2020 (2020)

### **Patient-Specific CT-Based Fluid-Structure-Interaction Aorta Model to Quantify Mechanical Conditions for the Investigation of Ascending Aortic Dilation in TOF Patients**

Heng Zuo , Yunfei Ling, Peng Li, Qi An, and Xiaobo Zhou



Research Article (12 pages), Article ID 4568509, Volume 2020 (2020)

### **DE-CNN: An Improved Identity Recognition Algorithm Based on the Emotional Electroencephalography**

Yingdong Wang , Qingfeng Wu , Chen Wang , and Qunsheng Ruan



Research Article (12 pages), Article ID 7574531, Volume 2020 (2020)

### **An Automatic Epilepsy Detection Method Based on Improved Inductive Transfer Learning**

Yufeng Yao  and Zhiming Cui 





Research Article (13 pages), Article ID 5046315, Volume 2020 (2020)

### **An EEG Database and Its Initial Benchmark Emotion Classification Performance**

Ayan Seal , Puthi Prem Nivesh Reddy, Pingali Chaithanya, Arramada Meghana, Kamireddy Jahnavi, Ondrej Krejcar , and Radovan Hudak

Research Article (14 pages), Article ID 8303465, Volume 2020 (2020)

### **Research and Verification of Convolutional Neural Network Lightweight in BCI**

Shipu Xu , Runlong Li, Yunsheng Wang, Yong Liu, Wenwen Hu, Yingjing Wu, Chenxi Zhang , Chang Liu , and Chao Ma 


Research Article (11 pages), Article ID 5916818, Volume 2020 (2020)

### **An Epilepsy Detection Method Using Multiview Clustering Algorithm and Deep Features**

Qianyi Zhan  and Wei Hu 



Research Article (11 pages), Article ID 5128729, Volume 2020 (2020)

### **Research of Epidemic Big Data Based on Improved Deep Convolutional Neural Network**

Wendong Wang 


Research Article (10 pages), Article ID 3641745, Volume 2020 (2020)

### **Spatial-Frequency Feature Learning and Classification of Motor Imagery EEG Based on Deep Convolution Neural Network**

Minmin Miao , Wenjun Hu, Hongwei Yin, and Ke Zhang 

Research Article (13 pages), Article ID 1981728, Volume 2020 (2020)

### **An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm**


Wentao Wu , Daning Li, Jiaoyang Du, Xiangyu Gao, Wen Gu, Fanfan Zhao, Xiaojie Feng, and Hong Yan

Research Article (10 pages), Article ID 6789306, Volume 2020 (2020)







# Contents

## Multimodal MRI Brain Tumor Image Segmentation Using Sparse Subspace Clustering Algorithm

Li Liu , Liang Kuang, and Yunfeng Ji




Research Article (13 pages), Article ID 8620403, Volume 2020 (2020)

## Factors Influencing Information Service Quality of China Hospital: The Case Study of since 2017 of a Hospital Information Platform in China

Lei Jiao , HuaPing Xiao , XiaoZhuo Zhu , and Xu Zhao 



Research Article (17 pages), Article ID 2089024, Volume 2020 (2020)

## An Indirect Multimodal Image Registration and Completion Method Guided by Image Synthesis

Huan Yang , Pengjiang Qian , and Chao Fan 

Research Article (10 pages), Article ID 2684851, Volume 2020 (2020)

## Attention Optimization Method for EEG via the TGAM

Yu Wu  and Ning Xie 


Research Article (11 pages), Article ID 6427305, Volume 2020 (2020)

## Detection of Solitary Pulmonary Nodules Based on Brain-Computer Interface

Shi Qiu , Junjun Li, Mengdi Cong , Chun Wu, Yan Qin, and Ting Liang 





Research Article (10 pages), Article ID 4930972, Volume 2020 (2020)

## The EEG-Based Attention Analysis in Multimedia m-Learning

Dan Ni , Shuo Wang, and Guocheng Liu


Research Article (10 pages), Article ID 4837291, Volume 2020 (2020)

## A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI

Qi Xiong , Xinman Zhang , Wen-Feng Wang , and Yuhong Gu 



Research Article (10 pages), Article ID 9812019, Volume 2020 (2020)

## EEG Signal and Feature Interaction Modeling-Based Eye Behavior Prediction Research

Pengcheng Ma and Qian Gao 

Research Article (10 pages), Article ID 2801015, Volume 2020 (2020)

## Association between Timing of Surgical Intervention and Mortality in 15,813 Acute Pancreatitis

Lan Lan , Jiawei Luo, Xiaoyan Yang, Dujiang Yang, Mengjiao Li, Fangwei Chen, Nianyin Zeng , and Xiaobo Zhou





Research Article (8 pages), Article ID 1012796, Volume 2020 (2020)

## Fast Enhanced Exemplar-Based Clustering for Incomplete EEG Signals

Anqi Bi, Wenhao Ying, and Lu Zhao 



Research Article (11 pages), Article ID 4147807, Volume 2020 (2020)

## Cross-Subject Seizure Detection in EEGs Using Deep Transfer Learning

Baocan Zhang , Wennen Wang, Yutian Xiao, Shixiao Xiao , Shuaichen Chen, Sirui Chen, Gaowei Xu , and Wenliang Che 




Research Article (8 pages), Article ID 7902072, Volume 2020 (2020)

**A Fast Subpixel Registration Algorithm Based on Single-Step DFT Combined with Phase Correlation Constraint in Multimodality Brain Image**

Jianguo Li  and Quanhai Ma 

Research Article (10 pages), Article ID 9343461, Volume 2020 (2020)

**A Novel Radial Basis Neural Network-Leveraged Fast Training Method for Identifying Organs in MR Images**

Min Xu , Pengjiang Qian , Jiamin Zheng, Hongwei Ge , and Raymond F. Muzic Jr.

Research Article (9 pages), Article ID 4519483, Volume 2020 (2020)

**A Combined Ultrasonic Backscatter Parameter for Bone Status Evaluation in Neonates**

Weiyang Mao , Yang Du , Chengcheng Liu , Boyi Li , Dean Ta , Chao Chen, and Rong Zhang 


Research Article (9 pages), Article ID 3187268, Volume 2020 (2020)

**Epilepsy Detection in EEG Using Grassmann Discriminant Analysis Method**

Hongbin Yu , Chao Fan , and Yunting Zhang


Research Article (14 pages), Article ID 2598140, Volume 2020 (2020)

**Evaluation of Cerebral Blood Flow Dynamics in Transient Ischemic Attacks Patients with Fast Cine Phase Contrast Magnetic Resonance Angiography**

Yuzhao Wang, Duo Gao, and Huaijun Liu 

Research Article (6 pages), Article ID 4097829, Volume 2020 (2020)

**A Modified Skip-Gram Algorithm for Extracting Drug-Drug Interactions from AERS Reports**

Li Wang, Wenjie Pan, Qinghua Wang, Heming Bai, Wei Liu, Lei Jiang, and Yuanpeng Zhang 







Research Article (9 pages), Article ID 1747413, Volume 2020 (2020)

**Evaluating the Acute Effect of Stereoscopic Recovery by Dichoptic Stimulation Using Electroencephalogram**

Wei Shi , Luyang He , Bin Lv , Li Li , and Tongning Wu 

Research Article (7 pages), Article ID 9497369, Volume 2020 (2020)





**A Novel Deep Neural Network for Robust Detection of Seizures Using EEG Signals**

Wei Zhao , Wenbing Zhao , Wenfeng Wang , Xiaolu Jiang , Xiaodong Zhang , Yonghong Peng ,

Baocan Zhang, and Guokai Zhang 

Research Article (9 pages), Article ID 9689821, Volume 2020 (2020)

**Magnetic Resonance Image Denoising Algorithm Based on Cartoon, Texture, and Residual Parts**




Yanqiu Zeng , Baocan Zhang, Wei Zhao , Shixiao Xiao, Guokai Zhang , Haiping Ren , Wenbing

Zhao, Yonghong Peng , Yutian Xiao, Yiwen Lu, Yongshuo Zong, and Yimin Ding

Research Article (10 pages), Article ID 1405647, Volume 2020 (2020)



## Contents

### **Bayesian Estimation of Gumbel Type-II Distribution under Type-II Censoring with Medical Applications**

Kamran Abbas , Zamir Hussain, Noreen Rashid, Amjad Ali , Muhammad Taj, Sajjad Ahmad Khan, Sadaf Manzoor, Umair Khalil, and Dost Muhammad Khan 

Research Article (11 pages), Article ID 1876073, Volume 2020 (2020)

### **Influential Factors of an Asynchronous BCI for Movement Intention Detection**

Sura Rodpongpun , Thapanan Janyalikit, and Chotirat Ann Ratanamahatana 




Research Article (12 pages), Article ID 8573754, Volume 2020 (2020)

### **Research of Low-Rank Representation and Discriminant Correlation Analysis for Alzheimer's Disease Diagnosis**

Zhigang Li, Aimei Dong , and Jing Zhou

Research Article (8 pages), Article ID 5294840, Volume 2020 (2020)

### **Validity Prediction of Amplitude-Integrated EEG in Early Neuromotor Development Outcomes in High-Risk Neonates**

Jian Guo , Wentao Wang , and Qili Zhou 






Research Article (8 pages), Article ID 9438248, Volume 2020 (2020)

### **An Informative and Comprehensive Behavioral Characteristics Analysis Methodology of Android Application for Data Security in Brain-Machine Interfacing**

Xin Su , Qingbo Gong , Yi Zheng , Xuchong Liu , and Kuan-Ching Li 

Research Article (14 pages), Article ID 3658795, Volume 2020 (2020)

### **An Assessment of the Relationship between Structural and Functional Imaging of Cerebrovascular Disease and Cognition-Related Fibers**

Xiaoping Tang , Xinlan Xiao , Jianhua Yin , Ting Yang , and Bingliang Zeng 

Research Article (12 pages), Article ID 4347676, Volume 2020 (2020)

## Retraction

# Retracted: Tap Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy

### Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

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

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

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

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

### References

- [1] C. Wang, Y. Gu, X. Men, P. Sun, and M. Chen, "TAP Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 8072126, 9 pages, 2021.

## Research Article

# Research on Ultrasonic Image Recognition Based on Optimization Immune Algorithm

Xueqiang Zeng<sup>1</sup> and Sufen Chen<sup>2</sup> 

<sup>1</sup>*School of Computer & Information Engineering, Jiangxi Normal University, Nanchang 330022, China*

<sup>2</sup>*School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China*

Correspondence should be addressed to Sufen Chen; csf@nit.edu.cn

Received 23 June 2020; Accepted 29 April 2021; Published 18 May 2021

Academic Editor: Chenxi Huang

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

With the rapid development of science and technology, ultrasound has been paid more and more attention by people, and it is widely used in engineering, diagnosis, and detection. In this paper, an ultrasonic image recognition method based on immune algorithm is proposed for ultrasonic images, and its method is applied to medical ultrasound liver image recognition. Firstly, this paper grays out the ultrasound liver image and selects the region of interest of the image. Secondly, it extracts the feature based on spatial gray matrix independent matrix, spatial frequency decomposition, and fractal features. Then, the immune algorithm is used to classify and identify the normal liver, liver cirrhosis, and liver cancer ultrasound images. Finally, based on the deficiency of the immune algorithm, it is combined with the support vector machine to form an optimized immune algorithm, which improves the performance of ultrasonic liver image classification and recognition. The simulation shows that this paper can effectively classify the normal liver, liver cirrhosis, and liver cancer ultrasound images. Compared with the traditional immune algorithm, this paper combines the immune algorithm with the support vector machine, and the optimized immune algorithm can effectively improve the performance of ultrasonic liver image classification and recognition.

## 1. Introduction

Liver diseases threaten human health at all times. The diagnosis and treatment of liver diseases have always been the focus of medical research. At present, diseases caused by liver diseases have become one of the main causes of human death, including liver cancer, fatty liver, and schistosomiasis. If these diseases can be detected at the same time and treated at the same time, the possibility of a cure can be greatly increased.

One of the methods of examining liver disease is the scanning of medical images. B-ultrasound is one of them, it is one of the most widely used and simple clinical applications, and it has been paid more and more attention. Because the B-ultrasound inspection equipment is relatively cheap, the treatment cost is low, there is no adverse reaction, it can be repeatedly checked, and the penetration rate in China is high. B-ultrasound is to display the echo signal as a two-dimensional

image in the form of a light spot. The size of the echo is represented by the brightness of the light spot. According to the gray level of the light spot, the two-dimensional structure image with a well-defined layer is a grayscale modulation type. B-ultrasound can obtain various cut-off patterns of various organs in the human body and is more suitable for diagnosis of various organ diseases such as liver, gallbladder, kidney, bladder, uterus, and ovary. Since the ambiguity of B-mode images brings certain difficulties to image recognition, improving the correct recognition rate has been a research hot spot at home and abroad. How to effectively extract information from B-mode images and achieve accurate recognition is the key to improving the diagnostic level.

Therefore, many researchers in China have painstakingly studied how to improve the recognition efficiency and accuracy of ultrasound images and reduce the number of repeated treatments and examinations. This topic needs to be solved urgently, and it also has far-reaching significance.

For a long time, in the process of B-ultrasound image recognition, doctors have established diagnostic experience to analyze and judge pathology through a large number of visual observations, which makes the judgment of images have no objective evaluation criteria, so there is a poor repeatability, and the accuracy is different due to the doctor's diagnosis level. In order to better improve the accuracy and reliability of B-mode image recognition, it is urgent to develop the image recognition function of the ultrasonic diagnostic system. This requires a certain processing of the B-mode image in order to obtain certain quantitative parameters, reducing the errors and workloads that occur when doctors use the naked eye to interpret medical images.

At present, the classification and recognition of liver ultrasound images are the focus and difficulty of many scholars at home and abroad. There are many literatures about the classification of prostate and breast images in the classification of ultrasound images; in particular, for the recognition of breast tumors, many scholars have done a lot of research. This provides a good contrast and reference for our work in the field of liver ultrasound image recognition. The recognition and classification of hyperhepatic images that we have studied is a very challenging subject. In this respect, domestic and foreign scholars have also done some exploratory work. Because of its complexity, ultrasound images have many difficulties in classification. Firstly, the image is noisy and the contrast of the lesion is low, which undoubtedly has a great impact on the effect of feature extraction. Secondly, feature selection is also a huge difficulty, and how to find the case to be identified with a high recognition rate of features is also a problem for many researchers.

In this paper, based on immune algorithm, an ultrasonic image recognition method based on immune algorithm is proposed, which is applied to ultrasonic liver image recognition. An optimized immune algorithm is formed by combining immune algorithm with support vector machine. The main contributions of this paper are as follows.

- (1) The ultrasonic image of the liver is grayscale, and the region of interest is selected
- (2) Feature extraction is achieved by spatial gray independent matrix, spatial frequency decomposition, and classification features
- (3) The ultrasonic images of normal liver, liver cirrhosis, and liver cancer were classified and recognized by immune algorithm, and a method of ultrasonic liver image recognition based on immune algorithm was proposed
- (4) An optimized immune algorithm is formed by combining immune algorithm with support vector machine

## 2. Related Work

With the development of information technology, researchers combine digital image processing technology, information technology, and medical diagnosis knowledge. Firstly, the

digital image processing technology is used to process the medical image, and then, the doctor's knowledge of pathological diagnosis is used to make scientific judgment, which can improve the accuracy and reliability of diagnosis [1–3].

The classification of medical images is to find the law of sample distribution from known training samples and to apply this rule to the identification of new samples. A good classifier can not only better explain the known samples but also make better predictions and judgments on new samples and even unobservable phenomena. This is what is commonly referred to as generalization ability.

At present, more and more researchers use optimization algorithms for medical image recognition. Commonly used algorithms are Particle Swarm Optimization (PSO) [4] and Ant Colony Optimization (ACO) [5], artificial bee colony (ABC) algorithm [6], genetic algorithm [7], simulated annealing algorithm, and immune algorithm. The immune algorithm was inspired by the biological immune mechanism. The earliest immunology dates back more than 200 years. In 1974, Jarne's immune network theory was awarded Nobel, and people began to become interested in immune network theory. Edward Jenner discovered the vaccinia vaccine in 1796 and used it to treat the deadly infectious disease of human smallpox. In 1986, Farmer's paper "Immune System, Adaptation and Machine Learning" first developed the immune mechanism and artificial intelligence, and then, the artificial immune system was officially put on the road. In the early 1990s, Bersini added AI ideas to the control problem [8]. In 1990, Ishida introduced AI in the problem of fault diagnosis [9]. After 1999, Castro and Gaspar designed two algorithms based on the principle of clonal selection: clonal selection algorithm and pattern tracking algorithm [10]. In the same year, Dasgupta and other scholars designed a negative selection algorithm using the negative selection mechanism of immune tolerance alone [11]. These algorithms are based on the initial biological immune mechanisms that have emerged in just a few decades, and it can be seen that immunology has great potential to solve problems. More research and application of immunology theory is the development trend of today's high-end field.

An immune algorithm is an algorithm that is based on the principles of immunology and is used in engineering applications. The development of immunology has been continuously advanced through the successive submission of various immune algorithms. Because the immune algorithm is a problem-oriented method, the new algorithm is generated based on the relevant principles of the immune system, or the original technology is improved to make it more suitable for engineering applications. Representative examples include Castro's proposed clonal selection algorithm [12] and artificial immune network algorithm [13] and Forrest's negative selection algorithm [14].

The characteristics of the immune algorithm are as follows: each element has intelligence, has high autonomy, and can judge whether other elements are autologous types; the immune system elements selectively recognize nonself types; and diversity is generated by gene combination. Self-learning is at the network component stage, as long as the new nonself appears, nonself learning [15–17]. Identification



is a passive approach, trying to identify nonself, and communication between units is the affinity between them, not hard links. The immune algorithm simulates the adaptive and artificial immunity of the human immune system, which is a means to strengthen the human immune system. A concentration-based selection update strategy is adopted to prevent the occurrence of premature phenomenon and ensure that the search process is proceeding toward global optimization. The search target of the immune algorithm has a certain degree of dispersion and independence, and the diversity search is realized.

### 3. Proposed Method

**3.1. Ultrasonic Liver Image Preprocessing.** In the process of liver hyperimage recognition, feature extraction is the premise of image recognition. In a broad sense, image feature extraction is a kind of transformation, that is, the low-dimensional space is used to represent the high-dimensional image sample space. Feature selection and extraction are very important, which directly affect the design, performance, and accuracy of image recognition classifier. Its basic task is to find out the most effective features from many features. In this chapter, according to the characteristics of liver ultrasound image, the liver ultrasound image is preprocessed, including transforming the image into gray image and extracting the region of interest (ROI) [18].

**3.1.1. Characteristics of Ultrasound Liver Images.** First, the ultrasonographic features of normal liver, liver cirrhosis, and liver cancer are described as follows.

- (1) Ultrasound images of normal liver showed fine light spots and uniform distribution of liver parenchyma, sometimes sparse, scattered slightly strong light spots and short linear echoes, and the outline was complete, the boundary was neat, and the boundary was clear. The corresponding image features were described as normal liver texture rules, clear, uniform gray distribution, and its texture distribution was dense
- (2) Ultrasound medical images of cirrhosis showed uneven, thin wave, stepped and serrated echoes, and the echo light spots in liver parenchyma were coarser and unevenly distributed
- (3) Ultrasound imaging of liver cancer and low echo sonography of liver parenchyma. The corresponding image features are irregular but fresh, and the range of grayscale distribution increases

In this paper, the description of these features is regarded as the expert experience, combining with the feature descriptors of the region, the feature of never changing moment, the feature of frequency spectrum, the statistical moment feature of texture, and the description of texture based on gray level cooccurrence matrix. The changes of these four kinds of liver tissues were quantified, and the B-ultrasound images were further classified and recognized by these quantized values.

**3.1.2. Grayscale Processing.** Because of insufficient exposure, the gray change range of the whole image is usually very small. If the gray change of the real scene is really small, in order to increase the contrast of all or part of the image, the gray change range can be enlarged by gray transformation. In this way, the original image with small gradient becomes rich in gray level, which improves the visual perception conditions and achieves the purpose of image enhancement.

**3.1.3. Selection of Regions of Interest.** Because the current conditions cannot automatically extract the region of interest, the region of interest extraction in this paper is manually extracted under the guidance of doctors with rich clinical experience. Region of interest extraction follows the following principles [19–21].

- (1) Each image selected two rectangular areas, one located inside the space occupying lesion as A1 and one located in the normal liver as A2
- (2) A1 and A2 are as far as possible to maintain the same depth, for the fan array probe images, as far as possible in the same arc; for the linear array probe images, as far as possible in the same horizontal position. And A2 is closer to A1 as far as possible
- (3) A1 and A2 may vary in size. A1 may include as many space-occupying lesions as possible, but not beyond the scope of space-occupying lesions. A2 may be as large as possible within the allowable scope
- (4) When choosing A2, try to avoid large intrahepatic ducts, such as blood vessels, bile ducts, and gallbladder

**3.2. Ultrasonic Liver Image Feature Extraction.** The entropy, Fourier energy spectrum, angle second-order matrix, and contrast are calculated from the spatial grayscale independent matrix of  $1 \times 1 \text{ cm}^2$  region. The  $M$  band wavelet transform is used to calculate the fractal feature vectors.

The following is the parameter definition:  $f(i, j)$  is the density of pixels in ROI.

- (1) Entropy  $H$

$$H = - \sum_{i=1}^{i-1} \sum_{j=1}^{j-1} P(i, j) \log P(i, j). \quad (1)$$

Entropy represents the degree of nonuniformity or complexity of texture in images. When the complexity is high, the entropy value is large; otherwise, the entropy value is smaller or 0 [22].

- (2) Angular Second Moment (ASM)

Use spatial grayscale independent matrix (cooccurrence matrix).  $C(a, b; d, \theta)$  is a conditional probability function based on second-order estimation, which represents the probability of grayscale from  $a$  to  $b$ ,  $d$  is the sampling of the system, and the direction is specified by angle  $\theta$ . The ASM,

calculated by  $C(a, b; d, 0^\circ)$ , is defined as follows:

$$C(a, b; 1, 0^\circ) = \frac{1}{T(1, 0^\circ)} \text{cardinality} \cdot \{[(k, l), (m, n)] \in \text{ROI} : |k - m| = 1, l = n, f(k, l) = a, f(m, n) = b\}. \quad (2)$$

$T(1, 0^\circ)$  is the total number of pairs of pixels in the ROI, ROI is the sampling interval  $d = 1$ , and the direction angle is  $0^\circ$ . Get ASM:

$$\text{ASM} = \sum_a \sum_b C(a, b; 1, 0^\circ). \quad (3)$$

Second-order moment is a measure of the uniformity of image gray distribution. When the elements of gray level cooccurrence matrix concentrate on the principal diagonal line, the gray distribution of the image is more uniform from the local area; the whole texture is coarse, and the variance is larger, so the second-order moment is larger. When the texture of the image is more fine and the gray distribution is uniform, the energy value is larger. On the contrary, when the gray distribution of the image is very uneven and rough, the energy value is smaller [23].

### (3) Fractal feature vector based on wavelet transform of $M$ band

Pyramid-based wavelet transform provides more important features than ordinary multiresolution analysis. Therefore, the fractal feature vectors are based on ordinary multiresolution analysis and are defined as follows:

$$MF \equiv (D_f^{m,i}, D_f^{m,i+1}, \dots, D_f^{m-n+1,i}, \dots, D_f^{m-n+1,i+k}), \quad (4)$$

where  $D_f$  is the fractal of the  $i$ th subimage at level  $m$ .

For  $M$  band wavelet transform,  $M$  takes different values at different levels. The common multiresolution in formula (4) is implemented by a dual channel and 3-channel filter layer hybrid structure. The fractal dimension of the LH band (low-pass filter along abscissa and high-pass filter along ordinate) provides information for cirrhosis and hepatocellular carcinoma, and the characteristic order is as follows:

$$MF \equiv (D_f^{3,0}, D_f^{2,1}, D_f^{2,2}, D_f^{1,1}, D_f^{1,2}, D_f^{1,3}, D_f^{1,4}, D_f^{1,5}, D_f^{1,6}), \quad (5)$$

where  $D_f^{3,0}$  is the initial image fractal dimension,  $D_f^{2,1}$  is the fractal dimension of the LL image, and  $D_f^{2,2}$  is the fractal dimension of the LH image. The last 6 parts of the eigenvector are calculated from the overdrift diagram.

### 3.3. Immune Algorithm and Support Vector Machine

**3.3.1. Working Principle of Immune Algorithm.** In mathematical theory, immune algorithm is often used to deal with optimization problems (NBP) [24]. Figure 1 gives the basic work flow of the immune algorithm. The steps in the diagram, such

as antigen recognition and parameter determination, coding and initializing antibody groups, and memory cell evolution and recruiting new members, correspond to specific evolutionary processes in the biological immune system. In immunology, these processes are humoral immune responses, and in algorithm, a systematic immune algorithm is formed, that is, the optimization problem is the antigen in immunology. The undetermined solution of NBP is considered as the antibody in the immune process, and the immune algorithm is to solve the optimal solution of NBP.

According to the working principle of immune algorithm, we can get the steps of immune algorithm:

- (1) Identification of antigen: the immune system can get and determine the information of antigen invasion
- (2) Producing the initial parent population: memory cells recognize and bind to antigens, and some of the best antibodies in the initial paternal population (the optimal solution of NBP) are selected from the database
- (3) Computational affinity: the matching degree between antibody and antigen was analyzed
- (4) Memory cell differentiation: the antibody that is most compatible with the antigen produced by (3) is presented to the memory cells, and those antibodies that are less compatible with the antigen will be replaced by the newly produced antibodies that match the antigen
- (5) Antibody promotion and inhibition: the immune system promotes antibodies that are highly compatible with the antigen and inhibits high-density antibodies similar to the antibody
- (6) Antibody renewal: for new antigens not touched by the immune system, the cleared antibodies by (5) will be replaced by lymphocytes, and various types of antibodies can be obtained by crossover mutation operator

The immune algorithm has the following characteristics:

- (1) Solving the diversity of candidates: for different candidate solutions, similar to the least square algorithm, the immune algorithm can obtain a global optimal solution for the problem to be optimized
- (2) Learning and memory: by learning and memory, an optimal global solution of NBP can be obtained quickly, that is, the immune system combines the antibodies with the invading antigens in a more rapid way and clears them away
- (3) Efficient parallel search: immune algorithm and genetic algorithm (GA) are both optimization model algorithms simulating biological physiological mechanism, but their underlying mechanisms are different. The difference between them can be summarized as follows [25, 26]:

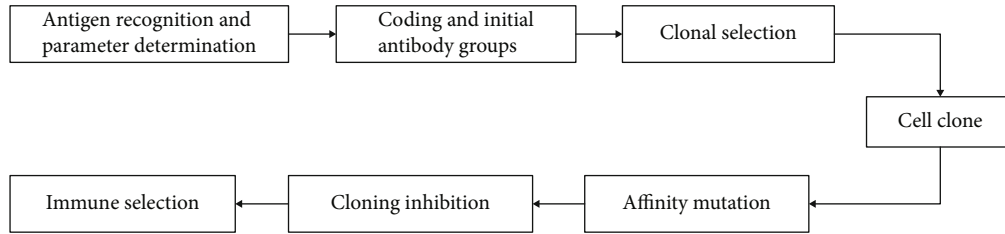


FIGURE 1: Working principle of immune algorithm.

- (1) From a biological point of view, GA is a competition between the evolution of individuals in a population and the heredity of their parents' genes. Immune algorithm is a kind of ability formed by an individual to adapt to the environment, the body to inhibit the invasion of external non-sels or remove their own pathological cells [27]
- (2) In the immune algorithm, antibodies and antigens, antibodies and different antibodies can interact with each other, forming a dynamic network system; GA does not consider the interaction between individuals
- (3) In immune algorithms, the body's genes can be optimized by themselves in some way; in GA, the changes in genes are determined by the environment
- (4) Genes in immune algorithms are carried out within the same organism, and the diversity is produced by the combination of genomes. GA often uses cross-over operator, and the individual gene is the result of the crossover of the previous generation [28]

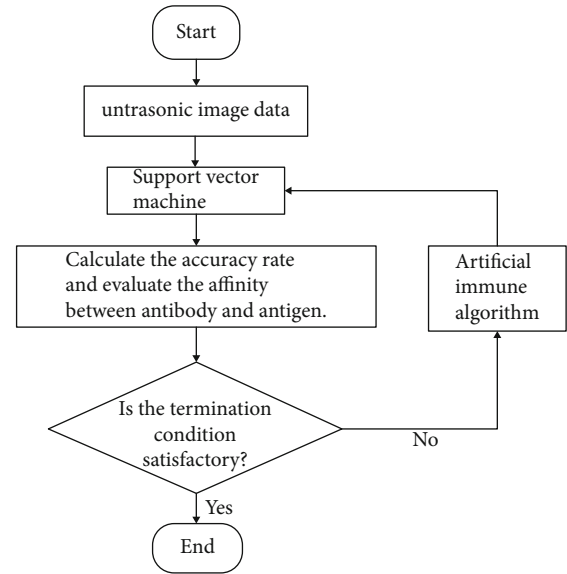


FIGURE 2: Flowchart of optimized immune algorithm.

**3.3.2. Support Vector Machine.** Support vector machine (SVM) was first proposed by V. Vapnik of AT&T Bell Laboratory. It is a machine learning method based on VC dimension theory of statistical learning theory and structural risk minimization principle. It has succinct mathematical form, intuitive geometric interpretation, and good generalization ability. It is a powerful tool to solve classification, regression, probability density estimation, and other issues [29]. It can automatically find the support vectors which have better distinguishing ability to classify, and the classifier constructed from this can maximize the distance between classes, so it has higher classification accuracy. The core content of support vector machine is mainly put forward in the 1990s. At present, the international discussion and further research on this theory are gradually widespread [30].

SVM is a new small sample learning method with solid theoretical foundation. It basically does not involve probability measure and law of large numbers, so it is different from the existing statistical methods. Essentially, it avoids the traditional process from induction to deduction, realizes efficient "transductive reasoning" from training samples to forecast samples, and greatly simplifies the usual classification and regression problems.

The final decision function of SVM is determined only by a few support vectors, and the complexity of calculation depends on the number of support vectors, not the dimension of sample

space, which in a sense avoids the "dimension disaster." A small number of support vectors determine the final result and help to grasp the key samples and "eliminate" a large number of redundant samples. Moreover, this method is not only simple but also has a good "robustness." This "robustness" is mainly reflected in the following: adding or deleting nonsupport vector samples has no effect on the model, and support vector sample set has a certain robustness [31].

The main principles of support vector machines can be roughly summed up as the following two points:

- (1) The corresponding relationship between the independent variables and the strain variables is mapped from the original low-dimensional vector space to the high-dimensional vector space, making it a linear separable state, so that the nonlinear eigenvectors can be linearly analyzed in the high-dimensional feature space by using the linear algorithm
- (2) Based on the structural risk minimization principle, find a hyperplane in the feature space by means of an optimal tool to divide the data and its components into two categories to obtain the optimal classification effect

Simply put, the problem for SVM is to find a hyperplane and effectively separate two different classes of data [32, 33].

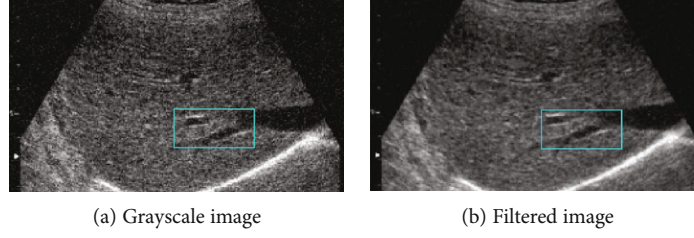


FIGURE 3: Ultrasound liver image preprocessing.

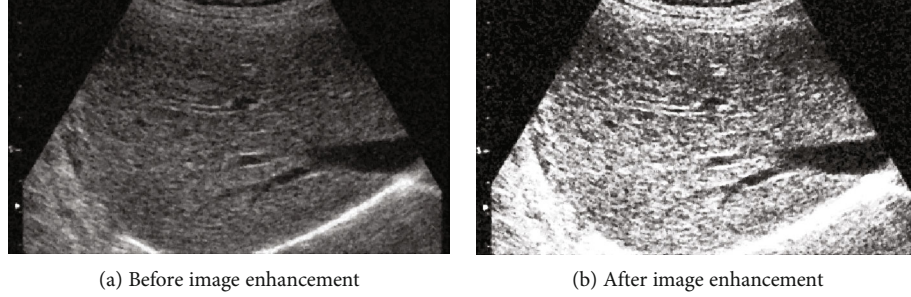


FIGURE 4: Image enhancement processing.

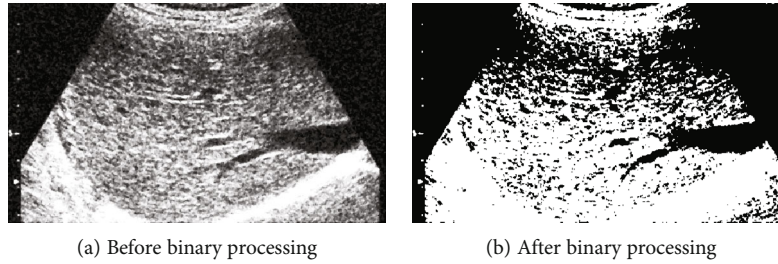


FIGURE 5: Image binarization processing.

**3.4. Ultrasonic Liver Image Classification and Recognition Based on Optimized Immune Algorithm.** In this paper, the most commonly used immune algorithm for learning and optimal computing is combined with SVM. The speed and accuracy of SVM classification are improved by the concepts of replication, mutation, and selection in the processing to preserve the optimal antibody. Finally, an immune algorithm combined with SVM is formed. The basic process of the operation of the immune algorithm is shown in Figure 2.

Firstly, after extracting and calculating the eigenvalues of the ultrasound liver images, the accuracy of SVM is calculated and the affinity between antibodies and antigens is evaluated. If the stopping condition is not satisfied, this paper uses the immune algorithm to filter the features and select the parameters of support vector machine classification and then passes them to support vector machine for analysis and recognition.

Because SVM itself is very sensitive to noise and poor eigenvalues of training samples and if the multiple eigenvalues used in this paper are not properly screened, the accuracy of SVM for liver classification may be reduced. Therefore, this paper combines the immune algorithm and support vector machine to form an optimized immune algorithm which can

improve the shortcomings of support vector machine and improve the accuracy of classification. The principle of this algorithm can be summed up as follows: firstly, the features before classification are brushed and the parameters of support vector machine are selected by immune algorithm so that the parameters  $C$  and  $r$  of radial basis kernel (RBF) can be transformed into higher dimensional space. The algorithm steps can be summarized as follows:

*Step 1 (initial antibody production).* The immune algorithm first produces an array of binary encoded initial antibodies, in which a parameter is selected for each  $n$ -bit. The total length of a group of antibodies is  $32 * n$  bits because each group of initial antibodies needs to select two parameters and thirty feature screening results of the support vector machine.

All the values in the binary initial antibody are selected by random number. After each parameter is generated, the original binary parameter is converted to decimal by the following formula:

$$x_i = d_i + \frac{u_i - d_i}{2^{l_i} - 1} \left( \sum_{j=1}^{l_i} a_j 2^{j-1} \right), \quad (6)$$



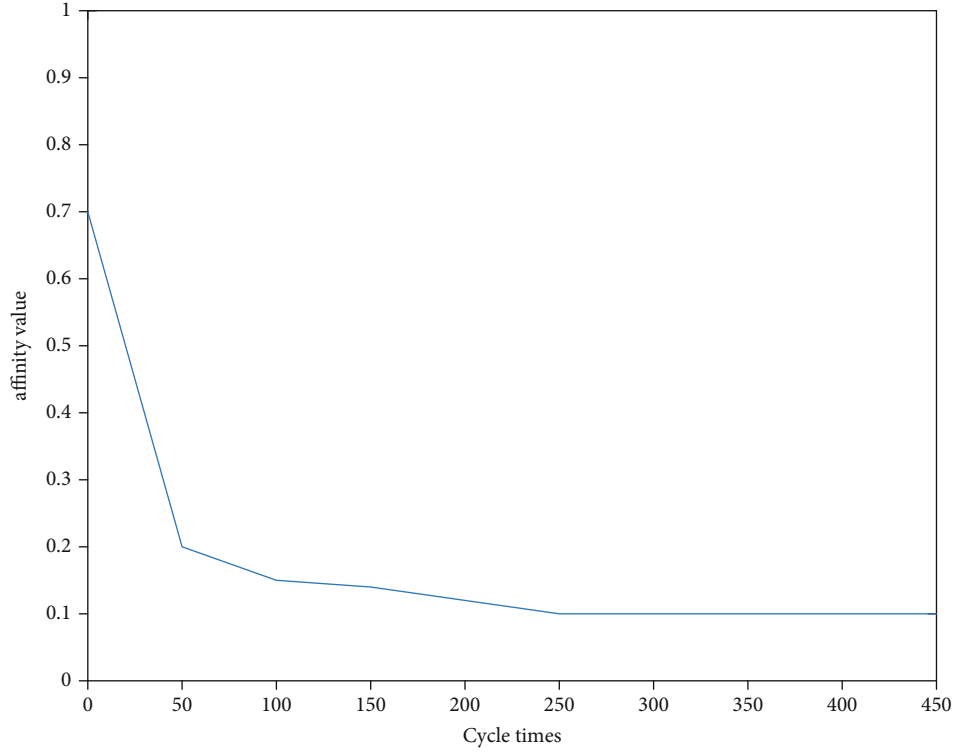


FIGURE 6: Change of affinity value with cycle number.

TABLE 1: Comparison of three algorithms for recognition data.

Characteristic number	Immune algorithm		Support vector machine		Optimized immune algorithm	
	Number of execution	Accuracy	Number of execution	Accuracy	Number of execution	Accuracy
5	405	87.1%	1990	91.2%	250	94.13%
6	424	88.5%	3312	91.9%	277	94.63%
7	456	89.9%	7154	92.5%	312	96.12%

where  $x_i$  is the value of every binary parameter converted to decimal,  $u_i$  and  $d_i$  are the upper and lower limits of the parameter, respectively,  $l_i$  is the length of each binary parameter, and  $a_j$  represents the value of the original binary parameter.

When all the parameters are decimal and then put into the support vector machine for calculation, the feature selection part will be randomly selected according to the given selection ratio. Finally, the recognition accuracy of the liver is regarded as the affinity between the antibody and antigen.

*Step 2* (replication and variation). After the affinity values of the initial antibodies and antigens in the first stage are calculated, the affinity values of these groups are first compared. The greater the affinity (that is, the higher the accuracy), the more likely this group of antibodies will have to replicate and mutate to find the optimal solution. Therefore, the size of antibody replicas with optimal affinity will be higher than that of generic antibodies. All the initial antibodies were duplicated and mutated by binary coding. The mutation rate was related to the coding length of each parameter in the binary initial antibody. The longer the given coding length

was, the smaller the mutation probability was. The replicated and mutated antibodies are then converted to decimal in the same way and put into SVM.

*Step 3* (select the antibody with the best affinity). The replicated and mutated antibodies are then individually compared with their initial antibodies and retain the most affinity antibodies for the next cycle. The cycle of the entire optimized immune algorithm will last until the affinity reaches our desired goal or the number of cycles set.

## 4. Experiments

In the experimental simulation work, the computer hardware configuration is as follows:

- (1) Processor: Intel i5 2.50 GHz
- (2) Memory: 4 GB
- (3) Operating system: Windows 7 64 Ultimate
- (4) The simulation software is Matlab 2016b

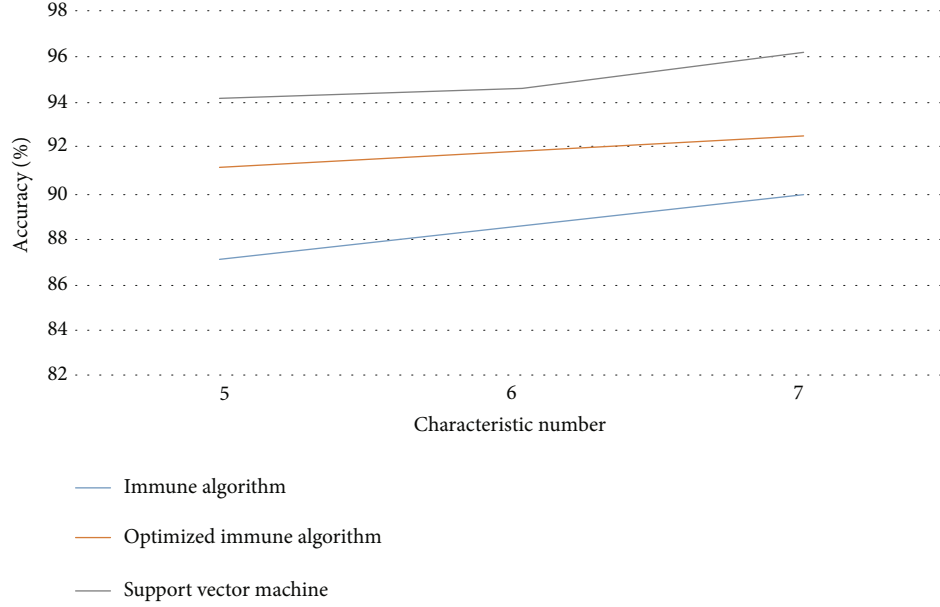


FIGURE 7: A broken line graph for comparing the accuracy of three algorithms.

## 5. Discussion

When extracting features from ultrasonic liver images, the image should be preprocessed in order to extract effective features and prevent the occurrence of false recognition caused by noise and other factors. Figure 3(a) gives the gray-scale image scanned by B-mode ultrasound, and the noise can be clearly seen from Figure 3(a) because of the equipment or other uncontrollable factors. In order to extract the image features better, this paper filters the image of Figure 3(a) and gets the image of Figure 3(b). Comparing the two images, it can be found that the features of Figure 3(b) are more obvious, and noise and other interference are effectively removed.

After filtering, the image is enhanced. The results are shown in Figure 4, in which Figure 4(a) is the result of filtering and Figure 4(b) is the result of image enhancement. From Figure 4, it can be seen that the gray level of the filtered image is relatively concentrated, and after image enhancement, the brightness of the image pixels is high or low, which makes the features of the ultrasound liver image more easily distinguishable.

Then, the image is binarized. The result is shown in Figure 5. Figure 5(a) is the image before binarization, and Figure 5(b) is the image after binarization. As can be seen from Figure 5, after binarization of the image, the data volume of the algorithm is less, the image sensitive area can be highlighted, and the gray level of the whole image is reduced to the binary dimension, which greatly simplifies the subsequent feature extraction algorithm.

In this paper, a total of 300 samples of liver ultrasound images including 100 normal livers, 100 cirrhosis, and 100 hepatocellular carcinomas were randomly divided into five groups. One of the samples was selected by turns for testing, and the other four groups were used as training samples of support vector machine. The parameters of the immune algorithm optimized in this paper are as follows: the initial

TABLE 2: Analysis of three kinds of liver data by three algorithms.

Ultrasound liver image	PCA dimension reduction method	Accuracy	
		Optimized immune algorithm	Support vector machine
Normal	97.9%	100%	98.21%
Cirrhosis	92.11%	94.13%	91.2%
Liver cancer	88.45%	93.34%	89.24%

number of antibodies is 5 groups, the replication rate of general antibodies in each cycle is 10 times, the replication rate of antibodies with optimal affinity is 30 times, and the cycle algebra is 500 times.

In Figure 6, a graph of the affinity of the optimized immune algorithm with algebra is given. From the graph, it can be seen that the affinity value decreases rapidly in the first 50 cycles, which means that the algorithm is searching globally fast to get close to the optimal solution. In 50 to 250 cycles, the affinity value decreases slowly, which means that the algorithm searches locally to find the optimal solution near the optimal solution; after 250 cycles, the affinity value does not change, which indicates that the algorithm has found the optimal solution.

In order to better illustrate the advantages of the algorithm, this paper will use immune algorithm, support vector machine, and the immune algorithm optimized in this paper to identify and analyze the samples; data comparison is shown in Table 1; according to the data in Table 1, the accuracy of the three algorithms compared with the broken line graph is shown in Figure 7.

As can be seen from Table 1, (1) with the increase of the number of features, the recognition accuracy of the three algorithms is improved; (2) the number of execution times of immune algorithm is less than that of support vector



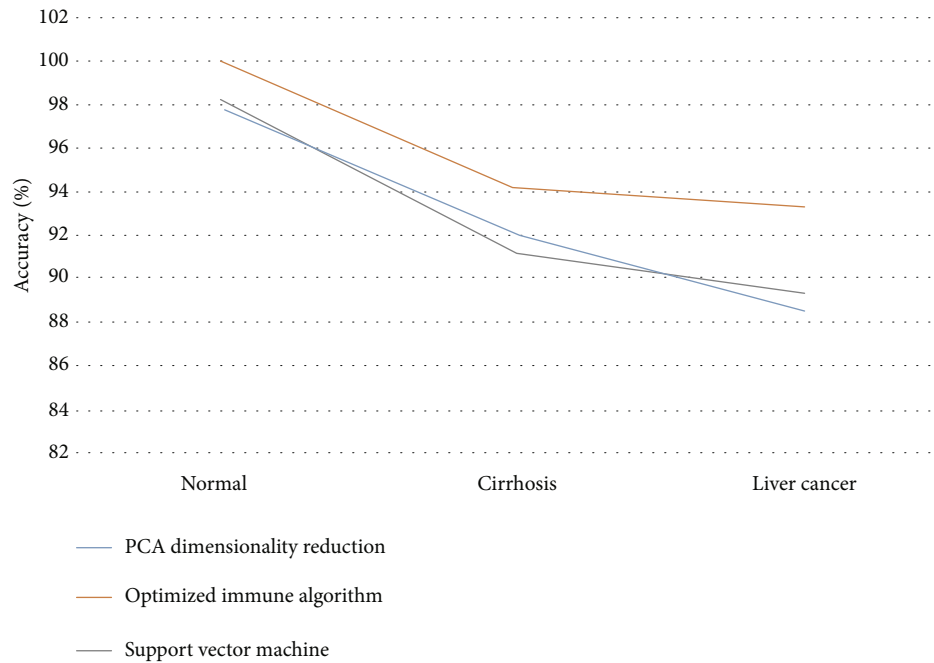


FIGURE 8: Comparison of three algorithms for three liver discrimination accuracy rates.

machine, but its accuracy rate is much lower than that of support vector machine; (3) compared with the immune algorithm, the optimized immune algorithm improves the execution times and the accuracy rate; (4) compared with the support vector machine, the optimized immune algorithm improves the execution times and the accuracy rate. Phenomena (2)–(4) illustrate that the search ability of the optimized immune algorithm is better than that of the immune algorithm and SVM. It shows that the immune algorithm can brush the features before classification and select the parameters of SVM, which can improve the shortcomings of SVM and improve the accuracy of classification.

In addition, the dimensionality of SVM is reduced by filtering the PCA feature number. Therefore, this paper compares the PCA dimensionality reduction method with the immune algorithm proposed in this paper. The comparison data is shown in Table 2, and the accuracy polyline graph is shown in Figure 8.

As can be seen from Table 2, the three algorithms have the highest resolution accuracy for normal liver images, followed by cirrhosis and liver cancer, and the optimized immune algorithm has a higher accuracy than SVM and PCA.

## 6. Conclusions

At present, ultrasound is more and more widely used, such as B-ultrasound in medicine. However, the performance of ultrasonic image recognition still needs to be improved. Therefore, this paper proposes a method of ultrasonic image recognition based on immune algorithm, which is applied to medical ultrasonic liver image recognition. Aiming at the deficiency of classification and recognition of immune algorithm and support vector machine, an optimized immune algorithm is formed to classify and recognize the normal

liver, liver cirrhosis, and liver cancer ultrasonic images. The simulation results show that the optimized immune algorithm can effectively improve the iteration times and accuracy of the immune algorithm and support vector machine, and the recognition accuracy is higher than that of the dimension reduction method of support vector machine by principal component analysis. Similarly, the simulation results show that the optimized immune algorithm performs better in the classification and recognition of normal liver, liver cirrhosis, and liver cancer ultrasound images. In conclusion, the optimized immune algorithm designed in this paper has great potential in the application of medical ultrasound liver image recognition. Because the immune system of an organism is extremely complex and precise, the theory of artificial immune system can be improved and simulated based on various theories of biological immune system. The algorithm used in this paper is based on the replication selection theory of immune system. In addition, there are many methods developed from other concepts of biological immune system, such as negative selection algorithm and artificial immune network, as well as applications combined with other famous algorithms. Therefore, in the follow-up research, we can try to use different artificial immune system theories combined with support vector machine, in order to better improve the operation efficiency of support vector machine and the accuracy of tumor recognition.

## Data Availability

Data can be available on request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the Natural Science Foundation of China under Grant Nos. 61866017 and 61966019, the Support Program for Outstanding Youth Talents in Jiangxi Province no. 20171BCB23013 and in part by the Natural Science Foundation of Jiangxi Province under Grant No. 20192BAB207027.

## References

- [1] D. Nagamalai, E. Renault, and M. Dhanuskodi, "Advances in digital image processing and information technology," *Communications in Computer & Information Science*, vol. 205, 2011.
- [2] G. Q. Yu, C. M. Sui, and B. D. Sui, "Applied-information technology with method of digital image noise processing," *Advanced Materials Research*, vol. 951, pp. 253–256, 2014.
- [3] D. U. Yun-Ming, Q. J. Shi, B. Wang, and S. Q. Tian, "Informationization teaching reform and practice based on student needs: a case study of digital image processing course website constructionsing," *Value Engineering*, 2016.
- [4] Y. Tang and X. Guan, "Parameter estimation for time-delay chaotic system by particle swarm optimization," *Chaos Solitons & Fractals*, vol. 40, no. 3, pp. 1391–1398, 2009.
- [5] E. Amiri, H. Keshavarz, M. Alizadeh, M. Zamani, and T. Khodadadi, "Energy efficient routing in wireless sensor networks based on fuzzy ant colony optimization," *International Journal of Distributed Sensor Networks*, vol. 10, Article ID 768936, 2016.
- [6] Q. K. Pan, "An effective co-evolutionary artificial bee colony algorithm for steelmaking- continuous casting scheduling," *European Journal of Operational Research*, vol. 250, no. 3, pp. 702–714, 2016.
- [7] Q. Long and C. Wu, "A hybrid method combining genetic algorithm and Hooke-Jeeves method for constrained global optimization," *Journal of Industrial & Management Optimization*, vol. 10, no. 4, pp. 1279–1296, 2017.
- [8] S. Senthamarai Kannan and N. Ramaraj, "A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 580–585, 2010.
- [9] T. Ishida, "Parallel, distributed and multi-agent production systems," *Lecture Notes in Computer Science*, vol. 878, pp. 416–422, 1995.
- [10] M. B. Gaspar, M. Castro, and C. C. Monteiro, "Effect of tooth spacing and mesh size on the catch of the Portuguese clam and razor clam dredge," *ICES Journal of Marine Science*, vol. 56, no. 1, pp. 103–110, 1999.
- [11] S. Singh and I. Chana, "Cloud resource provisioning: survey, status and future research directions," *Knowledge & Information Systems*, vol. 49, no. 3, pp. 1005–1069, 2016.
- [12] M. Marinaki and Y. Marinakis, "A hybridization of clonal selection algorithm with iterated local search and variable neighborhood search for the feature selection problem," *Memetic Computing*, vol. 7, no. 3, pp. 181–201, 2015.
- [13] R. J. Kuo, Y. S. Tseng, and Z. Y. Chen, "Integration of fuzzy neural network and artificial immune system-based back-propagation neural network for sales forecasting using qualitative and quantitative data," *Journal of Intelligent Manufacturing*, vol. 27, no. 6, pp. 1191–1207, 2016.
- [14] R. Chikh and S. Chikhi, "Clustered negative selection algorithm and fruit fly optimization for email spam detection," *Journal of Ambient Intelligence & Humanized Computing*, vol. 10, no. 1, pp. 143–152, 2019.
- [15] M. Gong, L. Jiao, H. Du, and L. Bo, "Multiobjective immune algorithm with nondominated neighbor-based selection," *Evolutionary Computation*, vol. 16, no. 2, pp. 225–255, 2008.
- [16] S. Wan and Z. Niu, "A learner oriented learning recommendation approach based on mixed concept mapping and immune algorithm," *Knowledge-Based Systems*, vol. 103, no. C, pp. 28–40, 2016.
- [17] Q. Lin, J. Chen, Z. H. Zhan et al., "A hybrid evolutionary immune algorithm for multiobjective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 5, pp. 711–729, 2015.
- [18] J. L. Brooks, A. Zoumpoulaki, and H. Bowman, "Data-driven region-of-interest selection without inflating type I error rate," *Psychophysiology*, vol. 54, no. 1, pp. 100–113, 2017.
- [19] J. Lv, L. Zhang, and S. Wang, "Region of interest extraction based on saliency detection and contrast analysis for remote sensing images," in *Image and Signal Processing for Remote Sensing XXII. Vol. 10004*, International Society for Optics and Photonics, 2016.
- [20] L. Zhang and Q. Sun, "Saliency detection and region of interest extraction based on multi-image common saliency analysis in satellite images," *Neurocomputing*, vol. 283, pp. 150–165, 2018.
- [21] L. Zhang and S. Wang, "Region-of-interest extraction based on local-global contrast analysis and intra-spectrum information distribution estimation for remote sensing images," *Remote Sensing*, vol. 9, no. 6, p. 597, 2017.
- [22] W. Zhang, J. Yang, Y. Fang, H. Chen, Y. Mao, and M. Kumar, "Analytical fuzzy approach to biological data analysis," *Saudi Journal of Biological Sciences*, vol. 24, no. 3, pp. 563–573, 2017.
- [23] S. Fekri-Ershad, "Pap smear classification using combination of global significant value, texture statistical features and time series features," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 31121–31136, 2019.
- [24] Y. Jin and K. Xia, "Intelligent location and recognition mechanism of abnormal point of medical image based on reliable transmission in medial CT scanner local big data networks," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 3, pp. 609–617, 2018.
- [25] P. Kavitha and S. Thirunavukkarasu, "Cancer recognition in medical image using watershed algorithm," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 12, pp. 3315–3322, 2018.
- [26] Y. H. Shi and Q. Wang, "The artificial intelligence-enabled medical imaging: today and its future," *Chinese Medical Sciences Journal*, vol. 34, no. 2, pp. 71–75, 2019.
- [27] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Systems with Applications*, vol. 57, no. 9, pp. 232–247, 2016.
- [28] A. Onan and S. Korukoglu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *Journal of Information Science*, vol. 43, no. 1, pp. 25–38, 2017.
- [29] A. Onan, "Mining opinions from instructor evaluation reviews: a deep learning approach," *Computer Applications in Engineering Education*, vol. 28, no. 1, pp. 117–138, 2020.
- [30] A. Onan, "An ensemble scheme based on language function analysis and feature engineering for text genre classification," *Journal of Information Science*, vol. 44, no. 1, pp. 28–47, 2018.

- [31] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurrency and Computation: Practice and Experience*, no. article e5909, 2020.
- [32] A. Onan, "Biomedical text categorization based on ensemble pruning and optimized topic modelling," *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 2497471, 22 pages, 2018.
- [33] A. Onan and M. A. Tocoglu, "A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification," *IEEE Access*, vol. 9, pp. 7701–7722, 2021.

## Retraction

# Retracted: Tap Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy

### Computational and Mathematical Methods in Medicine

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

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

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

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

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

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

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

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

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

### References

- [1] C. Wang, Y. Gu, X. Men, P. Sun, and M. Chen, "TAP Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 8072126, 9 pages, 2021.

## Research Article

# TAP Test Image Dynamic Tracking Study after Thyroid Cancer Surgery and after Radiotherapy and Chemotherapy

Changliang Wang,<sup>1</sup> Yongxue Gu,<sup>1</sup> Xiaojuan Men,<sup>1</sup> Peng Sun,<sup>2</sup> and Meili Chen<sup>3</sup> 

<sup>1</sup>Department of Galactophore Surgery, Weifang People's Hospital, Shandong 261041, China

<sup>2</sup>Weifang Medical University, Shandong 261041, China

<sup>3</sup>Weifang Brain Hospital Emergency Department, Shandong 261041, China

Correspondence should be addressed to Meili Chen; 15005367137@163.com

Received 28 May 2020; Revised 10 December 2020; Accepted 22 March 2021; Published 17 April 2021

Academic Editor: Chenxi Huang

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

Thyroid cancer is a relatively common endocrine gland malignant tumor; if improper treatment, there will be a high risk of recurrence or metastasis, and abnormal sugar chain glycoprotein (TAP) has a close relationship with the development of the disease; therefore, the purpose of this article is to discuss abnormal sugar chain glycoprotein (TAP) as thyroid cancer curative effect evaluation and radiation and chemotherapy after surgery clinical significance. In this paper, 95 patients with thyroid cancer diagnosed in a hospital were selected as the study objects and treated as the observation group. The clinical and follow-up data of the observation group were retrospectively analyzed. Meanwhile, 55 healthy patients were randomly selected as the control group. TAP, squamous cell carcinoma antigen (SCC) level, and carcinoembryonic antigen (CEA) level were detected in peripheral blood of 95 patients with thyroid cancer before and after treatment. The short-term efficacy was evaluated by chest CT examination, and the changes of the three markers before and after treatment and the correlation with the short-term efficacy of the patients were compared. According to the results of testing, the TAP positive expression in patients before radiotherapy can better predict the recent curative effect has certain clinical value; before radiotherapy TAP positive expression rate was significantly higher than that of healthy people, TAP positive expression quantity decreased obviously after radiation treatment, and patients with a recent radiotherapy curative effect is good or bad and negatively correlated with the degree of TAP protein positive expression; TAP high protein in patients with recent poor radiation effects, prompt the factor can be predicted in the near future curative effect of the molecular markers, and can TAP level for clinicians provide certain reference for targeted therapy.

## 1. Introduction

The incidence of thyroid cancer is increasing year by year, and the age of onset is getting younger [1, 2], which is a relatively common malignant tumor [3]. Early thyroid cancer with no obvious symptoms or signs was mostly found in the form of thyroid nodules in physical examination [4]. Current clinical though an ultrasound examination and puncture pumping fine needle biopsy detection methods, such as [5, 6], but some patients still cannot determine the nature of the lesion and the need to closely follow up even choose surgical resection in order to avoid delay in diagnosis or treatment,

which has increased the psychological burden of patients and physical trauma, so looking for tumor markers to assist in the diagnosis of benign and malignant thyroid nodule has important significance. Abnormal glucose-chain glycoprotein (TAP) is an abnormal complex that occurs in the malignant transformation stage of cells and is expressed after mutations of oncogenes and tumor suppressor genes. When cancer cells proliferate abnormally, TAP is released into the blood and is mostly in peripheral blood [7]. TAP test is a one-time combination detection of up to dozens of abnormal sugar chain proteins related to tumors, with high sensitivity and accuracy. TAP test has been widely used in the screening



and diagnosis of malignant tumors of the digestive system, breast, and gynecology but, rarely, reported in the diagnosis of thyroid cancer.

One-third of malignant tumors can be prevented, one-third can be cured if diagnosed early, and one-third can alleviate and prolong life [8]. The prevention and treatment of tumor mainly focuses on prevention and the combination of prevention and treatment [9]. In patients with early-stage tumors, the number of cancer cells is still at a reversible stage in the development of the tumor, the blood vessels that feed the tumor have not yet developed, and the tumor tissue is still fragile. Malignant tumors are the product of complex factors and gene mutations [10]. For a normal cell to accumulate these “three” changes, it generally takes a long time (many years of undetectable initial period). This provides valuable time and opportunity for early detection and effective blocking of tumors. There are three clinical methods for early detection and diagnosis: physical methods (X-ray, b-ultrasound, CT, MRI, PET-CT, etc.), biochemical tests (such as blood lipid and glucose), and histopathological examination [11–13]. Currently, the most accurate instruments can only detect tumors in diameter of  $>0.5$  cm. When the tumor mass from asymptomatic and gradually grow to be found by imaging, endoscopy and other examination is usually the clinical stage of the tumor, which has lost the optimal treatment period. That is to say, although these methods can assist the diagnosis and observation of tumors from different perspectives and to different degrees, they cannot fundamentally solve the problem of early detection and diagnosis of tumors. Domestic and foreign studies have shown that there is a close relationship between tumor marker TM and tumor [14, 15]. However, the sensitivity, specificity, and accuracy of a single TM test for tumor-assisted diagnosis or early diagnosis are poor, while TAP provides a new method for tumor diagnosis.

Scholars Li et al. assessed the level of the expression of CK20 and treatment effect and prognosis of postoperative colorectal cancer of correlation, the experiment to the postoperative follow-up of 62 patients with colorectal cancer, tumor tissue, and the abdominal cavity drainage liquid samples and blood sampling; in the fluid of the tumor specimens and abdominal cavity drainage, early cancer patients CK20 level lower than that of patients with advanced cancer, serum CK20 negative 3-year survival rate of patients was obviously higher than that of CK20 seropositive patients [16]. Scholars Wu and Huang discussed TAP and other serological biomarkers such as carcinoembryonic antigen (CEA), antigen 125 (CA125) sugar, and sugar antigen 19-9 (CA19-9) relationship, and its clinical application in patients with colorectal cancer (CRC) results showed that the TAP detection in CRC patients with high sensitivity and specificity of clinical monitoring in the process of CRC patients with chemotherapy can be used as a new independent index [17]. Scholars Li et al. combined TAP with hs-CRP level test to diagnose endometrial cancer in patients with endometrial thickness less than 8 mm. Their results showed that for suspected endometrial cancer patients with endometrial thickness less than 8 mm, the combined TAP and hs-CRP test could be used as a screening tool, providing new ideas for clinical diag-

nosis and treatment [18]. Scholar Fan studied the value of tumor markers C12 system in gastric cancer, using univariate and multivariate evaluation of tumor markers clinicopathologic correlation; using chi-square test to select key tumor markers in the diagnosis of gastric cancer is the most favorable combination; the result shows that the tumor marker biochip system has an important diagnostic value for gastric cancer, and CA19-9, CA242, CEA, and CA125 are four key indexes for tumor diagnosis [19]. Scholars Mo discusses the three tumor markers combined test value to the diagnosis of gastric cancer; the results showed that the levels of tumor markers gastric cancer group were obviously higher than that of control group; gastric cancer tumor markers combined detection sensitivity and accuracy were higher than pure detection, and CEA, CA199, and CA724 single detection has certain significance to the diagnosis of gastric cancer; the joint detection can improve the sensitivity and accuracy of diagnosis of gastric cancer [20]. Scholars Liu et al. studied the role of tumor markers in the differential diagnosis of benign and malignant ascites; the result shows that under the booking cutoff value of tumor markers, the diagnosis of ascites tumor markers' effect is better than that of the serum tumor markers; combined use of tumor markers and cytology can increase the diagnostic rate of the latter is 37%; in malignant ascites cytology negative, tumor markers are helpful to differentiate malignant ascites and benign ascites; ascites tumor markers are used in combination with sensitivity of 86%, specificity of 97%, which means that the detection of tumor markers may be useful auxiliary cytology [21]. Scholars Zhu et al. discussed in this paper the four kinds of tumor markers in serum and ascites and ascites/serum ratio value in the differential diagnosis of benign and malignant ascites, the experiment selected as the research object, and 76 patients divided into malignant ascites group (45 cases) and benign ascites group (31 cases); the experimental results show that compared with the single index, the joint detection in the serum and ascites tumor markers can significantly increase the sensitivity and specificity of diagnosis [22]. Scholars Trapé et al. studied the ascites tumor antigen 125 (CA125) and cytokeratin 19 soluble pieces (CYFRA21-1) and the tumor markers in the accuracy of the diagnosis of malignant ascites and analyzed 143 cases of patients with undiagnosed liver muscle inflammation CA125 and CYFRA21-1; the use of CA125 the sensitivity of the differential diagnosis of malignant ascites was 39.7% and specificity of 98.8%; the sensitivity of the differential diagnosis CYFRA21-1 was 50.0% and specificity of 97.6%, a combination of CA125 and CYFRA21-1. The sensitivity and specificity of these two tumor markers were 65.5% and 96.5%, respectively. In cytological negative patients, the sensitivity and specificity of these two tumor markers were 50% and 96.5%, respectively. The conclusion that the determination of tumor markers in treating ascites has a certain reference value for the diagnosis of ascites patients [23]. Scholars Cao and others studied the serum ALDH1A1 in nonsmall cell lung cancer diagnosis and prognosis of the potential clinical value in the experiment collected 100 patients with nonsmall cell lung cancer tumor resection before 60, 60 patients with benign lung disease and serum samples of healthy volunteers



and using sandwich ELISA method were retrospectively analyzed, their results show that the levels of serum ALDH1A1 associated with the occurrence and progress of nonsmall cell lung cancer, and serum ALDH1A1 detection is helpful to the diagnosis and prognosis of nonsmall cell lung cancer. Combined application of carcinoembryonic antigen and ALDH1A1 can significantly improve the diagnostic rate of nonsmall cell lung cancer [24]. Xiao and other scholars studied the lung tissue and tumor markers in serum HE4 expression levels and its role in the differential diagnosis of lung cancer, the experiment tested the 214 cases of lung cancer patients and 76 cases of healthy human lung adenocarcinoma, lung squamous carcinoma, and adjacent tissues in the expression of HE4, and results show that the HE4 as a biomarker of lung cancer, for the early diagnosis of lung cancer, can provide important reference [25].

Thyroid cancer is a relatively common endocrine gland malignant tumor; if improper treatment, there will be a high risk of recurrence or metastasis, and abnormal sugar chain glycoprotein (TAP) has a close relationship with the development of the disease; therefore, the purpose of this article is to discuss abnormal sugar chain glycoprotein (TAP) as thyroid cancer curative effect evaluation and radiation and chemotherapy after surgery clinical significance. In this paper, 95 patients with thyroid cancer diagnosed in a hospital were selected as the study objects and treated as the observation group. The clinical and follow-up data of the observation group were retrospectively analyzed. Meanwhile, 55 healthy patients were randomly selected as the control group. TAP, squamous cell carcinoma antigen (SCC) level, and carcinoembryonic antigen (CEA) level were detected in peripheral blood of 95 patients with thyroid cancer before and after treatment. The short-term efficacy was evaluated by chest CT examination, and the changes of the three markers before and after treatment and the correlation with the short-term efficacy of the patients were compared. According to the results of testing, the TAP positive expression in patients before radiotherapy can better predict the recent curative effect has certain clinical value; before radiotherapy TAP positive expression rate was significantly higher than that of healthy people, TAP positive expression quantity decreased obviously after radiation treatment, and patients with a recent radiotherapy curative effect is good or bad and negatively correlated with the degree of TAP protein positive expression; TAP high protein in patients with recent poor radiation effects, prompt the factor can be predicted in the near future curative effect of the molecular markers, and can TAP level for clinicians provide certain reference for targeted therapy.

## 2. Thyroid Cancer and TAP Test

**2.1. Diagnosis of Thyroid Cancer.** It is well known that the possibility of thyroid cancer should be considered when the following clinical signs occur: rapid growth of thyroid nodules, hard texture, adhesion and fixation with surrounding tissues and organs, or regional lymph node enlargement. Especially solitary nodules occurring in childhood or male adults; thyroid nodules in patients with a history of head,

neck, or chest radiotherapy; and thyroid nodules in patients with Grave's disease were characterized by low functional nodules on radionuclide scan and substantial masses on ultrasonography. Fine needle aspiration cytology (FNAC) should be performed in the above patients. In addition, the possibility of thyroid cancer should be suspected even if the following signs are clinically present.

**2.1.1. Cystic Masses.** It is generally believed that thyroid tumor cystic or solid is helpful to benign and malignant diagnosis, even some people think that cystic masses malignant possibility is not big, in the treatment that often uses the simple cyst excision nor postoperative pathologic examination, leading to postoperative thyroid cancer relapses, is common: thyroid cancer actually can present a substantial, cystic, or mixed, and a report into the lumen of the mastoid shaped gland thyroid carcinoma tissues form can reach 45%~60% adenocarcinoma (sac). Malignant lesions accounted for 32% of the 60 cystic thyroid masses reported by Resent. Therefore, a cystic thyroid mass cannot exclude malignant lesions.

**2.1.2. Multiple Mass.** Although thyroid cancer is dominated by a single nodule, the incidence of unilateral multicentric adenocarcinoma is reported to be 13%, and bilateral multicentric carcinoma is reported to be 6.0%-8.8%. Thyroid cancer, especially after radiotherapy, is often multicentric.

**2.1.3. Benign Thyroid Disease Coexists with Thyroid Cancer.** In addition to thyroid adenoma and nodular goiter can be cancerous, thyroid can also coexist with some benign lesions. It was reported that 0.76%~8.7% of hyperthyroidism was associated with thyroid cancer, with a greater likelihood of toxic nodular goiter and high functional adenoma stage associated with thyroid cancer. The incidence of Hashimoto's disease with thyroid cancer is 5%~17%. There have also been reports of the coexistence of thyroid tumor and thyroid cancer.

Since it is very difficult to diagnose thyroid cancer before surgery, in cases requiring surgical treatment, no matter whether the diagnosis result is benign or malignant before surgery, patients with conditions should have a quick biopsy during routine surgery to determine the nature of pathology and guide the surgical method. The quality of frozen sections and the difference of thyroid cancer tissues may lead to false-negative results. Therefore, in addition to intraoperative frozen sections and rapid disease examination, routine paraffin sections should also be performed postoperatively to prevent missed diagnosis.

In addition, attention should be paid to thyroid microcarcinoma or occult carcinoma. Patients often seek treatment for swollen lymph nodes in the neck, but there are no clinical symptoms at this time. Therefore, patients with metastatic adenocarcinoma in the neck lymph nodes should carefully examine the thyroid gland.

## 2.2. Treatment of Thyroid Cancer

**2.2.1. Treatment Principles of Thyroid Cancer.** The clinical manifestations, treatment methods, and therapeutic effects of different thyroid cancers are not completely the same. At present, it is generally accepted that the treatment plan

should be determined according to the extent of thyroid primary lesion expansion, pathological type, and patient age, so as to achieve both radical treatment of the lesion and quality of life. The operation should include the primary lesion and the metastatic lesion in the neck. Although the choice of operation for thyroid cancer is still controversial, the current consensus is as follows.

**2.2.2. Selection of Primary Cancer Foci.** Medullary carcinoma: total thyroidectomy and central lymph node dissection; undifferentiated cancer: preoperative diagnosed people generally do not advocate surgical treatment, only when the trachea is invaded, breathing difficulties, only consider surgery to remove the obstruction. If the postoperative pathological report is an undifferentiated tumor, radiotherapy should be added.

**2.2.3. The Choice of Surgical Method for Cervical Lymph Node Metastasis.** At present, the vast majority of scholars advocate neck cleansing, generally do not advocate the use of preventive neck cleansing, such as bilateral metastasis, then phased neck cleansing.

**2.2.4. The Pathological Report after Thyroid Nodule Resection Was the Management of Thyroid Cancer.** For patients who have undergone total lateral lobe resection, thyroxine inhibition therapy can be used under close observation. If subtotal resection of the affected lobe or local resection of the mass is performed, another operation should be performed immediately with the same surgical method as before. Because in the latter case, the residual gland has a high residual cancer rate.

**2.2.5. Radiotherapy after Surgery.** Undifferentiated thyroid cancer is highly sensitive to external radiation therapy. Among the 305 cases of thyroid cancer reported in XiangYa hospital, 25 cases were undifferentiated. Among them, 7 underwent palliative surgery and received external radiation at 10-20 d after surgery. During follow-up, 5 patients survived for 18 to 24 months. Two cases survived for more than 2 years. In addition, 150 cases of well-differentiated thyroid cancer were treated with radiotherapy after surgery, and the 5, 10, and 15 survival rates were 76.19%, 56.25%, and 66.67%, respectively. Therefore, if there is no contraband, radiotherapy should be used after thyroid cancer, but myeloid cancer is not sensitive to radiotherapy.

**2.3. TAP Detection.** Glycoprotein (glycoprotein) is composed of two parts, protein and glycan connected by a covalent bond. Human cells can synthesize a certain type and amount of glycoprotein. Glycosylation of proteins is the most common modification after protein translation. Studies have shown that 1% of human genome involved in glycosylation and modification of sugar chains and about 70% of human proteins contain more than one sugar chain. Glycoprotein sugar chains are associated with a variety of biological processes, such as transport and processing of new peptide chains, interference with glycoprotein biological half-life, and influence on glycoprotein bioactivity and molecular recognition.

TABLE 1: Distribution of clinical baseline data of 95 patients with thyroid cancer.

Gender	Number of cases	Constituent
Male	60	63.15%
Getting out	35	36.84%

Oncogene and tumor suppressor gene mutations occur in human body. Malignant carcinogenesis is widespread abnormal glycosylation of membrane glycoproteins in the process of changes in cellular glycosyltransferase and colloidal enzyme activities. Modified proteins with abnormal changes in sugar chain structure are mainly represented by the new synthesis of sugar chains, and the hidden exposure of sugar chains is incomplete. After sugar chain synthesis, a variety of abnormal sugar chain glycoproteins are produced. When malignant cells approach a certain number, TAP produced by them is released into the blood and can be detected in peripheral blood.

At present, it is difficult to detect abnormal glycoprotein sugar chain directly. CLU is a group of sugar-binding proteins that can specifically recognize and bind to specific glycosyl sequences in a specific structural monosaccharide or oligosaccharide. Different CLU can bind to different sugar chains.

The main components of TAP detection reagent are condensates (lectin combination), which can agglutinate dozens of TAP to generate crystal-like condensates with specific forms. The image analyzer of TAP detection can be significantly observed, and it is very different from the form formed by normal blood. In this way, we can determine whether or not the subject has TAP in their blood and how much TAP is in their blood.

TAP has great characteristics and advantages as a marker detection. Glycoprotein is mainly distributed on the cell surface, cell endocrine granules, and extracellular matrix. O-sugar chains and *n*-sugar chains, as small molecular substances, can be specifically identified by lectins (lectins combination) and bound by noncovalent bonds, resulting in condensation and precipitation. TAP test can realize the combination detection of more than 20 kinds of abnormal sugar chains related to tumor at one time to gather a variety of tumor signals.

#### 2.4. Clinical Significance of TAP Test Results

**2.4.1. The TAP Is Positive.** The TAP content in the samples was abnormal, and the crystal-like condensate was larger. There are more than 10 cancer cells in the body, and the tumor may not have formed at this time, but the probability of tumor occurrence is significantly higher than that of the tap-negative population. At this point, the patient should be further confirmed with the combination of the medical history, symptoms, signs, other TM tests, and imaging results. Interference factors of TAP positive were excluded. If not, it is recommended to review every two weeks. If the results of three consecutive tests all contain crystal-like condensates, great attention should be paid to them and close observation should be followed up (it is recommended to review every six

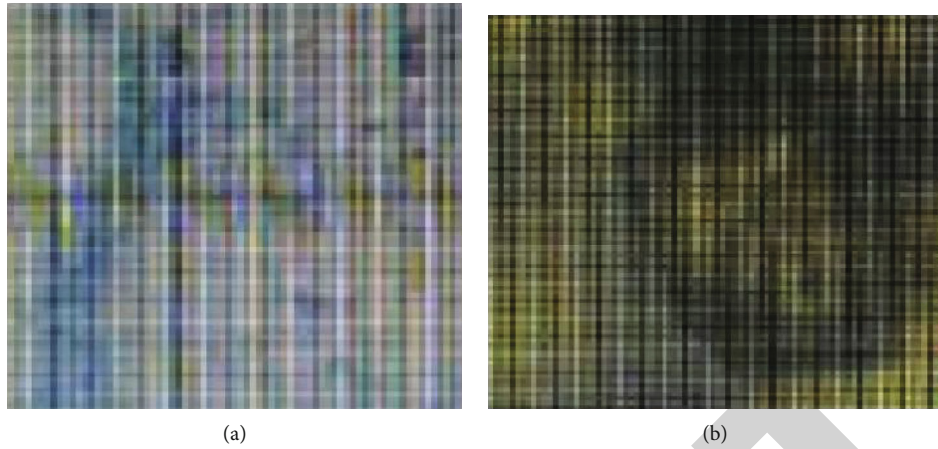


FIGURE 1: TAP detection image.

months). Concerns about chronic conditions can be addressed by using appropriate methods to improve immunity until the TAP test condensates disappear. TAP positive is not necessarily a tumor but can be seen as a warning signal of a high-risk group, which should be paid attention to. Positive people should actively find out the cause, to see whether there is a bad life habits, psychological pressure, and to apply effective health intervention, correction is possible to reverse, such as multiple strong positive should be in-depth clinical examination, in order to find early, early treatment.

**2.4.2. Weak Positive TAP.** The TAP content in the samples was abnormal, and the crystal-like condensate was small. Body cancer cell number is in 10 10 between and should prevent to turn positive. Eliminate whether there is positive interference disease, pay attention to chronic disease, and improve immunity appropriately.

**2.4.3. The TAP Is Negative.** The contents of TAP in the samples were normal without obvious agglomerates. This indicates that the number of cancer cells in the body is less than 1 million, which is normal, and the probability of tumor formation within one year is very low. Therefore, TAP test once a year can achieve the purpose of early prevention.

**2.4.4. Significance of TAP Expression Changes during and after Treatment.** The negative expression of TAP after surgery or other treatment indicates that the tumor may have been completely removed. After TAP negative expression or TAP weak positive expression, if TAP positive expression occurs again within a short period of time, it means that there is a residual tumor or the tumor has metastasized. After a long period of time after the negative expression of TAP, the positive expression of TAP was reestablished, indicating recurrence or metastasis.

### 3. Experimental Methods

**3.1. Clinical Data.** A total of 95 patients diagnosed with thyroid cancer admitted to a hospital from May 2016 to May 2018 were selected as the study objects as the observation group. All the patients had complete clinical and follow-up

TABLE 2: Expression of TAP in the observation group was compared with that in the control group before treatment.

Group	Number of cases	The TAP	
		Positive	Feminine
Observation group	95	71	24
Matched group	55	6	49
Chi-square value		63.1	
P value		<0.001	

TABLE 3: The Changes of tumor markers in patients with thyroid cancer before and after radiotherapy.

	The TAP ( $\mu\text{m}^2$ )	The CEA (ng/mL)	SCC (ng/mL)
Prior treatment	149.33	2.36	1.20
After treatment	65.32	2.40	1.19

TABLE 4: Changes of tumor markers before and after radiotherapy in patients with thyroid cancer with company's short-term efficacy.

Group	Period	The TAP ( $\mu\text{m}^2$ )	The CEA (ng/mL)	SCC (ng/mL)
PR ( $n = 53$ )	Prior treatment	159.97	2.36	1.30
	After treatment	53.31	2.35	1.27
CR ( $n = 19$ )	Prior treatment	155.76	2.39	1.17
	After treatment	47.57	2.31	1.13
NR ( $n = 23$ )	Prior treatment	119.51	2.35	1.01
	After treatment	107.67	2.61	1.35

data, including 60 male patients and 35 female patients, aged 45-75 years. In addition, 55 healthy patients who came to the hospital for physical examination from May 2016 to May 2018 were randomly selected as the control group, including 40 males and 15 females, aged from 30 to 70 years.

Inclusion criteria: the pathological types of patients were confirmed by histopathology. Local lesions cannot be removed surgically, or patients refuse surgery. All patients did not receive any antitumor therapy before enrollment.



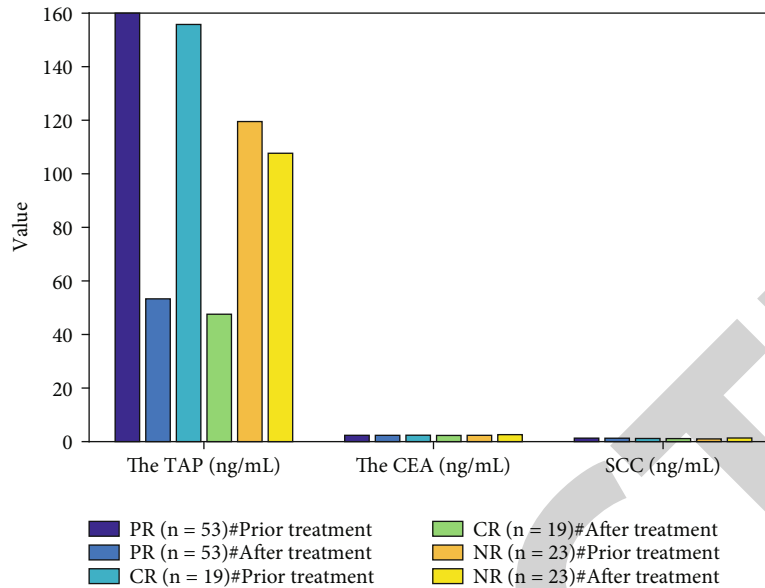


FIGURE 2: The changes of tumor markers before and after radiotherapy in patients with thyroid cancer with company's short-term efficacy.

Cargill score  $\geq 70$  can enter the liquid diet at least. Lesion length is  $< 10$  cm. The results of routine blood tests in the laboratory were all within the normal range, the electrocardiogram showed no abnormality, and the liver and kidney functions showed no abnormality. No other antitumor therapy was performed recently after radiotherapy, as shown in Table 1.

**3.2. TAP Detection Method.** In this study, tumor laboratory personnel were responsible for TAP test, TAP test comprehensive diagnostic instrument, and TAP test kit at different time points (before treatment and at the end of treatment). The operation was conducted in strict accordance with the operating instructions. Blood samples were taken from the tip of the study object's finger and made into 3 evenly thin blood slices, which were dried naturally. TAP detection reagent was shaken well, and three drops of each sample were added vertically to the blood slices with a dropper. The samples were placed under the TAP integrated slide reader microscope, and the 4 times flat field achromatic objective was used to scan and observe the 3 spots of the blood sample in the display screen at the last time, looking for the agglomerates with specific morphology. The TAP result is judged according to the area of condensate. TAP is in the blood and can be in condensing agent reaction generates specific image; negative image snowflake; dendritic, tree round-shaped, or loose sand; dark brown granules; no visible particle condensed matter, the following as shown in Figure 1(a); and positive condensed particle diameter greater than  $38\mu\text{m}^2$ , oval, round, or irregular polygon at the edge of the relatively complete crystal sample condensate, as shown in Figure 1.

TAP test reference range: normal TAP/no obvious condensate  $0-121\mu\text{m}^2$ ; smaller TAP anomaly/condensate,  $121-225\mu\text{m}^2$ ; and TAP anomaly/condensate was larger,  $> 225\mu\text{m}^2$ .

In addition, 10 mL of patients' fasting peripheral venous blood was collected in the morning and centrifugated at

3000 r/min for 12 min. CEA and SCC levels were detected by automatic electrochemical luminescence immunoassay.

SPSS software package was used to process the statistical data. Measurement data were expressed as mean  $\pm$  standard deviation, *T*-test was used, and counting data were expressed as percentage. Chi-square test was used, and  $P < 0.05$  was considered statistically significant. Correlation analysis of TAP positive expression and short-term efficacy was conducted by Spearman level correlation analysis;  $P < 0.05$  was considered statistically significant.

## 4. Discussion

**4.1. Analysis of TAP Expression in the Two Groups before Treatment.** Before treatment, the TAP positive expression rate in the observation group was 74.73% (71/95), significantly higher than that in the control group (10.91% (6/55), with a statistical difference ( $P < 0.05$ ) as shown in Table 2.

**4.2. Changes of Tumor Markers before and after Treatment.** The mean levels of TAP, CEA, and SCC in 95 patients before radiotherapy were  $149.33\mu\text{m}^2$ , 2.36 ng/mL, and 1.20 ng/mL, respectively. After radiotherapy, the TAP level was  $65.32\mu\text{m}^2$ , significantly lower than that before radiotherapy, and the difference was statistically significant ( $P < 0.01$ ), while CEA and SCC were 2.40 ng/mL and 1.19 ng/mL, respectively, showing no statistically significant difference compared with that before radiotherapy ( $P > 0.05$ ), as shown in Table 3:

**4.3. Changes of Tumor Markers in Patients with Different Short-Term Effects before and after Treatment.** After radiotherapy, the short-term efficacy of 95 patients with thyroid cancer was evaluated as CR in 19 cases (20%), PR in 53 cases (55.7%), and NR in 23 cases (24.2%). The TAP index in PR and CR patients was significantly lower after radiotherapy than before radiotherapy, and the difference was statistically significant ( $P < 0.05$ ). CEA and SCC levels of thyroid cancer

TABLE 5: Rank sum test of tumor markers before and after radiotherapy for thyroid cancer with company's short-term effects.

The test period and the items		NR	PR	CR
The prior treatment TAP	Mean rank	39.97	49.35	47.36
	<i>P</i> value		0.473	
After treatment the TAP	Mean rank	65.11	41.76	39.11
	<i>P</i> value		0.002	
The prior treatment CEA	Mean rank	46.35	46.76	51.96
	<i>P</i> value		0.793	
After treatment the CEA	Mean rank	51.63	45.19	46.31
	<i>P</i> value		0.693	
The prior treatment SCC	Mean rank	37.11	50.03	47.97
	<i>P</i> value		0.137	
After treatment SCC	Mean rank	47.65	43.16	48.63
	<i>P</i> value		0.793	

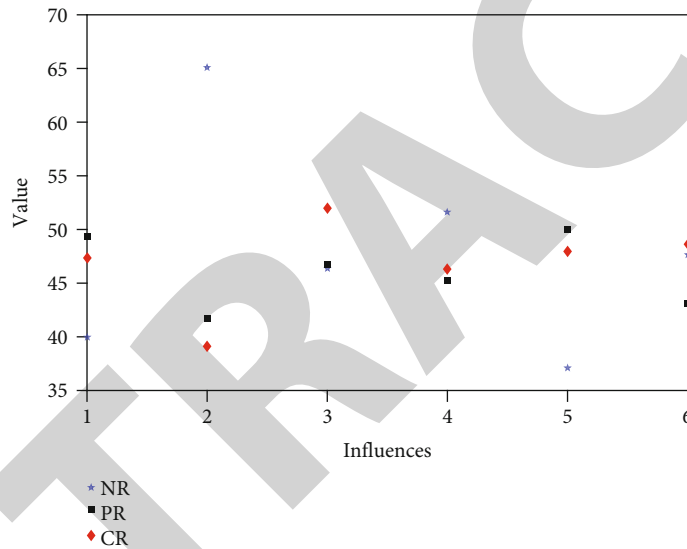


FIGURE 3: The rank sum test of tumor markers before and after radiotherapy for thyroid cancer with the company.

patients with the same efficacy evaluation showed no significant difference ( $P > .05$ ) as shown in Table 4 and Figure 2.

**4.4. Relationship between Different Short-Term Efficacy and TAP.** At the end of treatment, esophageal cancer patients with different short-term efficacy showed a difference in TAP after treatment, and the highest in nonremission (NR), followed by partial remission (PR), and the lowest in complete response (CR). However, there was no difference in SCC and CEA results among esophageal cancer patients with different short-term efficacy, as shown in Table 5 and Figure 3.

**4.5. Analysis of the Correlation between TAP Positive Expression and the Short-Term Efficacy of Treatment.** Of the 95 patients in the study, none had serious complications caused by radiotherapy, and all enrolled patients successfully completed the treatment. The short-term efficacy evaluation was mainly based on the results of CT examination. The specific situation of the efficacy evaluation is as follows: complete

response (CR) in 19 cases, accounting for 20%. There were 53 cases with partial remission (PR), accounting for 55.7%. There were 23 cases without remission (NR), accounting for 24.2%. Among the 95 patients, statistical analysis showed that there was a statistically significant difference in the effective rate of radiotherapy between those with positive TAP expression and those with negative TAP expression. Statistical Spearman level correlation analysis showed that there was a negative correlation between TAP positive expression and short-term efficacy.

## 5. Conclusions

- (1) Thyroid carcinoma is a relatively common endocrine gland malignant tumor; if treatment is not appropriate, there will be a high risk of recurrence or metastasis, and abnormal sugar chain glycoprotein (TAP) has a close relationship with the development of the disease; therefore, the purpose of this article is to discuss abnormal sugar chain glycoprotein (TAP) as

thyroid cancer curative effect evaluation and radiation and chemotherapy after surgery clinical significance

- (2) In this study, 95 patients diagnosed with thyroid cancer in a hospital were selected as the study objects and set as the observation group. The clinical and follow-up data of the patients were retrospectively analyzed. Meanwhile, 55 healthy patients were randomly selected as the control group. TAP, squamous cell carcinoma antigen (SCC) level, and carcinoembryonic antigen (CEA) level were detected in peripheral blood of 95 patients with thyroid cancer before and after treatment. The short-term efficacy was evaluated by chest CT examination, and the changes of the three markers before and after treatment and the correlation with the short-term efficacy of the patients were compared
- (3) The results of this study showed that the level of TAP positive expression in thyroid cancer patients was significantly higher than that in healthy people, suggesting that TAP can be used as a specific indicator for early thyroid cancer screening. After radiotherapy, the TAP positive expression level of patients with thyroid cancer gradually decreased, suggesting that TAP can be used as a specific indicator to determine the short-term efficacy of radiotherapy for thyroid cancer. The short-term radiotherapy effect of thyroid cancer patients is negatively correlated with the positive expression level of TAP protein. Thyroid cancer patients with high TAP protein have poor short-term radiotherapy effect, suggesting that this factor can be used as an indicator to predict the short-term radiotherapy effect of thyroid cancer patients. The detection of TAP positive expression in thyroid cancer patients before radiotherapy can better predict the short-term efficacy and has high clinical value. Clinicians can make individualized treatment plans according to the different expressions of this factor in different patients

## Data Availability

The data in the article are all real and available.

## Conflicts of Interest

There are no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

## References

- [1] C. M. Kitahara and J. A. Sosa, "The changing incidence of thyroid cancer," *Nature Reviews Endocrinology*, vol. 12, no. 11, pp. 646–653, 2016.
- [2] N. Li, X. L. Du, L. R. Reitzel, L. Xu, and E. M. Sturgis, "Impact of enhanced detection on the increase in thyroid cancer incidence in the United States: review of incidence trends by socioeconomic status within the surveillance, epidemiology, and end results registry, 1980–2008," *Thyroid*, vol. 23, no. 1, pp. 103–110, 2013.
- [3] C. la Vecchia, M. Malvezzi, C. Bosetti et al., "Thyroid cancer mortality and incidence: a global overview," *International Journal of Cancer*, vol. 136, no. 9, pp. 2187–2195, 2015.
- [4] M. B. Ukrainski, E. A. Pribitkin, and J. L. Miller, "Increasing incidence of thyroid nodules and thyroid cancer: does increased detection of a subclinical reservoir justify the associated anxiety and treatment?," *Clinical Therapeutics*, vol. 38, no. 4, pp. 976–985, 2016.
- [5] F. Teng, C. Yaqing, and D. O. Ultrasonography, "Diagnostic advances of thyroid cancer in ultrasonography," *China Oncology*, vol. 25, no. 4, pp. 316–320, 2015.
- [6] L. Li, B.-D. Chen, H.-F. Zhu et al., "Comparison of pre-operation diagnosis of thyroid cancer with fine needle aspiration and core-needle biopsy: a meta-analysis," *Asian Pacific Journal of Cancer Prevention*, vol. 15, no. 17, pp. 7187–7193, 2014.
- [7] L. Zhang, X. Guo, Y. Min, and J. Xu, "Tumor abnormal protein (TAP) examination contributes to primary diagnosis of bladder cancer," *International Journal of Clinical and Experimental Medicine*, vol. 8, no. 10, pp. 18528–18532, 2015.
- [8] L. F. Fajardo, B. Egbert, J. Marmor, and G. M. Hahn, "Effects of hyperthermia in a malignant tumor," *Cancer*, vol. 45, no. 3, pp. 613–623, 1980.
- [9] S. M. S. Islam and L. W. Niessen, "Population based cancer registry in the developing countries: a first step towards cancer control programs and research," *Journal of Cancer Research and Therapeutics*, vol. 11, no. 4, 2015.
- [10] F. Yang and Y. Yu, "Tumor microenvironment-the critical element of tumor metastasis," *Chinese Journal of Lung Cancer*, vol. 18, no. 1, pp. 48–54, 2015.
- [11] H. Liu, W. Li, X. Yi, P. Yigang, and G. Obin, "Diagnosis and differential diagnosis for solitary fibrous tumor in the abdomen and pelvis by CT," *Journal of Central South University*, vol. 42, no. 4, pp. 406–412, 2017.
- [12] Y. Zhang, Y. Huang, Y. Wu, L. Kong, S. Huang, and C. Hu, "Nasal endoscope, MRI, and ultrasound-guided fine needle aspiration in the diagnosis of primary head and neck tumor," *Chinese journal of otorhinolaryngology head and neck surgery*, vol. 49, no. 3, pp. 223–226, 2014.
- [13] W. Wei-min, L. Song, and Y. Jun, "Research on the value of diagnosis of malignant and benign by using MRI to assess Intraductal papillary mucinous tumor (IPMN)," *China Medical Equipment*, vol. 14, no. 8, pp. 5533–5546, 2015.
- [14] O. Ocal, V. Pashkov, R. K. Kollipara, J. B. Lorens, G. H. Swift, and R. A. Brekken, "Abstract 5182: a rapid in vivo screen for pancreatic ductal adenocarcinoma therapeutics using the tumor marker Rgs16::GFP," *Cancer Research*, vol. 76, Supplement 14, pp. 5182–5182, 2016.
- [15] S. Ferraro, R. Mozzi, and M. Panteghini, "Tumor marker ordering: do not lose control: a prospective clinical trial," *American Journal of Clinical Pathology*, vol. 144, no. 4, pp. 649–658, 2015.
- [16] W. X. Li, H. W. Xiao, X. Q. Hong, and W. X. Niu, "Predictive value of ck20 in evaluating the efficacy of treatment and prognosis after surgery for colorectal cancer," *Genetics and Molecular Research*, vol. 14, no. 2, pp. 5823–5829, 2015.
- [17] X. Y. Wu and X. E. Huang, "Clinical application of serum tumor abnormal protein (TAP) in colorectal cancer patients,"

## Research Article

# Task Transfer Learning for EEG Classification in Motor Imagery-Based BCI System

Xuanci Zheng,<sup>1</sup> Jie Li ,<sup>1</sup> Hongfei Ji ,<sup>1</sup> Lili Duan,<sup>1</sup> Maozhen Li,<sup>2</sup> Zilong Pang,<sup>1</sup> Jie Zhuang,<sup>3</sup> Lu Rongrong,<sup>4</sup> and Gao Tianhao<sup>4</sup>

<sup>1</sup>College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

<sup>2</sup>Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, UK

<sup>3</sup>School of Psychology, Shanghai University of Sport, Shanghai 200438, China

<sup>4</sup>Department of Rehabilitation, Huashan Hospital, Fudan University, Shanghai 200040, China

Correspondence should be addressed to Jie Li; [nijanice@163.com](mailto:nijanice@163.com) and Hongfei Ji; [jhf@tongji.edu.cn](mailto:jhf@tongji.edu.cn)

Received 17 May 2020; Revised 25 June 2020; Accepted 27 July 2020; Published 15 December 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Xuanci Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The motor-imagery brain-computer interface system (MI-BCI) has a board prospect for development. However, long calibration time and lack of enough MI commands limit its use in practice. In order to enlarge the command set, we add the combinations of traditional MI commands as new commands into the command set. We also design an algorithm based on transfer learning so as to decrease the calibration time for collecting EEG signal and training model. We create feature extractor based on data from traditional commands and transfer patterns through the data from new commands. Through the comparison of the average accuracy between our algorithm and traditional algorithms and the visualization of spatial patterns in our algorithm, we find that the accuracy of our algorithm is much higher than traditional algorithms, especially as for the low-quality datasets. Besides, the visualization of spatial patterns is meaningful. The algorithm based on transfer learning takes the advantage of the information from source data. We enlarge the command set while shortening the calibration time, which is of significant importance to the MI-BCI application.

## 1. Introduction

As a technique decoding brain activity, Brain-Computer Interface (BCI) based on electroencephalogram (EEG) enables people to interact with computers without the involvement of peripheral muscular activity, which builds a communication bridge between the brain and computer. With signal processing, pattern classification, machine learning, and other techniques, the BCI system translates different kinds of brain activities such as attentive mental states [1], motor imagery [2–5] (MI), and so on into machine instruction for controlling devices. For example, it could be used to control unmanned aerial vehicles [5], to help train soldiers [6], to help patients

with motor disabilities, such as amyotrophic lateral sclerosis [7], brainstem stroke [2, 8] to recover.

Among most studies about EEG signal classification, researchers are highly concerned about the EEG signal generated from motor imagery which has been widely used in many BCI applications [2–5]. Motor imagery is a BCI paradigm in which brain activity is generated at the sensorimotor cortex during the imagination of the limb movement [9] such as left hand (LH), right hand (RH), and both feet (F). During the MI experiment, the power of the alpha band (8–12 Hz) and beta band (13–30 Hz) increase or decrease in the sensorimotor cortex of the ipsilateral hemisphere and the contralateral hemisphere [10, 11]. The power suppression and



enhancement observed through EEG signal are, respectively, called event-related desynchronization (ERD) and event-related synchronization (ERS) [12]. The ERD/ERS patterns can be used for translating brain activity and classifying the imagination of limbs through machine learning.

However, most of MI-BCI promising applications are prototypes and still scarcely used outside laboratories. Among the reasons preventing MI-BCI from being widely put into practice, we want to give solutions to two of them, which are the long calibration time [13–16] and the lack of enough MI commands [17].

The first problem, namely, the long calibration time, means researchers require a large number of calibration trials for training a subject-specific and task-specific model, which is time-consuming in collecting data and training model. Due to the intersubject and intertask variability of the EEG signal [18], the training samples need to contain the EEG data from every MI commands for model construction. Thus, the data collection for BCI takes a lot of time. And the variability between datasets makes it difficult to give a model fit for every subject, which lengthens the time for the training model. As for the second question, the lack of enough MI commands makes it less available for BCIs. In the traditional MI experiment, there are almost 4 types of commands for subjects to imagine, that is left hand, right hand, both feet, and tongue [19].

In order to add more commands into the BCI system, the combination of single limbs command can be added into the command set. For example, on the basis of existing MI command (LH, RH, and F), we can obtain combined commands, movement of both left hand and right hand simultaneously (LH&RH), movement of the left hand and both feet simultaneously (LH&F), and movement of the right hand and both feet simultaneously (RH&F). However, researchers have to collect the same scale of dataset for new MI commands as which for old commands, which multiplies the time spent on data collection. The larger MI dataset also lengthen the time spent on model training. Besides, as for a new subject, the imagination of new MI commands is harder than typical commands, which leads them to train those commands for a long period of time before the recording. Due to the increasing number and complexity of the entire six commands, the expansion of the command set becomes less feasible.

In order to shorten the time on model construction as well as enlarge the command set, transfer learning would have a profound impact. Transfer learning algorithms use datasets, features, or model parameters [20] from the source domain for training the model in the target domain so as to reduce the scale of training data in the target domain, which reduces the sampling and training cost. In the classification of the EEG signal, there have been many algorithms based on task-to-task transfer and subject-to-subject transfer in response to intertask and intersubject variability. Among those algorithms, the features extractors based on the Common Spatial Pattern algorithm (CSP) are one of the mainstream techniques. The method of CSP designs spatial filters, which maximize the discriminability of two classes of data and make the variances in filtered data are optimal for classification [21–23]. Based on CSP

and transfer learning, researchers raised a lot of algorithms to generate spatial filters from the source domain, for example, Regularized CSP [24–26], stationary subspace CSP [27], Bayesian CSP [28]. Through adjusting parameters, finding the common subspace, and using the Bayesian model, similar features from source subjects can be shared and the feature extractors from the source domain can be used in the target domain.

The solution to calibration time extension caused by command set expansion is to decrease the scale of the dataset collected for new commands. Therefore, in order to increase the number of commands on the premise of not lengthening calibration time, it is proper to make the most of other available data for new filters and classifiers. While the intertask variability of EEG signals varies from one class to another, the principle feature characteristics remain invariant across classes. Thus, although it is unwise to simply add training set of old commands to that of new classes, the prior knowledge collected from the previous training set can be used into the building of filters and classifiers for new classes. Consequently, using transfer learning to construct model is an optimal solution. We regard data from typical commands and new commands as source data and target data, respectively. And we use source data to improve the property of our model for target data.

In this paper, we provide a spatial pattern transfer algorithm. We add a screening process before the utilization of spatial filters generated from datasets of old MI commands. In order to shrink the difference of the generated filters and objective filters, we reject the filters performing badly based on the fisher ratio [29, 30].

The remainder of this paper is structured as follows. In Section 2, we introduce our experimental paradigm, our dataset as well as our algorithm. In Section 3, experimental results are shown. The paper concludes with a discussion of the results in Section 4.

## 2. Materials and Methods

In this section, we show our experiment and our algorithm used for the data recorded in the experiment.

### 2.1. Experiment and Data Processing

**2.1.1. Experiment.** We recorded brain activity from 5 healthy subjects in the experiment. The BCI paradigm consisted of six different motor imagery tasks, namely the imagination of the left hand (class 1), right hand (class 2), both feet (class 3), left hand and right hand (class 4), left hand and both feet (class 5), and right hand and both feet (class 6). On the basis of traditional commands, subjects were more familiar with the combined commands comparing to imagining new body parts. Consequently, we chose LH&F, RH&F, and LH&RH as our new commands.

The experiment was comprised of 6 runs separated by short breaks. One run consisted of 60 trials (10 for each of the six classes), yielding a total of 360 trials for every subject (60 trials in each class).

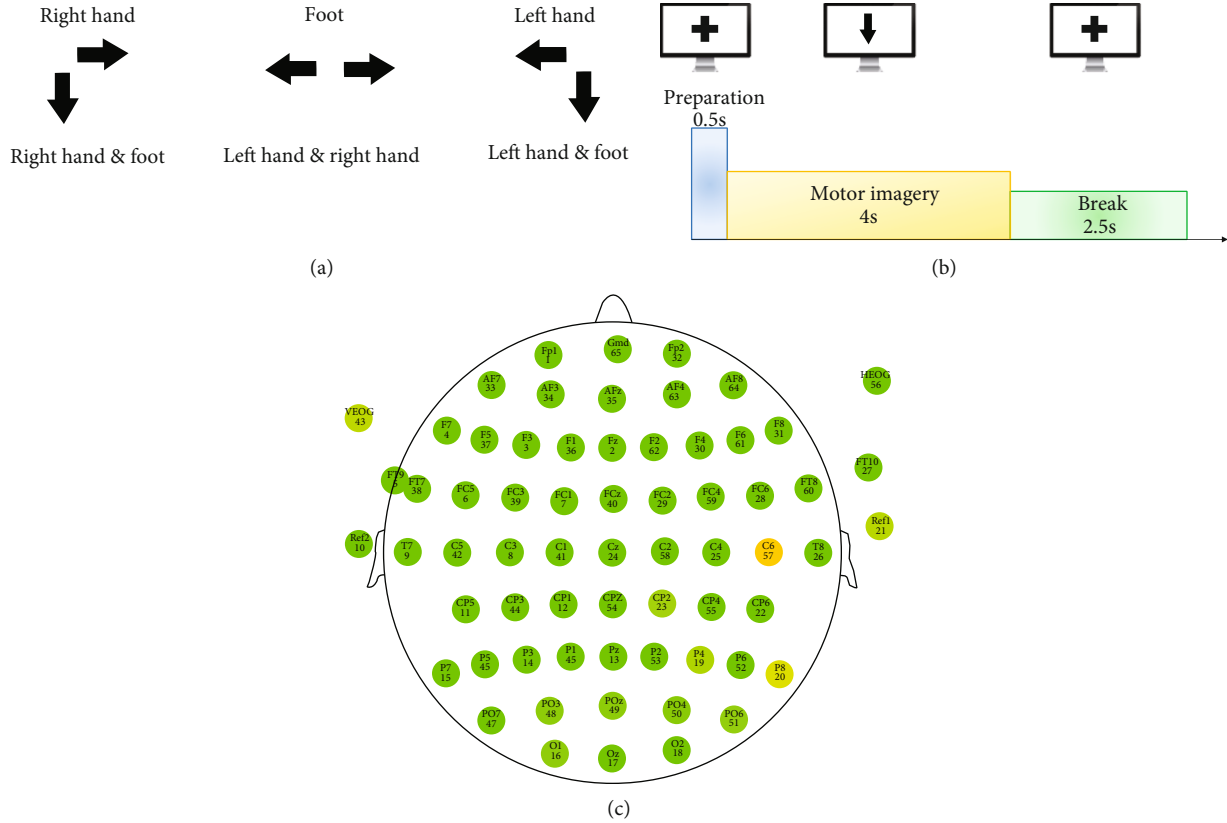


FIGURE 1: (a) The correspondence between the image on the screen and task of MI commands. (b) The time series of one trial in a run, and the image shown on the screen. (c) Electrode montage was used in the experiment.

The subjects in this experiment sat in a comfortable chair in front of a computer screen. At the beginning of a trial, a fixation cross appeared on the white screen. After 500 ms, an image in the form of one or two arrows pointing different directions corresponding to one of six classes appeared and stayed on the screen for 4 s. The relation between images and MI commands are showed in Figure 1(a). The image prompted the subjects to perform the MI task as requested until the fixation cross appeared again after the appearance of the image. The subject had a short break for 2.5 s and waited for the next trial. The paradigm is illustrated in Figure 1(b).

Multichannel EEG amplifiers with 64 channels band-pass filtered between 0.05 and 200 Hz and sampled at 500 Hz were used to record the EEG, whose montage is shown in Figure 1(c). Horizontal and vertical EOG signals were recorded to check for eye movements, which were not used for classification.

**2.1.2. Preprocessing.** During the process of preprocessing, we removed noise caused by eye and muscle movement and select effective rhythm (sensorimotor rhythm) in order to enhance the relevant information. Initially, we rereferenced the data. The left mastoid was chosen as reference electrode during the collection. In order to avoid laterality bias in the data, we rereferenced the data offline by changing the reference electrodes into the average of value of the left and right

mastoids. We removed contamination from bad channels by using the average of channels around the bad channel to replace them. Then, the whole time series of EEG data was band-pass filtered in 8-30 Hz, as written in [23]. We ran ICA to reject the EOG artifacts. We extracted epochs from the whole time segment located from 0.5 s before instructing the subject to perform MI to 2 s after instructing. Signals were baseline corrected over the interval 0-500 ms before instructing. Finally, the processed data was visually screened to discard any noise trials.

**2.2. Method.** In this subsection, we show the total scheme of our algorithm and detailed description of every essential component.

**2.2.1. The Total Scheme of the Algorithm.** The total scheme of the algorithm is organized as described in Figure 2, which can be divided into model training and model test. The training of the model contains four components. First, the EEG signal is collected during the motor imaginary experiment, preprocessed, and divided into source data and target data according to our standard. Second, the original spatial patterns are constructed from source data based on the CSP algorithm. Third, with the fisher ratio algorithm, we transfer the original spatial filters into spatial filters for target data. At last, features extracted by filters are given as input to a Support Vector Machine (SVM) classifier [31, 32] for training the model.

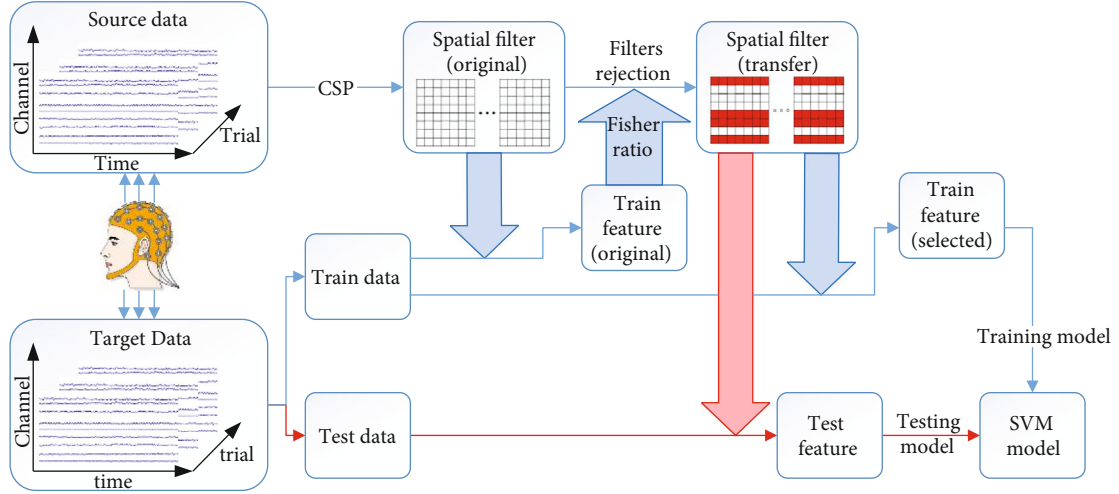


FIGURE 2: The total structure of our algorithm. The blue line stands for the training process, and the red line stands for the testing processing.

After the model has been trained, it can be used for the test data. The features of test data are also extracted by transfer spatial filters. And those features are used as input of classifier for the testing model.

**2.2.2. Data Transfer Standard on Division and Selection.** The preprocessed data will be divided into source data and target data. The source data comes from the old command, and target data from the new command. In order to improve the stability of the transfer model, we need to choose proper source data with the right label. In this section, we give our standard of choosing source data and the reason why we select those data depending on the target data. Initially, because of the intersubject variability of the spatial features between each subject, we select source data and target data from the same subject. And owing to the domain adaptation [18] standard in the transfer learning area, we choose the most similar data between source data and target data in terms of the imagery of body parts. For example, as for the classification of LH&F and RH&F, we use the dataset from LH and RH as source data. As for the classification of LH&RH and RH&F, we use the dataset from F and RH as source data. The standard for data selection will also be discussed in Section 4.4.

**2.2.3. Original Spatial Pattern Construction Based on CSP.** As for the classification of two distributions in a high-dimensional space, the CSP algorithm designs spatial filters, which maximize variance for one class and simultaneously minimize variance for the other class. Based on the simultaneous diagonalization of covariance matrices of two classes, spatial filters can lead to features whose variances are optimal for the discrimination.

In the traditional algorithm, in order to construct spatial filters extracting optimal features, the scale of training data is always large. By using source data, we can decrease the dependence on the training data. In order to make the most of the source data, we use source data for constructing the spatial filters. In the following section, we will take a particu-

lar classification as an example. We will use the source data (LH and RH) to construct the source spatial filters for the classification of target data (LH&F and RH&F).

The single-trial EEG signal from source data is represented as  $N \times G$  matrix  $X_c^i$ .  $N$  represents the number of channels, and  $G$  is the number of samples per channel.  $X_c^i$  is the  $i^{\text{th}}$  ( $i \in [1, K]$ ) trial of EEG signal matrix which belongs to the class  $c$  ( $c \in \{1, 2\}$ ). The class 1 refers to data from LH, and the class 2 refers to RH. The average spatial covariance matrix of each class can be calculated as

$$R_c = \frac{1}{K} \sum_i \frac{X_c^i (X_c^i)^T}{\text{trace}(X_c^i (X_c^i)^T)}. \quad (1)$$

$R$  is the sum of the covariance matrix from source data. And the composite spatial covariance matrix  $R$  can be factorized as

$$R = R_1 + R_2 = U \Lambda U^T. \quad (2)$$

$U$  is the matrix of eigenvectors, and  $\Lambda$  is the diagonal matrix of corresponding eigenvalues. Then, the whitening transformation is obtained as

$$P = \sqrt{\Lambda^{-1}} U^T. \quad (3)$$

$R_1$  and  $R_2$  from LH and RH are whitened as  $S_1$  and  $S_2$ .

$$S_1 = P R_1 P^T, S_2 = P R_2 P^T. \quad (4)$$

Then,  $S_1$  and  $S_2$  can be factorized as

$$S_1 = B \Lambda_1 B^T, S_2 = B \Lambda_2 B^T, \Lambda_1 + \Lambda_2 = I \quad (5)$$

$\Lambda_1$  and  $\Lambda_2$  are diagonal matrices.  $I$  is the identity matrix, and  $S_1$  and  $S_2$  share the same eigenvector. Therefore, this property makes the eigenvectors  $B$  effective for the

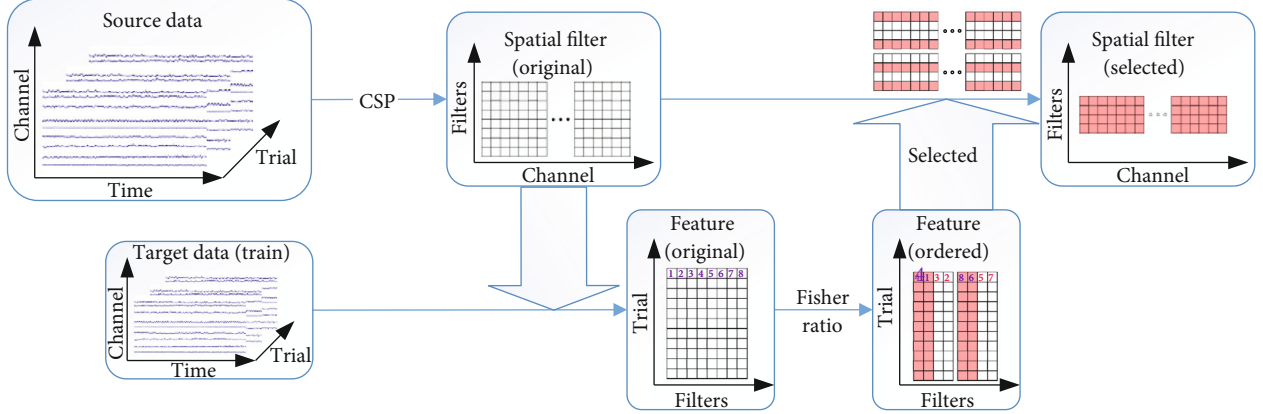


FIGURE 3: The transfer of spatial patterns based on fisher ratio. The features which are colored in pink refer to features in higher fisher ratio. And the filters in pink refer to selected filters.

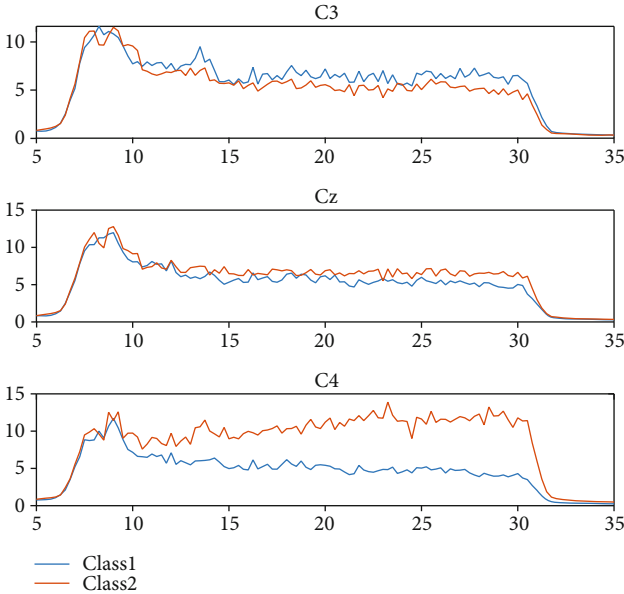


FIGURE 4: PSD analysis for imagining LH&F (class1) and RH&F (class2). The noticeable difference can be observed around 10-30 Hz in the average power spectrum at C3, C4, and Cz for two performed tasks. The blue line stands for the imagery of LH&F, and the red line stands for the imagery of RH&F.

discrimination of two classes.  $\hat{B}$  is the matrix of the  $m$  first and last eigenvectors in  $B$ . The spatial filters  $\omega$  can be calculated as

$$\omega = B\Lambda^T P. \quad (6)$$

**2.2.4. Pattern Transfer Based on Fisher Ratio.** The original spatial filters created from source data are in a large scale. There is no doubt that the original spatial filters are suitable for source data. However, not all filters can be used to extract features for target data. We need to select a valuable subset of filters from the original filters so as to transfer the source filters into target filters. In this section,

we briefly introduce pattern transfer based on fisher ratio. The filters we need can project target data into optimal features which have higher fisher ratio. The purpose of the fisher ratio is to find a subset of features, in which the distances between each data in different classes are as large as possible, while the distances between each data in the same class are as small as possible. Therefore, the fisher ratio is defined as the ratio of the variance between classes to the variance within classes. There is a dataset  $D$  which contains  $n$  samples ( $\omega_1, \omega_2, \dots, \omega_n$ ), which belongs to  $C$  classes. Each of the sample  $\omega$  has  $K$  features ( $x^k, k \in [1, K]$ ). There are  $n_i$  samples in one class.  $m_i^k$  is the mean value of  $x^k$  in one class. And  $m^k$  is the mean value of all  $x^k$ . Thus, the fisher ratio of the features in every dimensionality can be calculated as

$$J_k = \frac{S_{\text{between}}^k}{S_{\text{within}}^k},$$

$$S_{\text{between}}^k = \sum_{i=1}^C \frac{n_i}{n} (m_i^k - m^k)^2, \quad (7)$$

$$S_{\text{within}}^k = \frac{1}{n} \sum_{i=1}^C \sum_{x \in \omega_i} (x^k - m_i^k)^2.$$

The higher the fisher ratio of the feature is, the more discriminative the features between the two classes are.

Getting back to the pattern transfer, we think that the more optimal filter is, the higher the fisher ratio of feature is extracted. Consequently, to test the performance of each line of filters, we sort features generated through the projection of train data on original spatial filters. The score of each dimensionality in features calculated by the fisher ratio algorithm represents the performance of each spatial filter. Based on the principle of CSP, the original filters from source data should be divided into two groups which gives the larger eigenvalues of  $S_1$  and smaller eigenvalues of  $S_2$ , respectively. Consequently, the transfer filters should also be selected by the same amount in each group. We



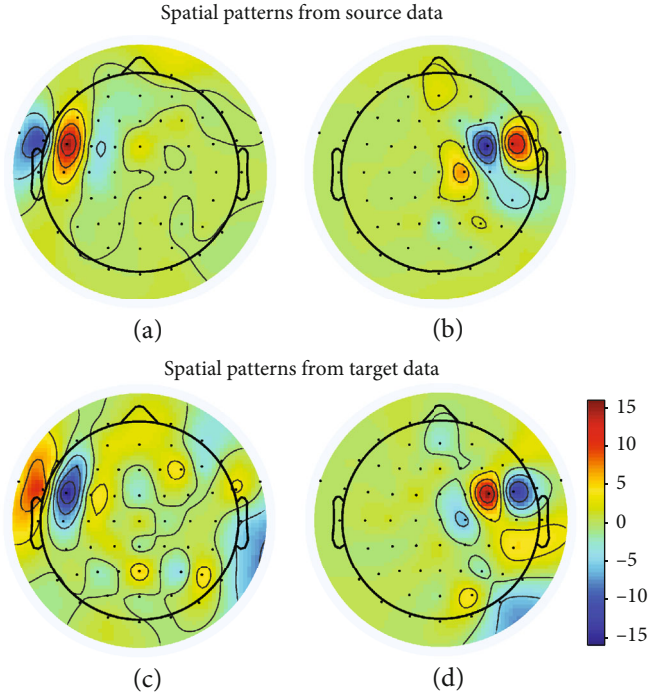


FIGURE 5: Spatial patterns of two most significant features in the transfer CSP and CSP algorithms, which were extracted from the 3rd subject's imagery of LH&F and RH&F. (a, c) are spatial patterns of LH&F and (b, d) are spatial patterns of RH&F.

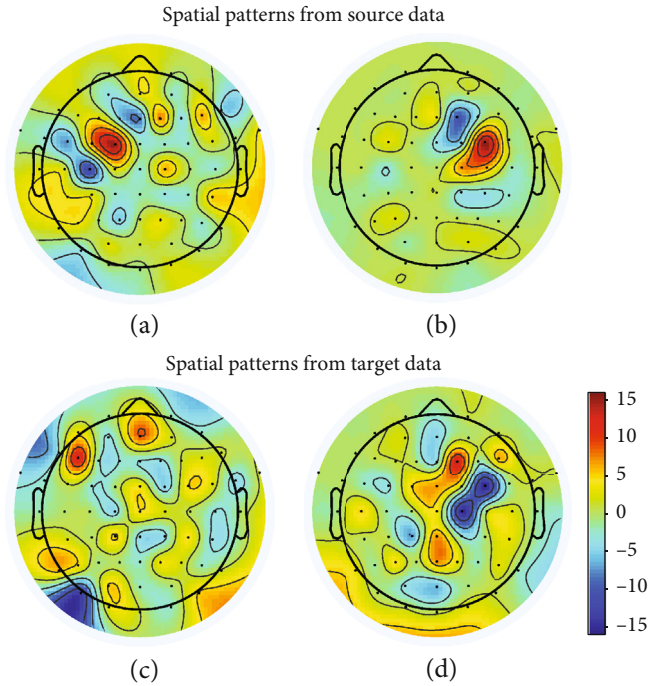


FIGURE 6: Spatial patterns of two most significant features in the transfer CSP and CSP algorithms, which were extracted from 1st subject's imagery of LH&F and RH&F. (a, c) are spatial patterns of LH&F and (b, d) are spatial patterns of RH&F.

divide the original spatial filters into two equal groups, which, respectively, are generated from the first and last eigenvectors through CSP. Then, we select the same

amount of filters, which have a higher score in each group, and combine them as transfer filters. The transfer of filters is described as Figure 3

TABLE 1: The average classification accuracy and accuracy range of three algorithms on the target dataset.

	Transfer CSP	Average accuracy (max-min) CSP	PSD
Sub 1	70.13% (79.58%-66.25%)	56.47% (70.83%-48.33%)	49.95% (56.25%-40.42%)
Sub 2	85.18% (91.43%-69.05%)	78.85% (92.38%-63.81%)	56.35% (70.00%-39.05%)
Sub 3	95.93% (98.79%-81.52%)	91.57% (99.39%-74.24%)	74.32% (88.79%-46.06%)
Sub 4	62.16% (78.61%-42.50%)	54.60% (69.72%-42.50%)	49.31% (56.39%-42.22%)
Sub 5	96.29% (99.39%-84.85%)	98.86% (100.00%-96.36%)	74.83% (91.82%-51.52%)

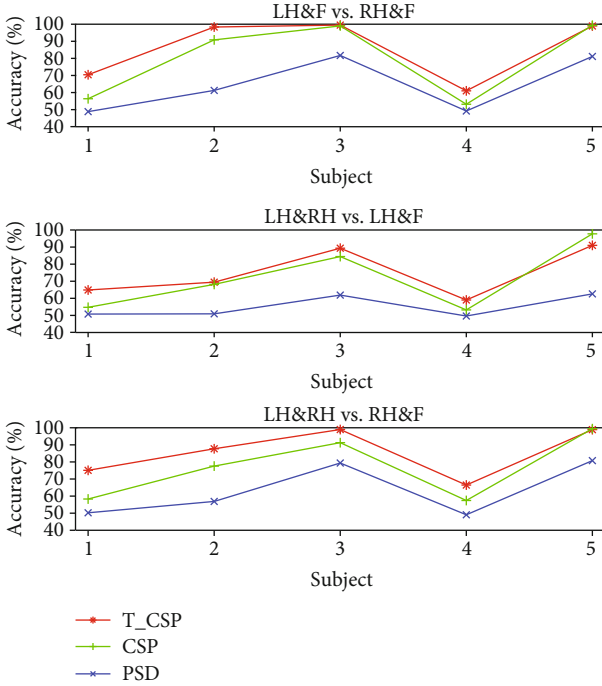


FIGURE 7: The average accuracy of classification on three different tasks (LH&amp;F vs. RH&amp;F, LH&amp;RH vs. LH&amp;F, and LH&amp;RH vs. RH&amp;F). The different color of lines refers to different algorithms. The x-axis refers to the number of subjects.

**2.2.5. SVM for Classification.** Before the classification, we use our transfer spatial filters for extracting features. With the transfer spatial filters  $\omega'$ , single-trial EEG signal  $X_c^i$  can be calculated as

$$Z = \omega' X_c^i. \quad (8)$$

The feature vector  $f_p^i$  can be calculated as

$$f_p^i = \log \left( \frac{\text{var}_p^i}{\sum_{p=1}^{2m} \text{var}_p^i} \right). \quad (9)$$

$\text{var}_p^i$  is the variance of  $p^{\text{th}}$  row in  $Z$  matrix.

Aiming at finding decision boundary between the class samples, SVM is an algorithm which can effectively prevent the defects of traditional classification algorithm, such as overtraining, dimension disaster, and local minima [31, 33].

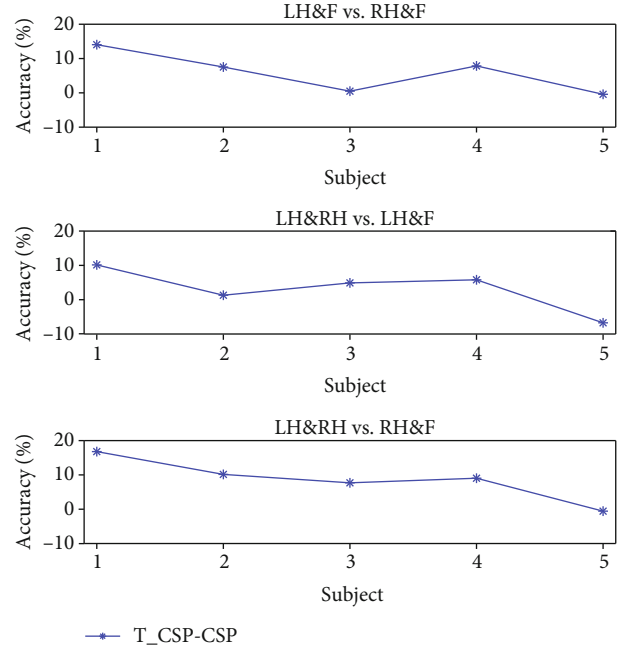


FIGURE 8: The difference of accuracy between our algorithm and CSP algorithm, the increase of accuracy performs remarkably in the 1st and 4th subjects whose data quality relatively worse in all of the subjects.

The SVM classifier maximizes the distance between decision boundary and margin. Due to the strength of the SVM classifier, it has been widely used in BCI classification [34, 35].

We utilized the entire 60 trials of source data from two selected classes which have been introduced in the previous section and 5 trials each class of target data for constructing spatial filters. In the selection of filters based on the fisher ratio, we chose the top 4 in the original filters which have higher scores. Then, we transferred the previous 5 trial target data into training features with the selected spatial filter. Combining with corresponding class information as the training label, we trained the classifier model.

As for the SVM classifier, the Gaussian radial basis function is used as the kernel function, and a five-fold cross-validation is used to choose suitable parameters for the testing data.

### 3. Results and Discussion

The performance of our algorithm in our dataset is compared with two popular algorithms, which are CSP and power



TABLE 2: The accuracy of classification on three different kinds of tasks (both old commands, both new commands, new commands and old commands) using the same source data.

Subject	LvsR	L&FvsR&F	FvsL&R	FvsR&F	FvsL&F	LvsL&R	LvsR&F	LvsL&F	RvsL&R	RvsR&F	RvsL&F
Sub1	61.9%	63.9%	69.5%	53.2%	65.1%	56.0%	63.4%	52.4%	63.2%	49.3%	67.6%
Sub2	98.6%	98.6%	68.9%	65.4%	81.9%	98.3%	98.5%	49.9%	61.9%	49.6%	99.6%
Sub3	99.1%	99.6%	70.2%	69.5%	86.3%	98.9%	98.7%	52.3%	69.0%	53.0%	99.8%
Sub4	63.1%	61.0%	71.9%	51.8%	67.9%	54.5%	62.2%	48.7%	63.3%	48.3%	64.0%
Sub5	100.0%	99.9%	50.2%	97.3%	99.9%	99.3%	99.7%	49.7%	97.3%	54.5%	100.0%

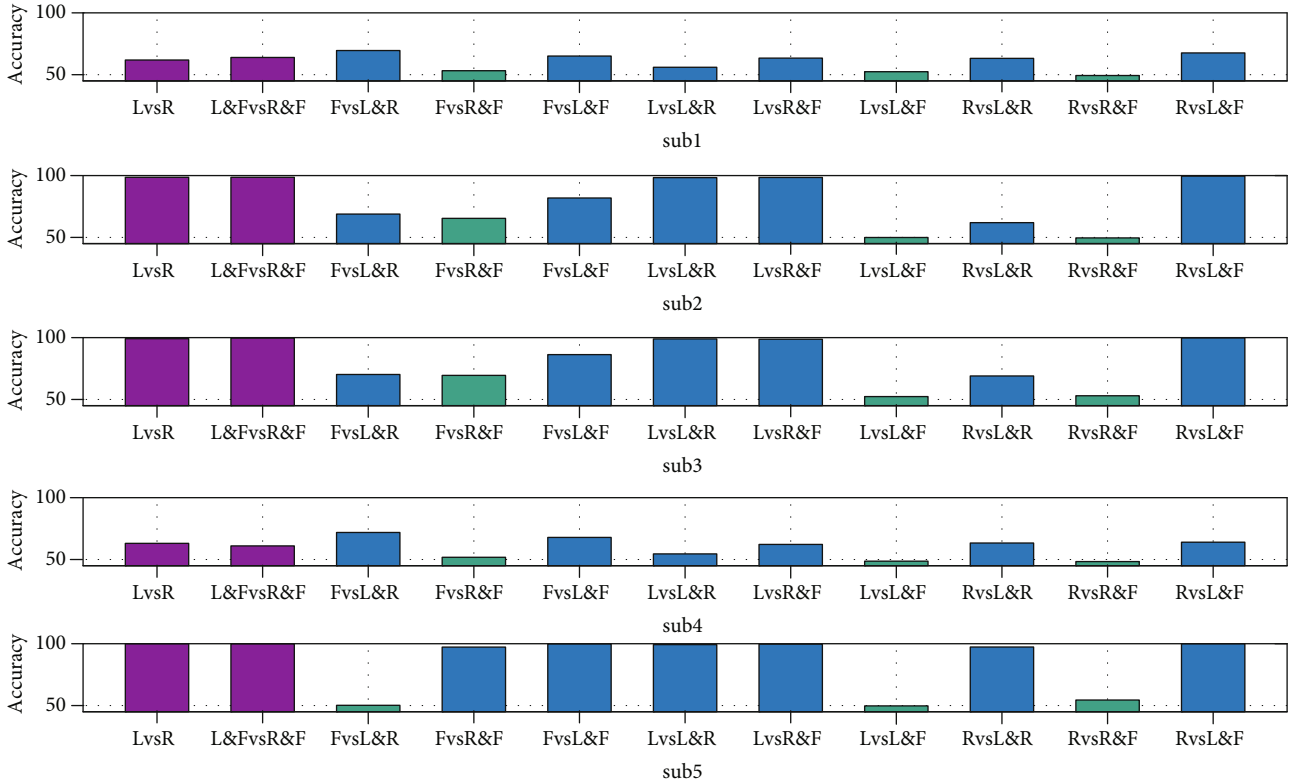


FIGURE 9: The accuracy of classification on three different kinds of tasks (both old commands, both new commands, new commands and old commands) using the same source data. The  $x$ -axis refers to different groups of tasks. The  $y$ -axis refers to accuracy. The bars in purple mean the first and second kinds of classification, and the rest refers to the third group. The green bars mean classification which has lower accuracy.

spectrum density (PSD) [36]. We utilize each algorithm to extract features from the testing set of the target data. In the CSP algorithm, we use spatial filters generated from target data. As for PSD, we use power spectral density values from one-way EEG data at particular channels (C3, C4, and Cz). We also calculate averaged PSD for the three channels. In order to test the robustness of the algorithm, we use the 12-CV 5 times on the target dataset of each subject. In order to decrease the scale of transfer data, in the cross-validation, we use 1/12 as training data and 11/12 as test data. We will compare the average accuracy of all algorithms. And the comparison of the different discriminative patterns from different algorithms will be showed in the following part. Besides, in order to make the comparison more particularly, the spatial patterns have been illustrated by focusing on the particular brain region.

**3.1. PSD Analysis for the Dataset.** In order to test the availability of the EEG dataset, PSD features at particular channels in the motor cortex (C3, C4, and Cz) are calculated based on temporal Fourier transform for each trial of EEG data. Noticeable differences can be observed in the averaged PSD for two target tasks. Imagining the LH&F movement leads to the decrease of alpha and beta bands' power at C4 and Cz channels and the increase at C3, whereas the contrary phenomena occur during imagining RH&F movement. The phenomenon from PSD analysis confirms the description of ERD/ERS. Figure 4 shows the average power spectrum at channel C3, Cz, and C4 for 3rd subject, evident differences are presented between the two tasks, especially around 10-30 Hz.

**3.2. Spatial Patterns Illustration.** In order to show the superiority of transfer spatial patterns, we visualize the spatial

patterns from our algorithm and traditional CSP algorithm and compare the visualization of them. The visualization of the two most discriminative spatial patterns extracted through two algorithms from the 3rd and 1st subjects is illustrated in Figures 5 and 6. In both figures, the spatial patterns in the first row are generated from our algorithm, and the spatial patterns in the second row are generated from CSP. As for the two subjects, the classification accuracies of 3rd subject for both algorithms is higher than the 1st subject. The accuracies for the 3rd subject in our algorithm and CSP are, respectively, 99.09% and 98.18%, which has a high accuracy in both algorithms and the optimization of the algorithm can only cause a little increase. As for the 1st subject, the accuracies are, respectively, 69.17% and 56.67%, which has a lower accuracy in the CSP algorithm but has a great increase in accuracy caused by the transfer algorithm. In Figure 5, the visualization shows that the highest discriminative weights can be viewed at FC5, FC6 channels from both patterns of our algorithm and CSP. The physiologically evidences from both our algorithm and CSP are similar to the result showed in PSD analysis. In Figure 6, the visualization shows that the patterns from our algorithm are clearer than patterns in CSP, which is fitter for the result in PSD comparing to the traditional algorithm. The optimization in spatial pattern also causes the increase in classification.

**3.3. Classification Results on the Testing Set.** The average classification results for all algorithms are listed in Table 1 and showed in Figure 7. As for every task of every subject, we run 12-CV for 5 times and collect the average accuracy and the range of the accuracy. Because of the different quality of data from each subject, the average accuracy from each subject in each task varies in a large range. However, we find that our algorithm performs better than two baseline algorithms on most subjects. In addition, it performs much better as for the dataset which performs badly in motor imagery tasks. Based on the accuracy in traditional algorithms, we find that the accuracies of the 1st and 4th subjects are relatively lower in all of the accuracies, which means the 1st and 4th subjects perform worse in all of the subjects. The difference on the accuracy between our algorithm and traditional algorithm is shown in Figure 8. We can find that the increase on accuracy of the 1st and 4th is more remarkable compared to all of the subjects.

**3.4. Classification Results on Different Task.** In order to check the performance of the algorithm when classifying different tasks, we use the spatial pattern transfer algorithm on different command tasks. We use data from LH and RH as source data for three different kinds of target data classification, which are target data from old commands, new commands, and both old and new commands. Aiming at the source data, we choose the LH vs. RH as the first group, and LH&F vs. RH&F as the second group. As for the third group which is the classification between old command and new command, we make the classification for each two groups which means LH vs. LH&F, LH vs. RH&F, LH vs. LH&RH, RH vs. LH&F, RH vs. RH&F, RH vs. LH&RH, F vs. LH&F, F vs. RH&F, and F vs. LH&RH. The classification results are listed in Table 2

and showed in Figure 9. Comparing the accuracy of all tasks using the same source data, we find that our spatial filters can classify both old commands and both new commands in a similar accuracy. However, as for the classification between new commands and old commands, the accuracy is seriously influenced by the similarity of the MI tasks. According to the comparison of accuracy in the second group, we find that each of the new commands has one type of old commands which is really similar and should be used as source data for the classification of new commands. For example, in the comparison of the LH vs. LH&F, RH vs. LH&F, and F vs. LH&F, we find that LH vs. LH&F caused a really low accuracy which contrast the accuracy of RH vs. LH&F and F vs. LH&F.

## 4. Conclusion

In order to add more MI commands, we add the combination of traditional MI commands into the BCI system. To reduce the time for researchers spending on collecting data and training model for new commands, this paper presented a new algorithm for feature extracting based on transfer learning. On one hand, it reduces the long calibration time through making the most of existing data and decreasing the number of training data in new MI commands. On the other hand, it increases the accuracy of the classification on less training samples in the BCI system. Our algorithm solves the problem of long calibration time and lack of MI commands, as well as increases the accuracy. Furthermore, because of its outstandingly increasing performance in subjects who have low accuracy in traditional CSP classification, it is helpful for subjects who cannot perform well in the BCI system. Our algorithm makes the BCI system more user-friendly for subjects. There is no need for subjects to train for a long time in order to adapt the MI tasks.

We compared our algorithm and traditional algorithms on the dataset collected by our laboratory. The spatial patterns from our algorithm are more physiologically reasonable in bad performing subjects. And our algorithm performs better than traditional algorithms.

In summary, our algorithm is better than traditional algorithms especially for subjects with less training samples and poor performances in traditional algorithm. With good performance and stability, our proposed algorithm can reduce the need for the training samples of new MI command by transferring the source data from the old MI command.

The method of spatial patterns is suitable for discriminating between 2 classes. Through changing the multiclass classification into binary classes, our algorithm can also be used in multiclass problem. In the future, with more state-of-the-art feature extractor from our algorithm, we can utilize it on feature extraction of deep transfer learning. With the help of a deep learning algorithm and the optimal features, the performance of classification may become better.

## Data Availability

The data used to support the findings of this study are not publicly available due to technology policy of Tongji

University but are available from the corresponding author upon reasonable request.

## Conflicts of Interest

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

## Authors' Contributions

Xuanci Zheng, Jie Li, and Hongfei Ji designed the whole research. Xuanci Zheng and Lili Duan performed the research. Xuanci Zheng, Jie Li designed the algorithm. Xuanci Zheng made data analysis. Xuanci Zheng and Maozhen Li wrote the manuscript. Jie Zhuang, Tianhao Gao, Rongrong Lu, and Zilong Pang made contribution after receiving reviews of the manuscript. Jie Zhuang, Rongrong Lu and Tianhao Gao helped refine the data analysing. Zilong Pang helped refine the code. Because of authors' negligence, Jie Zhuang, Rongrong Lu, Tianhao Gao and Zilong Pang was forgotten to be added into the author list before acceptance.

## Acknowledgments

This work was supported in part by the Science and Technology Commission of Shanghai Municipality under Grant 18ZR1442700, in part by Shanghai International Science and Technology Cooperation Fund 19490712800, and in part by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (No. TP2018056).

## Supplementary Materials

The supplementary material contains the code of the algorithm in the manuscript. The detailed description of the code can be seen in the README.md file in the RAR file (*Supplementary Materials*)

## References

- [1] F. Fahimi, Z. Zhang, W. B. Goh, T. S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI," *Journal of Neural Engineering*, vol. 16, no. 2, article 026007, 2019.
- [2] F. Pichiorri, G. Morone, M. Petti et al., "Brain-computer interface boosts motor imagery practice during stroke recovery," *Annals of Neurology*, vol. 77, no. 5, pp. 851–865, 2015.
- [3] K. K. Ang, C. Guan, K. S. G. Chua et al., "Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 5549–5552, Buenos Aires, Argentina, 2010.
- [4] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomedical Signal Processing and Control*, vol. 31, pp. 398–406, 2017.
- [5] A. Nourmohammadi, M. Jafari, and T. O. Zander, "A survey on unmanned aerial vehicle remote control using brain-computer interface," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 4, pp. 337–348, 2018.
- [6] C. N. Munyon, "Neuroethics of non-primary brain computer Interface: focus on potential military applications," *Frontiers in Neuroscience*, vol. 12, p. 696, 2018.
- [7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [8] D. T. Bundy, L. Souders, K. Baranyai et al., "Contralesional brain-computer interface control of a powered exoskeleton for motor recovery in chronic stroke survivors," *Stroke*, vol. 48, no. 7, pp. 1908–1915, 2017.
- [9] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013.
- [10] G. Pfurtscheller and F. H. L. da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.
- [11] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 7, no. 4, pp. 413–419, 1999.
- [12] G. Pfurtscheller, "Graphical display and statistical evaluation of event-related desynchronization (ERD)," *Electroencephalography and Clinical Neurophysiology*, vol. 43, no. 5, pp. 757–760, 1977.
- [13] J. Jin, S. Li, I. Daly et al., "The study of generic model set for reducing calibration time in P300-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 3–12, 2020.
- [14] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask learning for brain-computer interfaces," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 17–24, Chia Laguna Resort, Sardinia, Italy, 2010.
- [15] M. Krauledat, M. Tangermann, B. Blankertz, and K. R. Müller, "Towards zero training for brain-computer interfacing," *PLoS One*, vol. 3, no. 8, p. e2967, 2008.
- [16] W. Tu and S. Sun, "A subject transfer framework for EEG classification," *Neurocomputing*, vol. 82, pp. 109–116, 2012.
- [17] F. Lotte, A. Van Langenhove, F. Lamarche et al., "Exploring large virtual environments by thoughts using a brain-computer interface based on motor imagery and high-level commands," *Presence: teleoperators and virtual environments*, vol. 19, no. 1, pp. 54–70, 2010.
- [18] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Computational Intelligence Magazine*, vol. 11, no. 1, pp. 20–31, 2016.
- [19] V. Morash, O. Bai, S. Furlani, P. Lin, and M. Hallett, "Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries," *Clinical Neurophysiology*, vol. 119, no. 11, pp. 2570–2578, 2008.
- [20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [21] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial eeg classification in a

- movement task,” *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
- [22] C. Guger and H. Ramoser, “Real-time eeg analysis with subject-specific spatial patterns for a brain-computer interface (bci),” *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 447–456, 2000.
  - [23] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
  - [24] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “Regularized common spatial patterns with generic learning for EEG signal classification,” in *2009 Annual International Conference of the IEEE Engineering in medicine and biology society*, pp. 6599–6602, Minneapolis, MN, USA, 2009, September.
  - [25] H. Lu, H. L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, “Regularized common spatial pattern with aggregation for EEG classification in small-sample setting,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.
  - [26] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, “Correlation-based channel selection and regularized feature optimization for MI-based BCI,” *Neural Networks*, vol. 118, pp. 262–270, 2019.
  - [27] W. Samek, F. C. Meinecke, and K. R. Müller, “Transferring subspaces between subjects in brain-computer interfacing,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 8, pp. 2289–2298, 2013.
  - [28] H. Kang and S. Choi, “Bayesian common spatial patterns for multi-subject EEG classification,” *Neural Networks*, vol. 57, pp. 39–50, 2014.
  - [29] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford university press, 1995.
  - [30] T. N. Lal, M. Schroder, T. Hinterberger et al., “Support vector channel selection in BCI,” *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1003–1010, 2004.
  - [31] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
  - [32] K. P. Bennett and C. Campbell, “Support vector machines: hype or hallelujah?,” *Acm Sigkdd Explorations Newsletter*, vol. 2, no. 2, pp. 1–13, 2000.
  - [33] A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: a review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
  - [34] J. Atkinson and D. Campos, “Improving bci-based emotion recognition by combining eeg feature selection and kernel classifiers,” *Expert Systems with Applications*, vol. 47, pp. 35–41, 2015.
  - [35] H. Xu, W. Song, Z. Hu, C. Cheng, and J. Zhang, “A speedup SVM decision method for online EEG processing in motor imagery BCI,” *10th International Conference on Intelligent Systems Design and Applications, ISDA*, 2010, Cairo, Egypt, 2010, 2010.
  - [36] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms,” *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

## Research Article

# DRG-Oriented Mathematical Calculation Model and Method of Integrated Medical Service Cost

**Xiaowei Sun and Yi Zhu** 

*Jinling Hospital Department of Pharmacy, Nanjing University, School of Medical, Nanjing, 210002 Jiangsu, China*

Correspondence should be addressed to Yi Zhu; [sundayup686@163.com](mailto:sundayup686@163.com)

Received 29 June 2020; Revised 28 August 2020; Accepted 28 September 2020; Published 2 November 2020

Academic Editor: Yi-Zhang Jiang

Copyright © 2020 Xiaowei Sun and Yi Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the context of the new round of medical and health reform, in order to alleviate the problem of “difficult to see a doctor and expensive to see a doctor,” the state focuses on reducing the cost of medical services, so it puts forward the calculation and method research of medical costs. The purpose of this study is to calculate and predict the cost of medical services in a DRG-oriented integrated environment. In this study, activity-based costing and weighted moving average methods are used. First, basic data of medical services are collected, then all medical activities are confirmed and all service costs are collected, then a cost database is established, and a calculation model of medical costs is designed. Finally, calculation suggestions and optimization methods are put forward by analyzing the calculated data. The experimental results show that the actual demand of drugs predicted by the general moving average method is relatively insufficient, with the maximum error of 41%, the minimum of 5%, and the average error of 19.8%; the maximum error of drug demand predicted by the weighted moving average method is 24%, the minimum is 2%, and the average is 15.4%. It can be concluded that the prediction effect of the weighted moving average method is better than that of the ordinary moving average method, which plays a good and effective role in the prediction of medical cost. The activity-based costing method is more detailed and organized for the cost calculation and classification of medical services. It provides a certain value for the effective management and control of medical service cost.

## 1. Introduction

In order to restrain the sharp increase of medical expenses, the method of DRG payment has been adopted internationally. DRG will change the payment based on service items to the payment based on disease types. According to the international disease classification, inpatients are divided into several groups according to diagnosis, etc., to develop the corresponding payment benchmark. What is lower than the price of DRG is the “free profit” obtained by the hospital, and vice versa is the loss borne by the hospital. Therefore, how to effectively manage and control DRG-oriented medical service costs has always been the focus of scholars at home and abroad.

From the existing research, generally speaking, cost prediction and optimal control of medical expenses are limited to individual discussion. The management of medical expenses cannot be dealt with from the overall point of view.

At the same time, the existing method and theory of DRG direction of medical service cost management support are insufficient. From a comprehensive perspective, this paper studies the DRG-oriented cost management strategy of medical services, the construction of DRG-oriented general cost management system of medical services, the estimation and optimization control of disease cost, the determination and control of hidden cost, and the total cost of ownership of medical products, further improving the DRG-oriented medical service cost management development of medical service cost management methods and technical system control and medical safety cost management strategy.

About the calculation method of medical service cost, based on the mathematical theory, Szekely proposes a new method to use Benford’s law. They carefully check the data at the personal level to determine the specific cost for in-depth analysis. His method is to expand the mathematical theory and prove that large-scale data conforms to Benford’s



law. Then, they used the actual large-scale medical data of the national patient sample (NPS) of Korea health insurance assessment (HIRA) to test its applicability. For Benford's law, they considered the mean absolute deviation (MAD) formula to test large-scale data; this method cannot predict the budget cost effectively [1]. Hoyle believes that classification systems such as DRG and ICD are the basis of modern medical service planning, capital, operation management, and governance. Existing classification systems, such as ICD, DRGs, and ICF, have evolved for different reasons and reflect unique perspectives on health, disease, care, and "patient therapy." They proposed a practical method to add patient-centered "dimensions" to the existing classification. For example, they considered the meaning of comprehensive nursing classification from the perspective of ICD/DRG. By adding a dimension to closely link the continuous and stable patient attributes with service integration, they show that it is possible to develop stable and meaningful patient-centered categories across time and space, so as to support patient-centered integration by forging these connections, and the broader nursing integration projects can focus on maximizing and improving the patient's interests rather than being distracted by institutional boundaries. The implementation of this method needs an ideal external environment and various cooperation, and its enforceability is poor [2]. Hardin et al. are mainly engaged in BPCI and cooperate with a group of medical diagnosis-related groups (DRGs). They describe the experience of creating a response system for different personnel and diagnosis in medical DRGs and specifically determine the organizational factors to achieve the successful implementation of bundling. Their experience shows that the interprofessional collaboration program is successful, although it is still in its early stages, and the information observed from their project strategies may provide potential insights for organizations considering participating in BPCI programs. This method is vague and inaccurate in the calculation of medical cost [3].

This study first gives an example of some problems in the current cost management of medical service projects. At the same time, the key technology of DRG-oriented integrated medical service cost management and the composition and classification of medical cost are described. This study also introduces the principle of activity-based costing and the principle of single disease cost prediction based on a support vector machine. Through this principle and weighted moving average method to collect and process data, through the model and calculation method to effectively analyze and predict the medical service cost, the existing scheme is further optimized, so as to better promote the development of DRG-oriented medical service cost management.

## 2. Medical Service Cost and Mathematical Calculation Method

*2.1. Main Problems in Cost Management of Current Medical Service Projects.* At present, the cost management system of many hospitals is relatively chaotic, by which the refining cost management needs of medical service projects are difficult to meet [4].

*2.1.1. The Management Did Not Pay Attention to the Cost Management of the Medical Service Project.* The cost management of a medical service project is a comprehensive, complex, and human-oriented project. Only when the administrator invests more energy can the unified work instruction be established and play the role of cost management. At present, the hospital management focuses on the calculation of income and cost, but due to the neglect of the supervision of events, the lack of understanding of the control function of management accounting, and the lack of leadership in the cost management of medical service projects, the implementation of detailed cost management is not easy [5, 6].

*2.1.2. Weak Cost Management Foundation of Medical Service Project.* First, there are insufficient cross-border cost management talents to meet the needs of industry and financial integration. Although the financial director of the hospital has accounting knowledge, he only pays attention to the basic cost calculation and accounting function. Due to the lack of management and evaluation awareness and function of management accounting, the meticulous management of hospital expenses is limited. Secondly, the support of the cost management information system is insufficient. Traditional manual methods cannot meet the needs of cost management; information technology plays an important role in the whole medical system. However, the current hospital information system is not perfect, there is no connection between the systems, and data dispersion, collection, and induction are not easy, which directly affects the accounting and control of medical service cost [7].

*2.1.3. The Cost Calculation of Current Medical Service Projects Is More Limited.* At present, the cost calculation method of the hospital has fallen behind, most of which are based on departments, which cannot reflect the total cost of the hospital as a whole, and has no contribution to the implementation of refined cost management. At the same time, the cost calculation of the department has fallen into a misunderstanding. First of all, as the benchmark of department bonus using cost calculation, the competition between departments becomes fierce. Secondly, no matter the renewal of machines, investment in scientific research, and other expenses, they will over pursue to reduce costs. Misunderstanding of cost calculation affects the long-term development of hospitals [8].

*2.1.4. Imperfect Cost Management and Guarantee Mechanism of Medical Consumables.* The cost management of medical consumables is one of the important links in the cost management of medical service projects. According to the survey, the cost of medical consumables in hospitals is increasing year by year, and the proportion of the total cost of medical services is also increasing. According to the price management department, the medical consumables with a unit price lower than 1000 yuan in public hospitals can be increased by 10%. For medical consumables over 1000 yuan, the price increase rate is 8%, and the maximum price increase is 800 yuan. If improper management results in the loss or



damage of consumables, the cost of the hospital will rise. Generally speaking, the actual use of consumables in the hospital has been confirmed clinically in the whole process. In fact, the intermediate monitoring link of functional departments does not work, which will lead to not only medical corruption but also waste, directly affecting the decision-making budget of the hospital. Therefore, for the detailed management of medical expenses, it is particularly important to strengthen the construction of a cost management guarantee system of consumables [9, 10].

## 2.2. Key Technologies of DRG-Oriented Integrated Medical Service Cost Management

**2.2.1. Disease Cost Prediction Method.** Disease cost prediction is the basis of reducing disease cost. The influencing factors of disease cost are distributed in all stages of the process from hospitalization to discharge. The complexity of the disease treatment process will bring many factors and difficult data to cost. At the same time, the importance of factors is very different. Each factor interacts and influences each other. The selection of factors affecting traditional cost is mainly based on experts, lacking theoretical guidance. The detailed estimation method used in traditional cost forecasting needs a lot of time and cost. The parameter prediction method is due to the generalization ability and insufficient adaptability of the model, and the accuracy of cost prediction is not good enough, which almost has no effect on reducing the cost of diseases. At present, a high level of cost prediction method is needed to predict the cost of diseases. The purpose of disease cost prediction is not only to clarify the process of disease cost but also to clarify the inherent law through the analysis of prediction cost and get improved feedback [11, 12].

**2.2.2. Disease Cost Control Method.** Disease cost management is an important part of DRG-oriented comprehensive medical service cost management. The rationality and scientificity of management methods not only directly affect the effect of hospital cost management but also determine the stability of a hospital operation system. The cost management of disease includes the whole process of disease design, implementation, and improvement, so the cost management of disease must also be implemented through the whole life cycle of disease.

**2.2.3. Disease Cost Prediction and Control.** The hospital's cost management system is a multioutput service system. The environment of medical service activities is complex, and the cost is affected by various factors. Therefore, there are various reasons for different costs. It is difficult to get satisfactory results with simple statistical tools. The existing disease cost management method usually uses the accounting period as the analysis period. Through the analysis of financial accounting indicators, the medical service process is monitored and controlled. This is usually afterwards management, which has a great impact on the timeliness of cost management. By introducing the idea of predictive control and combining the management chart with artificial intelligence

technology, the timeliness of disease cost management can be greatly improved [13].

**2.2.4. Cost Difference Identification of Diseases.** If we successfully predict and calculate the cost of different kinds of diseases, we need to enable the diagnostic function of different costs to distinguish different types of costs. From the point of view of diagnosis pattern recognition of disease cost difference, according to the principle of management chart, different patterns of disease cost can be divided into six categories: normal, rising trend, falling trend, rising stage, falling stage, and cycle mode. Because the determination of the cost difference of disease is combined with various experiences and knowledge of hospital information systems, it is necessary to make full use of real-time information and knowledge, determine the classification of the cost difference of disease, find out the cause of cost as soon as possible, and take timely countermeasures. The action is adjusting the medical service process to a controlled state [14, 15].

**2.2.5. Cost-Oriented Optimization Strategy of Medical Service Operation.** The diagnosis system of disease cost difference is only to predict and recognize the disease cost difference, and the specific scheme should be decided by the relevant medical personnel. The allocation of various resources in hospitals directly determines the cost level of diseases. Optimizing the allocation of resources plays an important role in reducing the cost of diseases. In order to cope with the surgical pressure caused by the disease payment system, the introduction of queuing theory of surgical research, mathematical planning and other theories, optimization of hospital facilities, and personnel allocation can greatly reduce the cost of disease [16].

## 2.3. Composition and Classification of Medical Cost

**2.3.1. Composition of Hospital Cost.** Hospital cost refers to the expenses of various medical services borne by the active TCM hospital. Whether the hospital adopts the former cost usage or the activity-based cost method, the total expenses of the hospital are divided into drug expenses and medical expenses. Medical expenses and material expenses are medical product and material expenses; intangible asset depreciation and fixed asset depreciation required for undertaking in the hospital are fixed asset depreciation and depreciation; wages and welfare expenses used in the hospital of traditional Chinese medicine for intangible assets and medical activities are 5 insurances and 1 fund for basic wages, subsidies, and bonuses, as well as other expenses borne by employees of all departments of the hospital; and management expenses are the daily business and hospital management expenses of the logistics department and management department of the hospital, including business expenses, transportation expenses, travel expenses, and business expenses. In terms of the total cost of the hospital, the proportion of drug and material costs is about 55%, the proportion of wages and welfare costs is about 25%, and the proportion of other costs is about 20%. Shown in Figure 1 is the composition framework of medical service cost [16, 17].

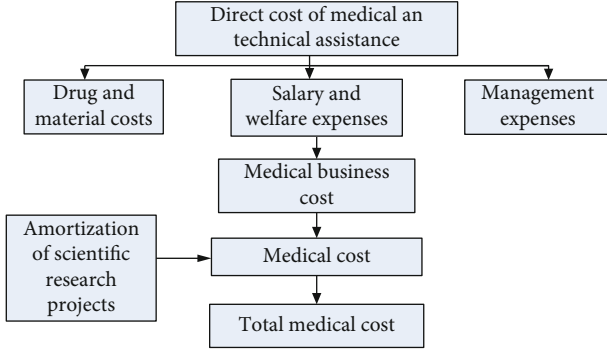


FIGURE 1: Structure of medical service cost.

**2.3.2. Classification of Hospital Cost.** According to different research purposes, hospital cost can be divided into disease cost, project cost, department cost, and hospital total cost. Disease cost refers to the treatment cost of a specific disease, and project cost refers to the expenses incurred in the medical service project. Department cost refers to the daily medical service process in the department to ensure normal operation. The total cost of the hospital refers to the total expenses borne by all departments in the daily operation [18, 19].

Cost can be divided into variable cost and fixed cost according to different forms. Variable cost refers to the total cost of the hospital's business cost and daily business process. The more business, the higher the total cost. The fixed cost has nothing to do with the total cost of the hospital's daily business process. The cost will not change with the increase in transaction volume. A credit card can be divided into indirect cost and direct cost. Indirect cost refers to the cost that cannot be directly included in the cost object. This cost is related to multiple cost objects and needs to be allocated according to the allocation basis, including interest payment, bad debt provision, and lease fee. Direct expenses refer to the expenses that can be directly included in the cost object, such as salary, medicine expenses, and welfare expenses of employees [20].

**2.4. Principle of Activity-Based Costing.** The method of activity-based costing is different from the former costing method. The traditional cost accounting method is limited by the single distribution standard of cost distribution. The allocation method of activity-based cost uses various allocation benchmarks and main principles such as mechanical time and labor force. Consumption resource is operation, product consumption operation, and operation is the intermediate variable between resource and product. In the hospital industry, specifically, patients are registered in the hospital or online, diagnosed, and examined in the outpatient process. These processes are the occurrence of operations, which consume specific resources, and the allocation of product cost based on resource motivation is consumed. The cost calculated by this method becomes more accurate and accurate, and the decision reference value of the manager points up. The activity cost method is widely used in finance, communi-

cation, medical, and other fields, and now, it is the main choice of cost management methods for enterprises [21, 22].

## 2.5. Principle of Single Disease Cost Prediction Based on Support Vector Machine

**2.5.1. SVM Algorithm.** A support vector machine (SVM) is a machine learning method based on statistical learning theory. According to the limited sample data information, the structural risk is controlled to the minimum, and the nonseparable data set is mapped to the high-dimensional feature space. In this way, samples can be distinguished correctly in high-dimensional space, so as to solve the problem of low-dimensional space [23]. The cost estimation of a single disease is related to many factors. It is assumed that  $x_i \in E$  is the factor influencing the cost estimation and  $y_i$  is the cost estimation. The steps of single disease cost regression prediction by SVM are as follows.

Set the training sample set of size  $L$ :

$$\{(x_i, y_i) | i = 1, 2, 3, \dots, l, x_i \in E, y_i \in R\}. \quad (1)$$

Hyperplane in high-dimensional space can be expressed as

$$\omega x + b = 0. \quad (2)$$

The idea of SVM for inferring regression function is to map the data in input space to high-dimensional feature space through nonlinear mapping  $x$  and carry out linear regression in high-order element space [24].

SVM uses the following estimation functions:

$$f(x) = \omega x + b. \quad (3)$$

SVM uses the SRM (structural risk minimization) criterion and uses the loss function of  $\varepsilon$  function to measure the risk, which is defined as

$$L(y, f(x, \delta)) = L(|y, f(x, \delta)|_\varepsilon). \quad (4)$$

The regression estimation problem is defined as minimizing the risk of nonresponse loss function ( $\varepsilon \geq 0$ ). If SVM is used to predict the cost of a single disease, it can be transformed into the minimum value problem of the next convex plan

$$\begin{aligned} \min \omega(a_i, a_i^*) &= 0.5(a_i - a_i^*)^T K(x_i, x)(a_i - a_i^*) \\ &+ \varepsilon \sum_{i=1}^l (a_i + a_i^*) + \sum_{i=1}^l (a_i + a_i^*) y_i. \end{aligned} \quad (5)$$

$$\text{Among } \sum_{i=1}^l (a_i - a_i^*) = 0,$$

$$0 \leq a_i, a_i^* \leq C, i = 1, 2, 3, \dots, l. \quad (6)$$

$K(x_i, x)$  refers to the problem of finding the vector  $\omega$  in a quadratic programming optimization.

In the case of nonlinear separability, for a specific kernel function, any sample in a specific sample group may be the support vector  $\omega$ . General kernel functions include polynomial kernel function, sparse generating kernel function, and radial basis kernel function.

Through the above work, for the cost of a single disease, the SVM regression prediction function can achieve the following results:

$$f(x) = \sum_{i=1}^l \omega^* K(x_i, x) + b = \sum_{i=1}^l (-a_i + a_i^*) * K(x_i, x) + b. \quad (7)$$

In order to use SVM to solve linear and nonlinear regression problems, it is necessary to determine the insensitive value, penalty factor  $C$ , and kernel function parameters. Choosing different kernel functions will produce different support vectors. This has a great influence on the accuracy of cost estimation. In order to make the RBF kernel function perform well in prediction accuracy and speed, this chapter uses radial basis kernel function RBF:

$$K(x_i, x) = \exp \left( -\frac{\|(x - x_i)\|^2}{2\sigma^2} \right). \quad (8)$$

Cost difference analysis and management chart theory also play an important role in disease cost management. The hospital managers and medical staff explained the difference of DRG cost from medical and economic aspects. The hospital is a prolific organization, and the decision of the clinician indirectly affects the cost of the hospital. The target cost of the disease is

$$C = \sum_{i,j} n \times p_i \times r_{ij} \times c_j. \quad (9)$$

The difference in disease cost is

$$Y = \sum_{i,j} (n \times p_i \times r_{ij} \times c_j - n' \times p'_i \times r'_{ij} \times c'_j). \quad (10)$$

**2.5.2. Mathematical Calculation Model and Method of Medical Cost. Simple moving average method:**

$$F_{t+1} = \frac{\sum_{i=t-n+1}^t A_i}{n}, \quad (11)$$

where  $F_{t+1}$  is the demand forecast of time period  $t + 1$ ,  $n$  is the number of time periods to calculate the moving average (here  $n = 4$ ),  $A_i$  is the actual demand in time period  $i$ , and  $w_i$  is the weight in time period  $i$  ( $\sum w_i = 1$ ).

**Weighted moving average method:**

$$F_{t+1} = \sum_{i=t-n+1}^t w_i A_i. \quad (12)$$

### 3. Calculation and Treatment of Medical Service Cost

**3.1. Data Collection and Sorting.** The top three collect and check the income data through the hospital information system and financial software. The hospital's expenses are divided into medical expenses, management expenses, and other expenses. In 2019, the total expenditure is 1773122500 yuan, including 15449825475 yuan for medical services and 223297025 yuan for management.

The medical service fee items include Western medicine, traditional Chinese medicine, Chinese patent medicine, and other medicine fee items. There are charging standards for separately charged materials such as cylinder, embedded consumables, vacuum blood sampling needle, oxygen fee, and blood fee, but there is no need for heating fee, diagnosis and treatment fee, incidental fee, application fee, diagnosis and treatment record fee, and external inspection and other medical fees.

As for the department of research and education, it usually does not provide medical services beyond the cost calculation range of pure research-based research institutions in hospitals. In the pharmaceutical sector, there is a special purchasing platform for pharmaceutical products, whose revenue is controlled by the difference between the purchase price and the sales price. The requested item of the pharmaceutical product is not in the cost calculation object. Departments providing services to hospital staff will provide simple diagnosis and prescription services to hospital staff, but this is not included in the cost calculation. As for the special diagnosis and treatment department, the registration fee of the special diagnosis and treatment clinic is different from the usual diagnosis and treatment fee. The diagnosis and treatment are also very simple, and there is no actual implementation project, but the data is very different when calculating the cost of such diagnosis and treatment department.

#### 3.2. Medical Cost Calculation Ideas

- (1) **Collection of Basic Data.** This is done by visiting the hospital, taking the staff of the hospital as the object to conduct a questionnaire survey, first learning the basic medical services of the hospital, then conducting a detailed interview of the hospital, deepening the understanding of the current cost management status and the existing problems of the hospital, and requiring data. Finally, the relevant cost data of the hospital will be collected by category in 2019.
- (2) **Confirmation Work.** The work flow chart of hospital B shows each process of the hospital, which is composed of multiple operations and summarizes the same operations.
- (3) **Aggregate Cost: Analyze the Aggregate Cost and Cost Driver.** In the previous cost methods, the current allocation benchmark is used to allocate the related resource cost through the inventory and payable wages recorded in the entry. Cost drivers are based

on the nature of resource costs and the types of operations, different for decision.

- (4) *Build a Cost Library.* A cost library is a collection of similar costs and cost drivers. A cost base can collect management fees and depreciation fees of fixed assets. The number of cost libraries is different from the previous cost accounting, which is the number of cost drivers. It is the benefit of ABC.
- (5) *Design Model.* After fully understanding the hospital, the mode of the hospital is mainly determined according to the accounting method of operating cost, including confirming the identification and classification of cost objects, resources, and operations; the relationship between different levels; the responsibility of cost objects, cost-driven products, and resources; and the circulation relationship between operations.
- (6) *Analysis Data.* Analyze the cost data of the hospital under the accounting of activity-based costing.
- (7) *Put Forward Suggestions.* Put forward relevant suggestions for the problems found in the process of activity-based costing accounting, such as the problem that the cost budget is only in the form and lack of corresponding reward and punishment measures, and improve the relevant organizational structure and establish reward and punishment measures for cost management.

#### 4. Calculation Method and Prediction Analysis of Medical Service Cost

*4.1. Prediction and Analysis of Medical Drug Cost by Moving Average Method.* Most medical institutions may have a big error between the future drug forecast and the actual demand. If the patient does not meet the drug demand, it will cause not only economic loss but also mental and physical injury. If the amount of one-time purchase is too much, it will not only restrict the flow and direction of funds but also lead to the overall rise of various related costs (the largest of all costs including procurement costs). Excess inventory is also easy to lead to the omission of custody and management.

The best way to forecast demand is to compare and analyze, collect and sort out the past consumption data (available and accurate) according to the four main factors (season, trend, cycle, and random) of forecast, follow this rule, and establish economic order quantity (EOQ) which is a single order quantity that can minimize the total order cost, storage cost (including holding cost), and economic loss caused by insufficient inventory. Only with accurate demand forecast can we realize the best combination of demand and supply on all links of the supply chain and reduce and avoid the phenomenon of affecting clinical business due to rebate, waste, or shortage.

From January 2019 to December 2019, the number of sales boxes of regional health service centers located in Dongcheng District of Beijing is as follows. It can be calculated by

the simple moving average method. The drug demand forecast is from January 2019. As shown in Table 1, the dynamic average method forecasts the demand for drugs.

Figure 2 shows the comparison between the actual demand and the predicted demand. As can be seen from Figure 2, the predicted demand for the drug was generally high in the first half of the year, while it was generally low in the second half of the year, and the highest demand reached 58 boxes in December, which may be caused by more diseases caused by severe cold weather. It can be seen from the trend and fluctuation of the error curve in the figure that there is still a little shortage in predicting the actual demand of drugs, with the maximum error of 41% and the minimum error of 5%, among which the average error of 19.8%. This result is relatively general for the prediction of drug use, and it can have a certain effect, but the effect is not good. This also makes the hospital have a certain deviation in the medical cost of purchasing drugs.

*4.2. Prediction and Analysis of Medical Cost by Weighted Moving Average Method.* Figure 3 shows the prediction of drug demand by the weighted moving average method.

It can be seen from Figure 3 that after the weighted moving decomposition method for drug demand prediction, the result is relatively good. It can be seen from the curve trend that the fitting degree of actual demand and forecast demand curve is quite high, especially at the beginning of the year with high demand, and in the year with low demand, the forecast effect is good. From the error curve, the highest is 24%, the lowest is 2%, and the average is 15.4%. It can be concluded that the prediction effect of this method is better than that of the ordinary moving average method. It plays a good role in the prediction of medical cost.

When the weighted moving average method is used to predict the result, the value of the tracking signal is 55.2%. Such a big mistake may be due to improper weight selection, so the weight will be reselected for calculation, but the result will be further deteriorated. The reason for this big mistake is that it is not appropriate to select the weight, or after selecting parameters from the horizontal direction, you need to specify the weight separately.

*4.3. Calculation and Analysis of Medical Cost by Activity-Based Costing.* The business cost, indirect cost, health material cost, and other expenses included in the inspection activity cost are related to the number of cases of medical service, so the number of cases of each medical item inspection can be calculated according to the number of cases. As shown in Figure 4, the distribution of health material cost is based on activity-based costing.

As shown in Table 2, the distribution table of health material cost is given.

It can be seen from Figure 4 that in the cost of health materials, other costs account for more than 7000 in total, and the cost of each of the seven projects is different, including the highest cost of obstetrics and the lower cost of thyroid and neck blood vessels. It can be concluded that activity-based costing is particularly clear in the calculation and statistics of medical service cost.



TABLE 1: Prediction of drug demand by the moving average method.

Month	Actual demand (box)	Forecast demand (box)	Error	Error rate
1	50	46	4	8%
2	36	48	-12	34%
3	45	47	-2	5%
4	34	48	-14	41%
5	29	40	-11	36%
6	18	24	-6	33%
7	31	38	-7	22%
8	45	40	5	11%
9	43	38	5	13%
10	38	34	4	9%
11	46	39	7	14%
12	58	44	14	25%

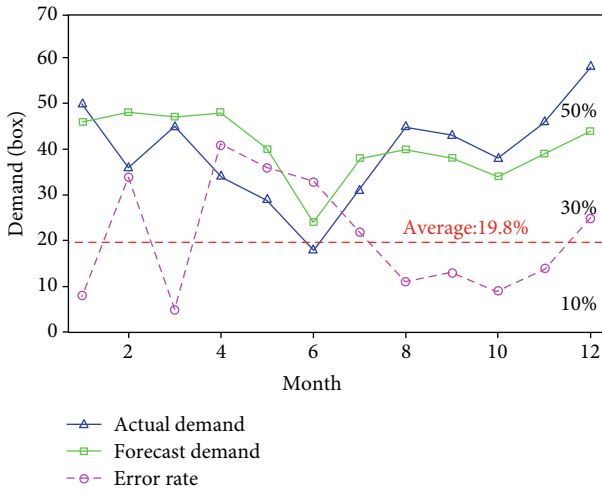


FIGURE 2: Comparison between actual demand and predicted demand.

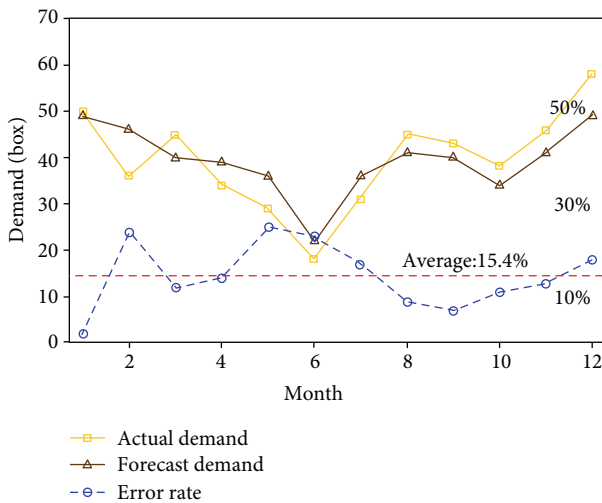


FIGURE 3: Prediction of drug demand by the weighted moving average method.

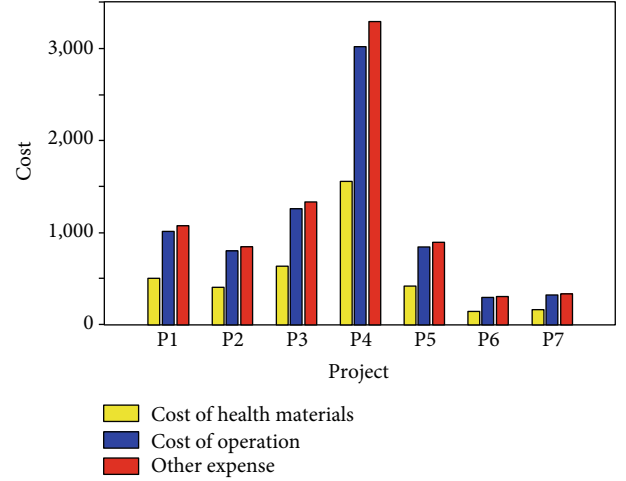


FIGURE 4: Distribution of health material cost based on activity-based costing.

**4.4. Comparative Analysis of Medical Cost Calculation Methods.** Based on the improved SVM data mining method, the DRG disease series cost prediction method, its objective data, prediction accuracy, and speed will become higher. The prediction results can be used as a reference value for the cost of new disease cases. Doctors not only modify the treatment plan to reduce the cost, so that the hospital can correctly grasp the future cost level and change trend, but also formulate the cost optimization control standard to provide a scientific and intuitive basis for the decision of hospital management. The model fully considers the multifactor, nonlinear characteristics of disease cost and the similarity of medical service composition of each case. According to the improved SVM proposed DRG disease series cost prediction model, it can be reused. After obtaining the new instance, we can directly use the influence factor information to predict the cost according to the specific value of the disease instance with the minimum condition attribute set in the model.

As shown in Figure 5, the ROC chart of the medical cost prediction algorithm is shown.

According to Figure 5, our algorithm is superior to the BP algorithm and SVM algorithm in the prediction of medical cost. The results show that this method is effective and excellent and can obtain effective knowledge which is directly used on the basis of knowledge. Therefore, this method has a wide range of application prospects in the DRG direction of disease cost prediction. It can provide an effective auxiliary means for the cost prediction of single disease and DRG disease series and has important theoretical and practical significance.

The prediction curve and actual curve using three methods are shown in the figure. The calculation time of the BP neural network, standard SVM algorithm, and white paper algorithm is 15.97 seconds, 10.32 seconds, and 5.71 seconds, respectively. Compared with the BP neural network and SVM algorithm, our algorithm can greatly improve the accuracy and speed of prediction, so it has a better application prospect in the field of DRG disease series cost prediction.

TABLE 2: Distribution of health material cost.

Project	Number of cases	Cost of health materials	Cost of operation	Other expenses
Digestive system (P1)	946	505.96	1011.92	10703
Urinary system (P2)	1532	398.32	796.64	8426
Gynaecology (P3)	2420	629.2	1258.4	13310
Obstetrics department (P4)	5978	1554.28	3108.56	32879
Breast (P5)	1624	422.24	844.48	8932
Thyroid (P6)	558	145.08	290.16	3069
Neck vessels (P7)	604	157.04	314.08	3322

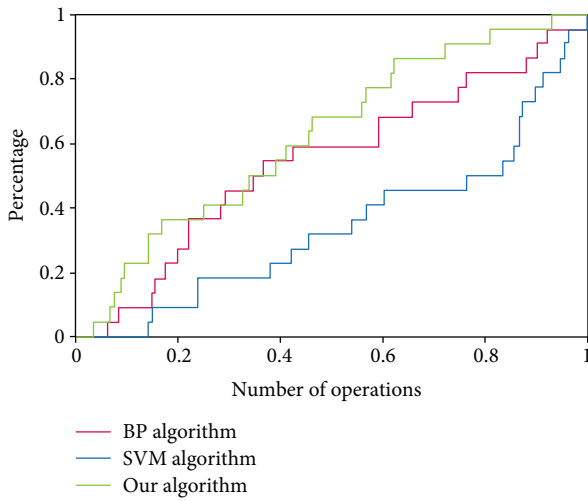


FIGURE 5: ROC chart of the medical cost prediction algorithm.

## 5. Conclusion

In order to give full play to the role of medical expense management in hospital management and solve the problems and defects in the implementation of DRG in China, this chapter combines the idea of comprehensive management to study the structure of DRG-oriented comprehensive medical expense management. This paper puts forward the management characteristics, DRG-oriented integrated medical service cost management mode, and its structure chart and investigates the characteristics and problems of disease cost prediction and control under the integrated medical service cost management mode in detail.

This study discusses the structure and function of DRG-oriented integrated medical service cost management, constructs the theoretical framework of DRG-oriented integrated medical service cost management mode, and constructs the framework of DRG-oriented integrated medical service cost management mode. The model describes the structure and characteristics of the system framework in detail. The basic theory of DRG-oriented comprehensive medical service cost management is complete. Due to the introduction of the DRG-oriented integrated medical service cost management mode for a relatively short time, the basis of previous research is relatively weak, especially the compo-

sition of the framework and the principles of the implementation framework, which needs further theoretical research.

In the DRG payment system, medical service cost management has become an important means for hospitals to successfully implement the DRG payment system and improve their economic and social interests. However, the current cost management of medical services is basically limited by cost calculation, and the scope of the management concept is very narrow. This study summarizes and analyzes the impact of DRG on the cost management of medical services and the main cost management issues of DRG in the process of hospital introduction in China. DRG is studied from the perspectives of cost, hidden cost, financing cost, and medical safety cost. The general medical cost management strategy promotes the development of DRG-oriented medical cost management, further improves and perfects the theory of medical cost management, and provides guidance for hospitals to improve the level of medical cost management.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

There is no potential competitive advantage in our paper.

## Authors' Contributions

All the authors have reviewed the manuscript and agreed to submit it to your journal.

## Acknowledgments

This work was supported by the Postdoctoral foundation of Jiangsu province.

## References

- [1] J. A. Park, M. Kim, and S. Yoon, "Evaluation of large-scale data to detect irregularity in payment for medical services," *Methods of Information in Medicine*, vol. 55, no. 3, pp. 284–291, 2018.
- [2] P. Hoyle, "Integrated categories for integrated care - the importance of patient centred categories," *International Journal of Integrated Care*, vol. 17, no. 3, p. 16, 2017.



- [3] L. Hardin, A. Kilian, and E. Murphy, "Bundled payments for care improvement," *The Journal of Nursing Administration*, vol. 47, no. 6, pp. 313–319, 2017.
- [4] L. S. Ronga, S. Jayousi, C. Aragno, M. Bagni, L. Carniato, and F. Fongher, "Satellite integrated heterogeneous architecture for professional health services," *IET Communications*, vol. 12, no. 1, pp. 9–17, 2018.
- [5] T. H. Kim, E. C. Park, S. I. Jang, S. Y. Jang, S. A. Lee, and J. W. Choi, "Effects of diagnosis-related group payment system on appendectomy outcomes," *Journal of Surgical Research*, vol. 206, no. 2, pp. 347–354, 2016.
- [6] T. D. Azad, M. Kalani, T. Wolf et al., "Building an electronic health record integrated quality of life outcomes registry for spine surgery," *Journal of Neurosurgery Spine*, vol. 24, no. 1, pp. 176–185, 2016.
- [7] Autoren (alphabetisch), German Association of Chairmen in Academic Radiology (KLR), and German Roentgen Society (DRG), "White paper: radiological curriculum for undergraduate medical education in Germany," *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, vol. 188, no. 11, pp. 1017–1023, 2016.
- [8] Y. J. Liu, F. Fang, J. Q. Fang, J. R. Zhang, X. L. Chi, and H. D. Chen, "Analgesic effect and mechanism of electroacupuncture on rats with chronic inflammatory pain," *Zhongguo Zhong xi yi jie he za zhi Zhongguo Zhongxiyi jiehe zazhi= Chinese Journal of Integrated Traditional and Western Medicine*, vol. 36, no. 6, pp. 690–695, 2016.
- [9] M. N. Ward, "New technology codes in FY 2019 ICD-10-PCS," *Health Management Technology*, vol. 39, no. 5, pp. 10–10, 2018.
- [10] S. Schleder, L. M. Dendl, C. Niessen, C. Stroszczyński, and A. G. Schreyer, "Acceptance of medical apps and e-books among German radiologists," *Der Radiologe*, vol. 57, no. 2, pp. 1–7, 2017.
- [11] Y. L. Hsieh, C. Z. Hong, S. Y. Liu, L. W. Chou, and C. C. Yang, "Acupuncture at distant myofascial trigger spots enhances endogenous opioids in rabbits: a possible mechanism for managing myofascial pain," *Acupuncture in Medicine*, vol. 34, no. 4, pp. 302–309, 2016.
- [12] A. Kazemi, M. Rahmati, R. Eslami, and V. Sheibani, "Activation of neurotrophins in lumbar dorsal root probably contributes to neuropathic pain after spinal nerve ligation," *Iranian Journal of Basic Medical Sciences*, vol. 20, no. 1, pp. 29–35, 2017.
- [13] J. M. Hensel, V. H. Taylor, K. Fung, and S. N. Vigod, "Rates of mental illness and addiction among high-cost users of medical services in Ontario," *Canadian Journal of Psychiatry*, vol. 61, no. 6, pp. 358–366, 2016.
- [14] C. D. Newgard, Z. Yang, D. Nishijima et al., "Cost-effectiveness of field trauma triage among injured adults served by emergency medical services," *Journal of the American College of Surgeons*, vol. 222, no. 6, pp. 1125–1137, 2016.
- [15] H. M. Colón Jordán, J. E. Laborde Rivera, H. A. Marín Centeno, and C. E. Albizu-García, "Medical costs of persons with drug use disorders among Medicaid managed care beneficiaries in Puerto Rico," *Journal of Behavioral Health Services & Research*, vol. 43, no. 2, article 9469, pp. 293–304, 2016.
- [16] S. Eftekhari, N. Yaraghi, R. Singh, R. D. Gopal, and R. Ramesh, "Do health information exchanges deter repetition of medical services?," *ACM Transactions on Management Information Systems*, vol. 8, no. 1, pp. 1–27, 2017.
- [17] E. Liffick, N. F. Mehdiyou, J. L. Vohs, M. M. Francis, and A. Breier, "Utilization and cost of health care services during the first episode of psychosis," *Psychiatric Services*, vol. 68, no. 2, pp. 131–136, 2017.
- [18] J. V. Peter, K. Thomas, L. Jeyaseelan et al., "Cost of intensive care in India," *International Journal of Technology Assessment in Health Care*, vol. 32, no. 4, pp. 241–245, 2016.
- [19] T. Pauley, J. Matienzo, J. Ventura, and J. Barbita, "Cost of an integrated care program to reduce ED visits during diabetic prisoner court hearings," *Journal of Correctional Health Care*, vol. 23, no. 3, pp. 283–296, 2017.
- [20] Z. Tan, H. Wei, Y. Chen, M. du, and S. Ye, "Design for medical imaging services platform based on cloud computing," *International Journal of Big Data Intelligence*, vol. 3, no. 4, pp. 270–278, 2016.
- [21] B. K. Chopra, "Paramedical personnel of the armed forces medical services: the unsung heroes," *Medical Journal Armed Forces India*, vol. 72, no. 1, pp. 5–7, 2016.
- [22] H. Li and S. Chen, "Detection ability mathematical model and performance evaluation method in visible-light photoelectric detection system," *IEEE Sensors Journal*, vol. 17, no. 6, pp. 1649–1655, 2017.
- [23] G. Chen, J. Stang, and M. Moghaddam, "A conformal FDTD method with accurate waveport excitation and S-parameter extraction," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 10, pp. 4504–4509, 2016.
- [24] N. A. Lutsenko and S. S. Fetsov, "Numerical model of time-dependent gas flows through bed of granular phase change material," *International Journal of Computational Methods*, vol. 17, no. 6, article 1950010, 2020.

## Review Article

# Intelligent Rehabilitation Assistance Tools for Distal Radius Fracture: A Systematic Review Based on Literatures and Mobile Application Stores

Yalan Chen,<sup>1,2</sup> Yijun Yu,<sup>3</sup> Xin Lin,<sup>4</sup> Zhenwei Han,<sup>3</sup> Zhe Feng,<sup>1</sup> Xinyi Hua,<sup>1</sup> Dongliang Chen,<sup>3</sup> Xiaotao Xu,<sup>1</sup> Yuanpeng Zhang,<sup>1</sup> and Guheng Wang<sup>1b</sup>

<sup>1</sup>Department of Medical Informatics, School of Medicine, Nantong University, Nantong 226001, China

<sup>2</sup>Department of Hand Surgery, Affiliated Hospital of Nantong University, Nantong 226001, China

<sup>3</sup>Nantong University Xinglin College, Nantong University, Nantong 226236, China

<sup>4</sup>Bachelor of Nursing, University of Technology Sydney, Sydney, NSW 2007, Australia

Correspondence should be addressed to Guheng Wang; [jiawenjiachen@126.com](mailto:jiawenjiachen@126.com)

Received 14 June 2020; Revised 6 September 2020; Accepted 11 September 2020; Published 29 September 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Yalan Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Objective.** To systematically analyze the existing intelligent rehabilitation mobile applications (APPs) related to distal radius fracture (DRF) and evaluate their features and characteristics, so as to help doctors and patients to make evidence-based choice for appropriate intelligent-assisted rehabilitation. **Methods.** Literatures which in regard to the intelligent rehabilitation tools of DRF were systematic retrieved from the PubMed, the Cochrane library, Wan Fang, and VIP Data. The effective APPs were systematically screened out through the APP markets of iOS and Android mobile platform, and the functional characteristics of different APPs were evaluated and analyzed. **Results.** A total of 8 literatures and 31 APPs were included, which were divided into four categories: intelligent intervention, angle measurement, intelligent monitoring, and auxiliary rehabilitation games. These APPs provide support for the patients' home rehabilitation guidance and training and make up for the high cost and space limitations of traditional rehabilitation methods. The intelligent intervention category has the largest download ratio in the APP market. Angle measurement tools help DRF patients to measure the joint angle autonomously to judge the degree of rehabilitation, which is the most concentrated type of literature research. Some of the APPs and tools have obtained good clinical verification. However, due to the restrictions of cost, geographic authority, and applicable population, a large number of APPs still lack effective evidence to support popularization. **Conclusion.** Patients with DRF could draw support from different kinds of APPs in order to fulfill personal need and promote self-management. Intelligent rehabilitation APPs play a positive role in the rehabilitation of patients, but the acceptance of the utilization for intelligent rehabilitation APPs is relatively low, which might need follow-up research to address the conundrum.

## 1. Introduction

Distal radius fracture (DRF) is one of the common upper limb fractures, which could be accounted for about 15.7% of upper limb fractures [1]. It is more common in elderly women; this may be due to the effects of osteoporosis and eventually lead to falls. Due to the influence of osteoporosis and the increase of age, slight external force could often result

in complex comminuted fracture, and with the possible injury of intercarpal ligament for postmenopausal women.

Majority of the patients with this type of fracture are mainly middle-aged and elderly women, most of whom still have a misunderstanding that they need to rest and cannot exercise during the recovery phase. Exercise intervention has been proven to be one of the most effective ways to improve health and promote recovery [2]. Appropriate

rehabilitation exercise can accelerate local blood circulation; accelerate the repair of surrounding soft tissues; prevent muscle atrophy, joint cavity adhesion, and stiffness; and reduce the occurrence of complications such as dysfunction of joints near the fracture site [3].

Elderly patients with fractures have poor compliance with postoperative rehabilitation training. In addition, in the short-term hospitalization period, it is difficult for patients to master the comprehensive knowledge and skills of rehabilitation training and to guarantee the progress and quality of independent rehabilitation training after discharge. Moreover, some traditional rehabilitation methods require patients to conduct at a fixed time and place, and some of the method is relatively boring. Too much repetitive exercise makes it difficult for patients to persist for a long time and lack motivation.

mHealth is a medical and public health practice supported by mobile devices that fills the above gap and promotes unprecedented opportunities for professional clinical diagnosis and treatment recommendations [4, 5]. As a software application accompanying mHealth, commonly known as “application” (APP), it has become one of the main ways for patients to participate in health and rehabilitation management. It can realize home rehabilitation training for patients, self-diagnose the rehabilitation progress, and obtain professional rehabilitation guidance, which greatly promotes the development of participatory medicine [6, 7]. For example, the Serious Game APP could be regarded as effective mean for intervention in home exercise projects [8], while the protractor software could measure the angle of arm rotation and abduction for patients and evaluate the degree of recovery by collecting each measurement data.

At present, the practical utilization of intelligent rehabilitation assistance tools related to DRF is not very extensive. In many cases, patients find it difficult to choose appropriate tools according to the rehabilitation needs when they are faced with a variety of APPs on different mobile platforms. Therefore, we summarize the up-to-date APPs related to DRF in both literatures and different mobile platforms. Through evidence-based retrieval and data extraction, we systematically analyze the functions, characteristics, and limitations of these APPs, to provide evidence-based references for clinical and patients.

## 2. Materials and Methods

**2.1. Data Retrieval.** A systematic review of the literature was performed by searching PubMed, the Cochrane Library, Wan Fang Data, and VIP Data for relevant literatures on February 20, 2020. The selected research is mainly about smart or intelligent APPs related to the rehabilitation of DRF patients. Based on the relative review studies and discussions with hand surgeons, as well as adjustments in the actual retrieval process, the following retrieval strategies were finally formulated: (distal radius fracture [tiab] OR DRF [tiab] OR fractur\* [tiab]) AND (smartphone [tiab] OR application [tiab] OR app\* [tiab] OR iphone [tiab]) NOT (Amyloid precursor protein [tiab]).

The intelligent rehabilitation tools on the APP market are mainly divided into several categories such as health education and functional exercise. “distal radius fracture”, “rehabilitation”, “functional exercise” or “health education” were searched in Chinese and English separately in APP store, Google player, and Huawei platform.

**2.2. Inclusion and Exclusion Criteria.** Inclusion criteria include the following: (1) the study about the intelligent rehabilitation APPs or tools for the DRF patient, especially intervention research; (2) APPs or tools that assist the intelligent rehabilitation for the DRF patient on any platform; (3) rehabilitation is the main function or included function of the APPs or tools; and (4) the APP was applicable from intelligent mobile terminals.

Exclusion criteria include the following: (1) APPs or tools limited to the operation and use of medical staff, (2) APPs with no score or score less than 3 and download volume less than 100, (3) APPs or tools for rehabilitation of other fracture types, and (4) APPs or tools without detailed description and usage records.

**2.3. Study Selection and Data Extraction.** Data from the included studies were extracted by the two investigators (J.Y. and D.C.) using standardized and piloted design formats. The preset extraction indicators mainly include basic information such as the name of the APP, the platform used, and the function. The literature research also needs to extract the clinical research methods and results, and the platform information needs to include user usage, downloads, and ratings. Discrepancies in the process of study selection and data extraction were resolved through a group discussion with two other authors (Y.C. and G.W.).

The extraction of the intelligent rehabilitation APPs involved in the literature, classified them according to the functions and types of different APPs, and analyzed the advantages and disadvantages.

## 3. Results

**3.1. Eligible Studies and Characteristics.** The flow of the search strategy is shown in Figure 1. A total of 2960 articles were searched in the four databases. Subsequently, 883 non-human research articles were screened out through sifting and screening. The remaining 2002 articles were analyzed by abstract and title, then 146 for full-text analysis, and 8 articles were finally included (Table 1). Ninety-seven intelligent rehabilitation APPs were sifted through according to the description by the publisher and the rating. Finally, 31 APPs related to the rehabilitation of DRF were included (Supplementary Table 1).

**3.2. Brief Analysis of Intelligent Rehabilitation Tools.** It can be seen from the result that 25 (68%) smart APPs are applicable to both iOS and Android, with 7 (19%) smart APPs are suitable for iOS system, and 5 (13%) smart APPs are suitable for Android system (Figure 2(a)). Most of the APPs are free.

**3.3. Analysis of Downloads and User Evaluations of Different APPs.** The number of APP downloads from the APP market

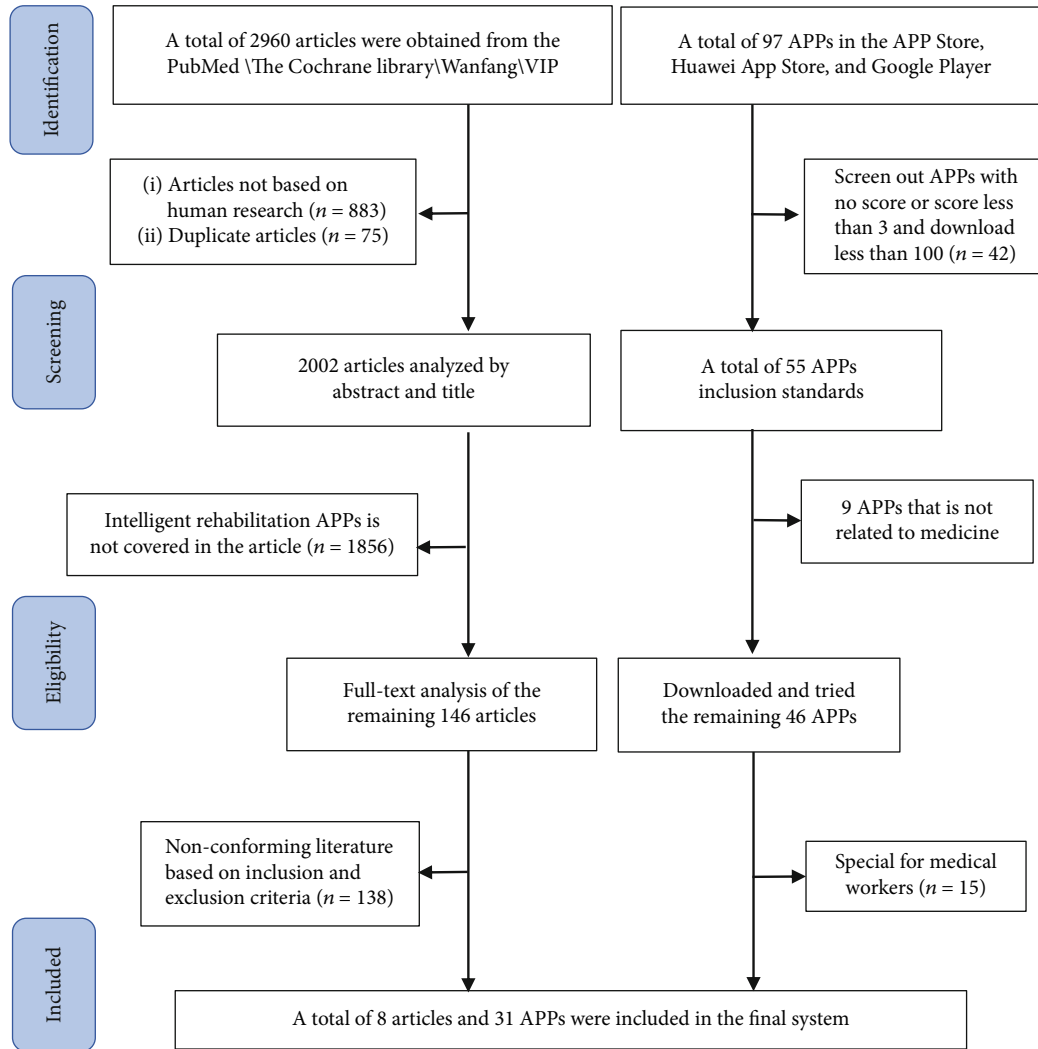


FIGURE 1: The selection process of literatures and APP stores.

is concisely described in Figure 2(b). The top three APPs with more than 100,000 downloads are the intelligent intervention APP “Fisioterapia a tu alcance” and two intelligent monitoring APPs “OASIS Healthcare (HongHua Medical)” and “Gold nurse (gold nurse).” The three software not only have the highest downloads but also got a user score higher than 4 points, which was highly praised by users.

**3.4. Classification and Function Characteristics of Different APPs.** Intelligent rehabilitation APPs were divided into five categories according to different functions: intelligent intervention, angle measurement, intelligent monitoring, and rehabilitation game [17] (Figure 2(c)).

Intelligent intervention software was accounted for the most of the results retrieved in the APP market. Intelligent intervention is mainly used in the later stage of fracture, which is a critical period for functional exercise after reaching the clinical healing standard. Such APPs would be able to provide patients with functional exercise and develop personalized rehabilitation services by combining physiotherapy (PT) and home exercise programs (Heps).

The second proportion of the results retrieved in the APP market is the angle measurement APP which holds the largest proportion among selected literatures regarding intelligent rehabilitation APPs. The degree of rehabilitation of DRF patients is generally diagnosed by the angle of extension, flexion, pronation, and supination of the wrist and forearm. The angle measurement tool allows the patient to measure and record the rehabilitation angle without leaving home and can make independent judgments or share the data with the doctor in real time so that the doctor can understand the degree of the patient’s rehabilitation. The ulnar deviation measurement of the two APPs is demonstrated in Figure 3.

There are not many rehabilitation games related to fractures in the APP store and literature studies. The way of games is conducive to divert the patient’s attention, in the process of playing games to enhance wrist finger movement and joint training and reduce the discomfort in the training process.

The intelligent monitoring APP involves more complicated processes and links, providing patients with an

TABLE 1: Brief information of the included studies.

Authors (year)	APP	System	Study type	Sample size	Method	Effect
Mohammad Reza Pour Ahmadi (2016)	G-Pro [9]	iOS Android	Cross-sectional observational study	38 men 32 women	The wrist flexion, extension, radial deviation, and ulnar deviation joint activity were measured by the universal goniometer and g-pro APP in iPhone 5, respectively.	The G-pro® APP has good to excellent reliability ( $ICC \geq 0.73$ ), and it is effective at the same time as the general goniometer ( $r \geq 0.80$ ) used to measure wrist ROM.
Jacob Modest BS (2018)	Compass [10]	iOS	Cross-sectional study	30 healthy 30 injured	The compass APP included in the conventional goniometer iPhone 5 was used to measure the joint activity during wrist flexion, extension, pronation, and supination, respectively.	Patients with wrist injuries can use this technology to perform self-measurements, and the measurements are comparable to standard UG ROM assessments performed by health providers.
Nuphar Lendner (2017)	Gyroscope [11]	iOS	Cross-sectional study	153	The flexion, extension, radial deviation, and ulnar deviation of the wrist were measured by common goniometer and free Gyroscope APP in iPhone 4, respectively.	It can improve compliance and is reliable and easy to use. Before using such an APP, it must be evaluated.
Robert H. Wellmon (2016)	Goniometer Records [12]	iOS Android	Descriptive validation study	NA	The research passed Goniometer Records of three smartphones; Goniometer Pro has two APPs and common goniometer. The inclinometer measures the joint activity of each standardized angle, respectively.	The app determines the error inherent in the measurement, which has nothing to do with patient factors and is attributed to the smartphone. The data shows that the installed APP can replace the conventional UG or inclinometer.
Susan Reid (2018)	DrGoniometer [13]	iOS	Clinical measurement study	30 fractured 30 healthy	Patients with DRF, and the forearm supination angle was measured by the universal goniometer and DrGoniometer, respectively. Healthy subjects were also measured by these two methods.	Whether it is a fractured forearm or a healthy forearm, the DrGoniometer and the universal goniometer are highly reliable ( $ICC$ range 0.74-0.88). DrGoniometer is an effective alternative tool.
Henriëtte A.W. Meijer (2019)	ReValidate! [14]	iOS	Cross-sectional study	45; 43	Patients recovering from wrist injury and professional wrist injury therapists completed a full-level game, respectively, and the experimenter finally scored the software, respectively.	ReValidate! is a beneficial and interesting experience for the rehabilitation of wrist fractures. It is tailored to patients and provides functional measurements for patients, which may become a very useful exercise tool in future rehabilitation.
Lori Algar OTD (2014)	Tilt Maze \ Labyrinth [15]	iOS Android	NA	NA	NA	Skilled hand therapy can help recommend appropriate postures and encourage participation in therapeutic games, address specific client deficiencies, and reduce functional problems.
Xiaoxia Zhu (2019)	Rehabilitation assistant [16]	iOS Android	Randomized controlled trial	250; 250	The control group was treated with preintervention rehabilitation assistant APP, while the observation group was treated with postintervention rehabilitation assistant APP, and the patient satisfaction was compared	The health education effect and health education satisfaction of the rehabilitation assistant APP were higher than those before the intervention ( $P < 0.01$ or $P < 0.05$ ).



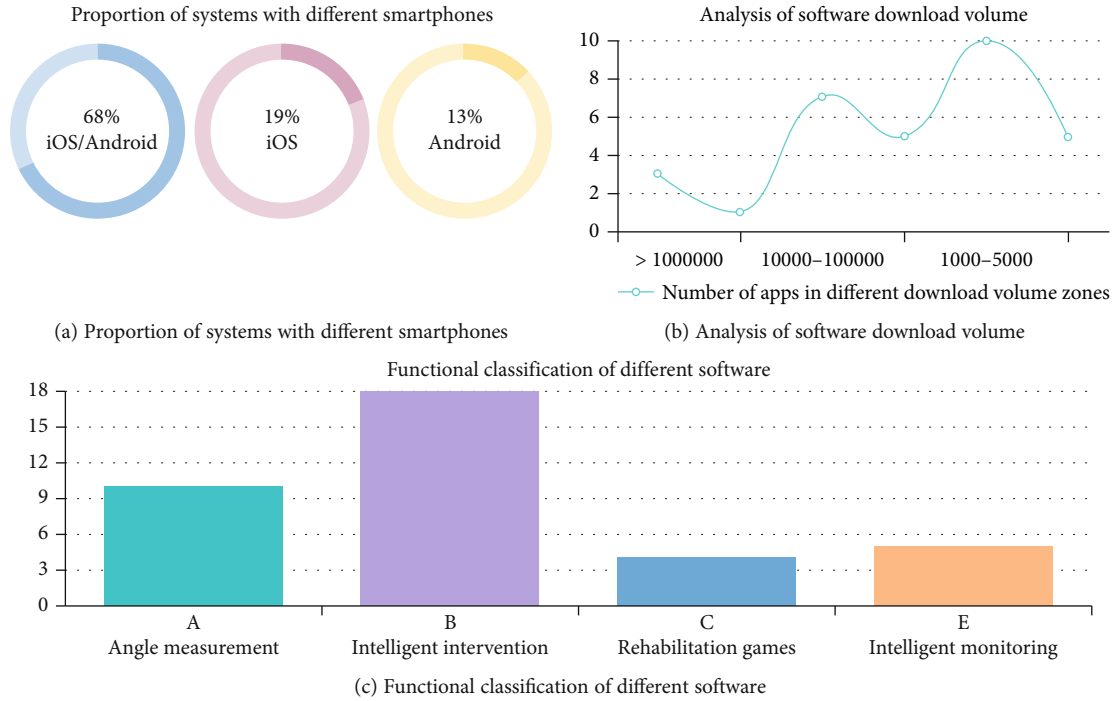


FIGURE 2: Brief analysis of intelligent rehabilitation tools.

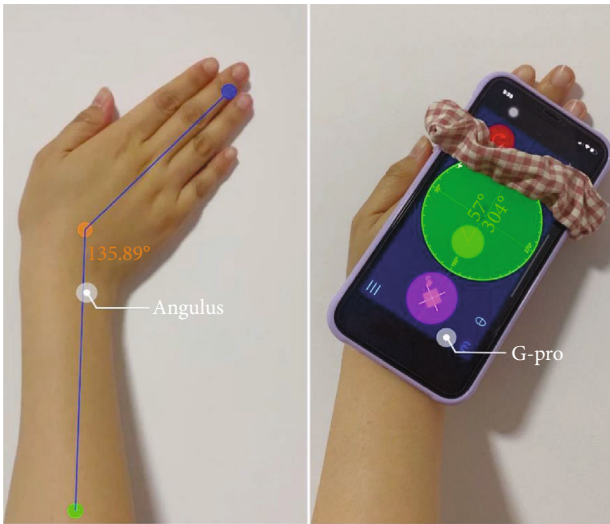


FIGURE 3: The ulnar deviation measurement demo of the Anglus and G-Prp.

emerging medical service model (Figure 4). It can combine online consultation with offline care to provide remote services for patients, ease the gathering and intersection of outpatient populations, promote closer communication between doctors and patients, and improve rehabilitation effects.

#### 4. Discussion

The popularization of mHealth technology has gradually brought a new experience and service mode to the public, which would help to address the disadvantages of the existing

medical system [18, 19]. This research combines scientific research with the practical APP market, comprehensively analyzes the current status of the utilization of DRF's intelligent rehabilitation APP, and discusses the characteristics of different types of rehabilitation APPs.

From the research results of the APP market, the most involved are intelligent intervention APPs, the average downloads are over 1000. The most popular APP is the intelligent monitoring APP, which scores more than 4 points. In contrast, the most involved in literatures are the angle measurement APPs, followed by rehabilitation game APPs. This may indicate that angular measurement and rehabilitation game APPs have more research value; meanwhile, intelligent intervention and intelligent monitoring are more valuable regarding commercial and economic areas. The APP involved in some literatures does not yet exist in the APP market, which may indicate that some APPs are still in the development stage. However, most of the APPs in the APP market lack effective evidence and real data support, and the actual medical value needs further verification.

*4.1. The Advantages and Disadvantages of Different Types of Rehabilitation APPs.* Studies have shown that the prognosis of patient with DRF would benefit from the combination of supervised PT and Heps [20]. Intelligent intervention APPs realize the combination of the two methods. On one hand, the real-time transmission of smart phone devices realizes the connection between patients and medical service providers, which is conducive to the development of personalized guidance programs [21]. Meanwhile, these APPs also provide professional rehabilitation exercise guidance which promotes correct functional exercise [22]. On the other hand, intelligence software runs in the background to supervise

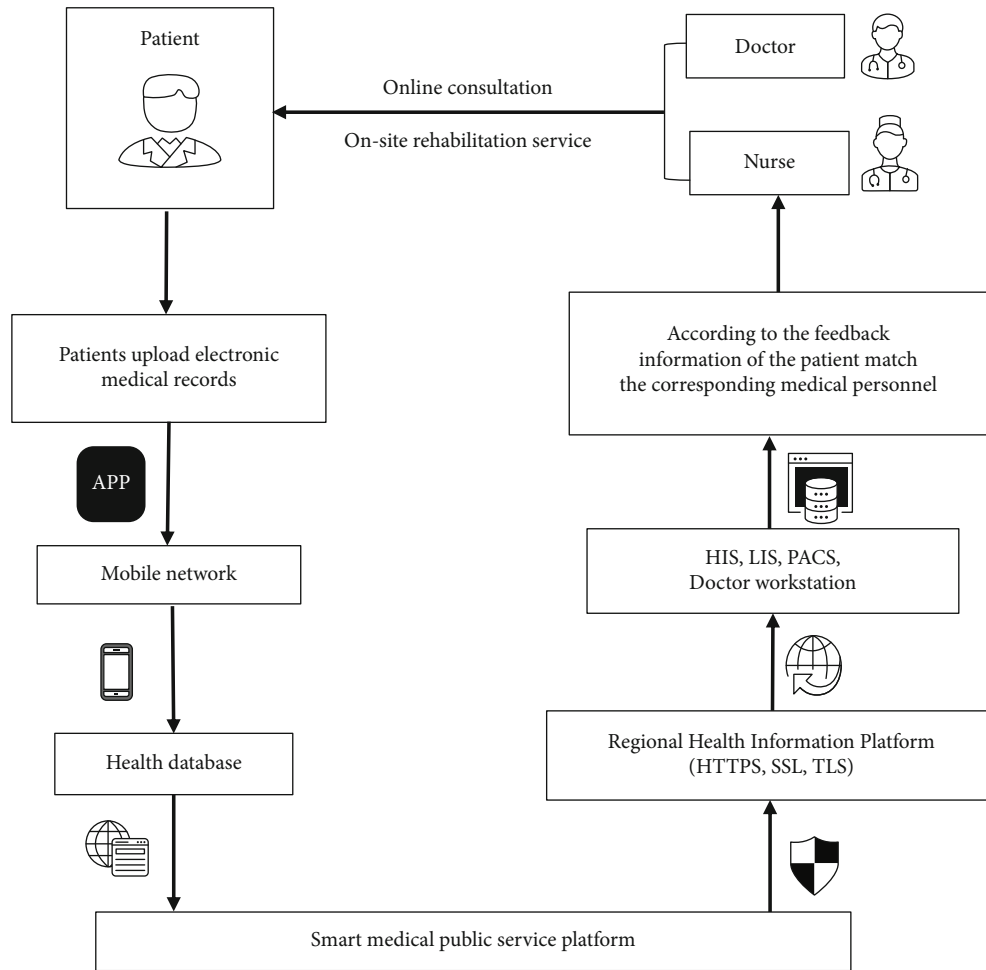


FIGURE 4: Briefly introduced the intelligent monitoring APPs. HIS: hospital information system; LIS: laboratory information management system; PACS: picture archiving and communication systems; HTTPS: Hypertext Transfer Protocol over Secure Socket Layer; SSL: Secure Sockets Layer; TLS: Transport Layer Security.

patients to complete corresponding functional exercises and provide users with timely feedback through charts and other forms of information to improve patients' emotional and psychological conditions in time [23].

Most of the patients with DRF would self-diagnose their condition and recovery process by judging proprioceptive recovery such as grip strength, pinch force, and wrist movement. Few patients would determine their rehabilitation process by measuring the angle of wrist movement [24, 25]. Nevertheless, the current understanding and knowledge in regard to wrist proprioception is still not enough [26–28]. Distal fractures of the radius might often involve with the wrist [29, 30], which would limit the range of movement (ROM) between the forearm and wrist in multiple motor planes, including wrist flexion and extension, carpal ulnar deflection, and forearm pronation [31, 32]. Hence, the ROM would be more accurate when determining the recovery process of DRF [11]. The angle measurement involved in this article mainly measures the range of motion and angle in two ways. For example, the measurement of ulnar deviation, one is to upload the angle photo or video of the part

to be measured to the APP and then measure based on the image data; the other is to fix the mobile device on the part to be measured on the wrist or forearm and measure the movement amplitude (angle) in real time (Figure 3). The accuracy and precision of these two methods are affected by the accuracy of the method used by the patients.

Patients with DRF usually experience symptoms such as limited wrist movement [33], pain, weakness, and even serious complications (nonunion [34] and malunion [35]). Professional rehabilitation care is beneficial for pain management and reduction of the occurrence of adverse complications [36]. Intelligent monitoring APP helps professional medical staff and patients to establish harmonious relationship [10], provide timely and professional medical services, and reduce the occurrence of adverse complications. Moreover, it can provide remote consultation and home care. Compared with the traditional medical model, it can greatly save patients' waiting time and improve patient satisfaction and rehabilitation flexibility.

The advent of the Internet era would allow non-text-based interactive information, such as video and images, to

become more acceptable [37]. Rehabilitation games are interactive APPs that are based on the internet and visual stimulation which could also combine with wearable sensors. For example, “ReValidate!” [14] is an APP that needs to be used in conjunction with wearable devices on the proximal end of the wrist and forearm. By monitoring the parameters of the patient’s wrist or upper limb motion range [32], it can safely and effectively help the DRF patient’s wrist rehabilitation [8]. However, the use of these APPs is affected by the patient’s cognition and acceptance and requires the guidance of professionals. At the same time, due to cost and geographic restrictions, the promotion of some APPs will be restricted by authority.

**4.2. Opportunity and Challenge.** mHealth care can provide personalized precision treatment and better people-centered care [38, 39]. This home-based rehabilitation model can better meet the needs of patients’ autonomous rehabilitation management. At the same time, it can greatly improve the flexibility of rehabilitation management, reduce the medical burden, and better respond to special events such as COVID-19 [40].

The market prospect of developing intelligent auxiliary rehabilitation tools from the perspective of patients is very broad. This study provides reference for patients with different needs to choose appropriate intelligent auxiliary rehabilitation tools: the services provided by intelligent intervention and monitoring APPs are more humanized with easier instruction to operate, which is suitable for middle-aged and elderly patients who cannot master APP skillfully; the rehabilitation gaming APPs are more interesting, but the operation and use process is relatively complicated, which are more suitable for younger patients; and the angle measurement APPs require relevant rehabilitation theory knowledge and are more suitable for patients to utilize under the guidance of medical staff or with the assistance of family members. Meanwhile, through this research, detailed information in regard to APP instruction and operation modes of different types of rehabilitation software was obtained, which provides a basis for subsequent clinical promotion and utilization.

With the continuous deepening of intelligent medical reform, artificial intelligence algorithms have been widely used in auxiliary diagnostic methods of clinical medicine [41, 42]. APP has gradually become a new trend in medical diagnosis, treatment, prevention, and management. However, the clinical application of intelligent APP is still relatively immature, and the practicability and reliability need further verification. For example, in the collection of medical images, noise, image ambiguity, and complex clustering of multiview data [43–46]; permissions and pertinence and limitations of the crowd [47]; and cumbersome operation steps may lead to misoperation and privacy protection of information and data [48], etc. This study is conducive for medical staff to know the utilization of DRF-related APPs, understand high-quality intelligent rehabilitation APPs, and encourage the practical utilization inside and outside the hospital, so as to promote the expected effect of intelligent-assisted rehabilitation tools in clinical and patient home rehabilitation management.

## 5. Conclusions

With the development of the mHealth medical model, intelligent rehabilitation APPs and tools are gradually being used in clinical and patient independent health management. Through systematic evidence-based analysis based on literature and APP platforms, this study integrates different types of intelligent rehabilitation APPs that are suitable for DRF rehabilitation and explores its effects on the rehabilitation of such fractures from different perspectives. Although the evidence is limited, it can still be clearly shown that APP-based rehabilitation intervention, angle measurement, and monitoring management can all improve the effect of rehabilitation training and actively promote patients’ self-rehabilitation management. This also enables clinical medical staff and DRF patients to make evidence-based choice according to the different characteristics and needs of APPs, which could meet the individual needs and improve the effectiveness of self-rehabilitation management. However, due to the restrictions of cost, geographic authority, and applicable population, a large number of APPs still lack effective evidence to support popularization. Therefore, in future research, large-scale user-centered clinical trials that will be added to evaluate the effectiveness and practicality of APPs are particularly important.

## Data Availability

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

## Conflicts of Interest

The authors declare no competing financial interests.

## Authors’ Contributions

The authors’ responsibilities were as follows: G.W., Y.C., and Y.Z. designed the research; Y.C. and Y.Y. designed the research; Y.C., Y.Y., and G.W. designed the pilot data extraction table; Y.Y., Z.H., Z.F., and X.H. performed literature search and study selection; Y.Y., D.C., and X.X. conducted data extraction; Y.Y., Y.C., X.L., and G.W. drafted the manuscript. All the authors completely consented with all the data in the study, critically revised the manuscript, and approved the final version.

## Acknowledgments

This work was supported by the Science and Technology Projects of Nantong City (grant number MSZ19166); the Natural Science Research Project of Xinglin College, Nantong University (grant number 2018K133); the College Student Innovation and Entrepreneurship Training Program of Xinglin College, Nantong University (grant number 202013993031xl); and the College Student Innovation Training Program Project of Nantong University (grant number 2020158).

## Supplementary Materials

The detailed information of 31 APPs related to the rehabilitation of distal radius fracture was contained in Supplementary Table 1, including the operation platform, user rating, download volume, payment method, and specific features description. (*Supplementary Materials*)

## References

- [1] G. H. Wang, R. G. Xie, and J. B. Tang, "Clinical data analysis of incidence and occurrence characteristics of upper limb fracture," *Chinese Journal of Hand Surgery*, vol. 28, no. 2, pp. 95–98, 2012.
- [2] D. Thomas and D. Zanin, "Hand rehabilitation after distal radius fracture," *Hand Surgery and Rehabilitation*, vol. 35s, pp. S156–S161, 2016.
- [3] C. Karagiannopoulos and S. Michlovitz, "Rehabilitation strategies for wrist sensorimotor control impairment: from theory to practice," *Journal of Hand Therapy*, vol. 29, no. 2, pp. 154–165, 2016.
- [4] C. K. Chow, N. Ariyaratna, S. M. Islam, A. Thiagalingam, and J. Redfern, "mHealth in cardiovascular health care," *Heart, Lung & Circulation*, vol. 25, no. 8, pp. 802–807, 2016.
- [5] J. Early, C. Gonzalez, V. Gordon-Dseagu, and L. Robles-Calderon, "Use of mobile health (mHealth) technologies and interventions among community health workers globally: a scoping review," *Health Promotion Practice*, vol. 20, no. 6, pp. 805–817, 2019.
- [6] K. Santo and J. Redfern, "The potential of mHealth applications in improving resistant hypertension self-assessment, treatment and control," *Current Hypertension Reports*, vol. 21, no. 10, p. 81, 2019.
- [7] D. Smahel, S. Elavsky, and H. Machackova, "Functions of mHealth applications: a user's perspective," *Health Informatics Journal*, vol. 25, no. 3, pp. 1065–1075, 2019.
- [8] H. A. Meijer, M. Graafland, J. C. Goslings, and M. P. Schijven, "Systematic review on the effects of serious games and wearable technology used in rehabilitation of patients with traumatic bone and soft tissue injuries," *Archives of Physical Medicine and Rehabilitation*, vol. 99, no. 9, pp. 1890–1899, 2018.
- [9] M. R. Pourahmadi, I. Ebrahimi Takamjani, J. Sarrafzadeh et al., "Reliability and concurrent validity of a new iPhone®) goniometric application for measuring active wrist range of motion: a cross-sectional study in asymptomatic subjects," *Journal of Anatomy*, vol. 230, no. 3, pp. 484–495, 2017.
- [10] J. Modest, B. Clair, R. DeMasi et al., "Self-measured wrist range of motion by wrist-injured and wrist-healthy study participants using a built-in iPhone feature as compared with a universal goniometer," *Journal of Hand Therapy*, vol. 32, no. 4, pp. 507–514, 2019.
- [11] N. Lendner, E. Wells, I. Lavi, Y. Y. Kwok, P. C. Ho, and R. Wollstein, "Utility of the iPhone 4 gyroscope application in the measurement of wrist motion," *Hand*, vol. 14, no. 3, pp. 352–356, 2019.
- [12] R. H. Wellmon, D. T. Gulick, M. L. Paterson, and C. N. Gulick, "Validity and reliability of 2 goniometric mobile apps: device, application, and examiner factors," *Journal of Sport Rehabilitation*, vol. 25, no. 4, pp. 371–379, 2016.
- [13] S. Reid and B. Egan, "The validity and reliability of DrGoniometer, a smartphone application, for measuring forearm supination," *Journal of Hand Therapy*, vol. 32, no. 1, pp. 110–117, 2019.
- [14] H. A. W. Meijer, M. Graafland, M. C. Obdeijn, J. C. Goslings, and M. P. Schijven, "Face validity and content validity of a game for distal radius fracture rehabilitation," *Journal of Wrist Surgery*, vol. 8, no. 5, pp. 388–394, 2019.
- [15] L. Algar and K. Valdes, "Using smartphone applications as hand therapy interventions," *Journal of Hand Therapy*, vol. 27, no. 3, pp. 254–257, 2014.
- [16] X. X. Zhu, J. J. Zan, Y. L. Luan, and L. Q. Fu, "Nursing management and effect evaluation of orthopedic rehabilitation assistant APP," *Today Nurse*, vol. 26, no. 3, pp. 178–180, 2019.
- [17] Z. Feng, Y. Yu, S. Wu et al., "Systematic evaluation and research on intelligent assistant rehabilitation tools after fracture surgery," *China Digital Medicine*, vol. 14, no. 11, pp. 38–41, 2019.
- [18] F. Lucivero and K. R. Jongsma, "A mobile revolution for healthcare? Setting the agenda for bioethics," *Journal of Medical Ethics*, vol. 44, no. 10, pp. 685–689, 2018.
- [19] G. Flodgren, A. Rachas, A. J. Farmer, M. Inzitari, and S. Shepperd, "Interactive telemedicine: effects on professional practice and health care outcomes," *Cochrane Database of Systematic Reviews*, vol. 2015, article Cd002098, no. 9, 2015.
- [20] H. Gutiérrez-Espinoza, D. Rubio-Oyarzún, C. Olguín-Huerta, R. Gutiérrez-Monclus, S. Pinto-Concha, and G. Gana-Hervias, "Supervised physical therapy vs home exercise program for patients with distal radius fracture: a single-blind randomized clinical study," *Journal of Hand Therapy*, vol. 30, no. 3, pp. 242–252, 2017.
- [21] W. B. Hansen and L. M. Scheier, "Specialized smartphone intervention apps: review of 2014 to 2018 NIH funded grants," *JMIR Mhealth Uhealth*, vol. 7, no. 7, article e14655, 2019.
- [22] M. Patel, "Action observation in the modification of postural sway and gait: theory and use in rehabilitation," *Gait & Posture*, vol. 58, pp. 115–120, 2017.
- [23] S. N. Etkind, B. A. Daveson, W. Kwok et al., "Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: does it make a difference? A systematic review," *Journal of Pain and Symptom Management*, vol. 49, no. 3, pp. 611–624, 2015.
- [24] D. Cella, W. Riley, A. Stone et al., "The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008," *Journal of Clinical Epidemiology*, vol. 63, no. 11, pp. 1179–1194, 2010.
- [25] M. J. Shauver, K. W.-C. Chang, and K. C. Chung, "Contribution of functional parameters to patient-rated outcomes after surgical treatment of distal radius fractures," *The Journal of Hand Surgery*, vol. 39, no. 3, pp. 436–442, 2014.
- [26] R. Wollstein, H. Harel, I. Lavi, R. Allon, and D. Michael, "Postoperative treatment of distal radius fractures using sensorimotor rehabilitation," *Journal of Wrist Surgery*, vol. 8, no. 1, pp. 2–9, 2019.
- [27] J. M. Aimonetti, D. Morin, A. Schmied, J. P. Vedel, and S. Pagni, "Proprioceptive control of wrist extensor motor units in humans: dependence on handedness," *Somatosensory & Motor Research*, vol. 16, no. 1, pp. 11–29, 2009.
- [28] E. Hagert, "Proprioception of the wrist joint: a review of current concepts and possible implications on the rehabilitation of the wrist," *Journal of Hand Therapy*, vol. 23, no. 1, pp. 2–17, 2010.



- [29] P. Wiemer, G. Köster, J. Felderhoff, and U. Weber, "Fractures of the distal radius. Changing therapeutic strategies," *Orthopaed*, vol. 28, no. 10, pp. 846–852, 1999.
- [30] P. C. Wong, "Vertebral column and os calcis fracture patterns in a confined community (Singapore). Notes on fracture incidences as clinical evidence of disturbed bone metabolism with increasing age," *Acta Orthopaedica Scandinavica*, vol. 37, no. 4, pp. 357–366, 1966.
- [31] Z. Yang, P. P. H. Lim, S. H. Teo, H. Chen, H. Qiu, and Y. H. Pua, "Association of wrist and forearm range of motion measures with self-reported functional scores amongst patients with distal radius fractures: a longitudinal study," *BMC Musculoskeletal Disorders*, vol. 19, no. 1, p. 142, 2018.
- [32] G. H. Wang, R. G. Xie, Y. L. Chen, and P. C. Xun, "Analysis of the factors affecting the angle and function of wrist joint after distal radius fracture," *Chinese Journal of Hand Surgery*, vol. 35, no. 4, pp. 288–292, 2019.
- [33] S. Batra and A. Gupta, "The effect of fracture-related factors on the functional outcome at 1 year in distal radius fractures," *Injury*, vol. 33, no. 6, pp. 499–502, 2002.
- [34] D. S. Lee and D. R. Weikert, "Complications of distal radius fixation," *The Orthopedic Clinics of North America*, vol. 47, no. 2, pp. 415–424, 2016.
- [35] N. H. Jenkins and W. J. Mintowt-Czyz, "Mal-union and dysfunction in Colles' fracture," *Journal of Hand Surgery*, vol. 13, no. 3, pp. 291–293, 1988.
- [36] S. C. Mears and S. L. Kates, "A guide to improving the care of patients with fragility fractures, edition 2," *Geriatric Orthopaedic Surgery & Rehabilitation*, vol. 6, no. 2, pp. 58–120, 2015.
- [37] S. A. Adams, "Use of 'serious health games' in health care: a review," *Studies in Health Technology and Informatics*, vol. 157, pp. 160–166, 2010.
- [38] G. E. MacKinnon and E. L. Brittain, "Mobile health technologies in cardiopulmonary disease," *Chest*, vol. 157, no. 3, pp. 654–664, 2020.
- [39] S. Agnihotri, L. Cui, M. Delasay, and B. Rajan, "The value of mHealth for managing chronic conditions," *Health Care Management Science*, vol. 23, no. 2, pp. 185–202, 2020.
- [40] R. Collado-Borrell, V. Escudero-Vilaplana, C. Villanueva-Bueno, A. Herranz-Alonso, and M. Sanjurjo-Saez, "Features and functionalities of smartphone apps related to COVID-19: systematic search in App stores and content analysis," *Journal of Medical Internet Research*, vol. 22, no. 8, article e20334, 2020.
- [41] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, p. 118, 2019.
- [42] Y. Zhang, H. Ishibuchi, and S. Wang, "Deep Takagi-Sugeno-Kang fuzzy classifier with shared linguistic fuzzy rules," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1535–1549, 2018.
- [43] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Transactions on Cybernetics*, vol. 45, no. 4, pp. 688–701, 2015.
- [44] Y. P. Zhang, F. L. Chung, and S. T. Wang, "Fast Exemplar-based clustering by gravity enrichment between data objects," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 8, pp. 2996–3009, 2018.
- [45] Y. Jiang, X. Gu, D. Wu et al., "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2020.
- [46] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [47] M. Tomlinson, M. J. Rotheram-Borus, L. Swartz, and A. C. Tsai, "Scaling up mHealth: where is the evidence?," *PLoS Medicine*, vol. 10, no. 2, article e1001382, 2013.
- [48] T. L. Lewis and J. C. Wyatt, "mHealth and mobile medical Apps: a framework to assess risk and promote safer use," *Journal of Medical Internet Research*, vol. 16, no. 9, article e210, 2014.



## Research Article

# EEG-Based Epilepsy Recognition via Multiple Kernel Learning

Yufeng Yao <sup>1,2</sup>, Yan Ding <sup>2</sup>, Shan Zhong <sup>2</sup>, and Zhiming Cui <sup>3</sup>

<sup>1</sup>The Institute of Intelligent Information Processing and Application, Soochow University, Suzhou 215006, China

<sup>2</sup>Department of Computer Science and Engineering, Changshu Institute of Technology, Changshu 215500, China

<sup>3</sup>Suzhou University of Science and Technology, Suzhou 215009, China

Correspondence should be addressed to Zhiming Cui; [zmcai@usts.edu.cn](mailto:zmcai@usts.edu.cn)

Received 25 June 2020; Accepted 3 August 2020; Published 29 September 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Yufeng Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the field of brain-computer interfaces, it is very common to use EEG signals for disease diagnosis. In this study, a style regularized least squares support vector machine based on multikernel learning is proposed and applied to the recognition of epilepsy abnormal signals. The algorithm uses the style conversion matrix to represent the style information contained in the sample, regularizes it in the objective function, optimizes the objective function through the commonly used alternative optimization method, and simultaneously updates the style conversion matrix and classifier during the iteration process parameter. In order to use the learned style information in the prediction process, two new rules are added to the traditional prediction method, and the style conversion matrix is used to standardize the sample style before classification.

## 1. Introduction

Due to the proposal of support vector machine (SVM) [1] and the development of related theories, the kernel method has become an effective method to deal with nonlinear fractional data. Since the performance of the classification algorithm depends largely on the representation of data, the kernel method uses relatively simple functional operations to map samples to higher dimensions, avoiding the design of feature space and complex inner product calculation in feature space. For example, in [2], a fast kernel ridge regression was proposed by using the kernel method. In the last decades, the kernel method has been applied in many fields of machine learning [3–5].

However, some data sets contain samples with uneven distribution, heterogeneous features, or irregular data; the single-kernel method using only a single feature space performs poorly. And since different kernel functions have their characteristics, even in the same application, the effect of using different kernel functions may be very different, which makes the selection of kernel functions and their parameters have an important influence on the performance of the algorithm. Since one kernel function often cannot meet the requirements in some practical application scenarios, multi-

kernel learning that combines multiple kernel functions has been attracting more attention [6].

The combination generated by multikernel learning can be the combination of the same kernel function under different parameters or the combination of many different kernel functions [7]. After years of research, compared with single kernel function, multikernel learning has stronger flexibility, higher interpretability, and better performance in data dimension reduction [8], text classification [9], domain adaptation [10], and other fields.

Although the multikernel learning algorithm fully combines the mapping ability of different kernel functions for data, essentially, it only uses the physical characteristics of samples that include similarity and distance and fails to take into account the implicit information in the stylized data set in the real situation. In practical application, in addition to the representative content information, the data set often contains a variety of style information, and samples with the same style often exist in the form of groups. For example, there are two ways of dividing the letters shown in Figure 1(a), i.e., by the content shown in Figure 1(b) and by the font shown in Figure 1(c), where each font is regarded as a style, and such data is regarded as stylized data.

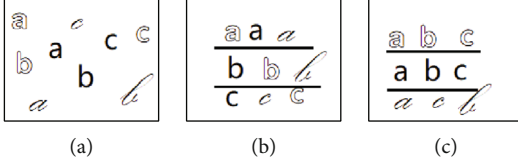


FIGURE 1: Example of stylistic data: (a) data set; (b) different contents; (c) different styles.

To mine the style information of data, scholars have done many types of research. The second-order statistical model proposed in the literature [11] is applied to the problem of number recognition, but it only has a good effect on the data subject to the Gaussian distribution, which leads to a great limitation in the application scenario of the algorithm. The bilinear discriminant model proposed in the literature [12] has achieved good results in behavior recognition data, but the computational cost of the algorithm is relatively high. The domain Bayesian algorithm proposed in the literature [13] improves the naive Bayesian algorithm to identify the style information in the sample group, but it needs to specify a clear data distribution type for the algorithm in advance. However, the distribution of data in real situations is often complex and difficult to be determined in advance. The algorithm proposed in the literature [14, 15] uses a single mapping to mine the style information of samples and achieves excellent results in regression and classification problems, but it makes limited use of the physical characteristics of samples. The time-series style model of mining sample historical information proposed in the literature [16] and the bilayer clustering model of user's age and gender information proposed in the literature [17] effectively make use of the style information in the data in the unsupervised problem, but the algorithm is only targeted at specific fields, and the use of style information is limited.

Inspired by the above scholars, we propose style regularization least squares support vector machine based on multiple kernel learning (SR-MKL-SVM) to excavate and utilize the physical similarities between sample points and the implied style information in samples. In addition to using the physical characteristics of each basic kernel function for data mapping to express the similarity between samples, the algorithm uses the style transformation matrix to represent and mine the style information contained in the data set and takes it into the objective function. In the training process, the alternate optimization strategy is used to update the style transformation matrix in addition to the classifier parameters, and the mined style information is used to synchronously update the kernel matrix. To use the sample style information obtained by training in the process of prediction, two new prediction rules are added on top of the prediction method of traditional multikernel least squares support vector machine. Because the style information contained in the sample is used effectively in the training and prediction process, the experiments of most of the stylized data sets show that SR-MKL-SVM is relatively recent and the classical multikernel support vector machine algorithm is effective.

## 2. Related Works

**2.1. Multikernel Learning.** Let  $\mathbf{x}$  and  $\mathbf{z}$  be two sample vectors;  $\Phi$  is a mapping function from the input space to the feature space. If there is a function  $k(\cdot, \cdot)$ , which can be defined as

$$k(\mathbf{x}, \mathbf{z}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{z}) \rangle = \Phi^T(\mathbf{x})\Phi(\mathbf{z}), \quad (1)$$

then we call  $k(\cdot, \cdot)$  the kernel function. Multikernel learning expects to achieve better mapping performance by combining different kernel functions. There are many ways to combine [6] kernel functions. In this study, we use the following way to find a final combined kernel function based on  $M$  basic kernel functions  $k_i(\cdot, \cdot)$ . If we use  $\mu_i$  to represent the kernel function coefficient, then the final combined kernel function is formulated as

$$k(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^M \mu_i k_i(\mathbf{x}, \mathbf{z}), \quad (2)$$

where

$$\sum_{i=1}^M \mu_i, \quad \mu_i > 0, i = 1, 2, \dots, M. \quad (3)$$

According to Mercer's theory, the combined kernel function generated by the above method still meets the Mercer condition.

**2.2. Least Squares Support Vector Machine Based on Multikernel Learning.** Let  $\mathbf{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$  be the training sample set;  $\mathbf{x}_j \in \mathbf{R}^d$  and  $y_j \in \{+1, -1\}$  are the label corresponding to  $\mathbf{x}_j$ . The objective function of the least squares support vector machine (LSSVM) proposed by Suykens [18] can be formulated as

$$\min_{\mathbf{w}, b, e} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\lambda}{2} \sum_{j=1}^n e_j^2 \quad (4)$$

$$\text{s.t. } y_j = \mathbf{w}^T \Phi(\mathbf{x}_j) + b + e_j, \quad j = 1, 2, \dots, n,$$

where  $\Phi(\mathbf{x}_j)$  represents the mapped  $\mathbf{x}_j$  in a high dimension,  $\mathbf{w}$  and  $b$  are the classification hyperplane parameters,  $e_j (j = 1, 2, \dots, n)$  is the error term, and  $\lambda$  is the regularization parameter.

The Lagrange multiplier  $\alpha$  is introduced into Equation (4), and its dual form can be further obtained by the Slater constraint specification:

$$\begin{aligned} \max_{\alpha} & -\frac{1}{2} \alpha^T \mathbf{K} \alpha - \frac{1}{2\lambda} \alpha^T \mathbf{a} + \alpha^T \mathbf{Y} \\ \text{s.t. } & \alpha^T \mathbf{I}_n = 0, \end{aligned} \quad (5)$$

MK-LSSVM.  
 Input: Dataset  $\mathbf{D}$ , threshold  $\sigma$ , parameter  $\lambda$   
 Output:  $\{\boldsymbol{\mu}, \mathbf{w}, b\}$   
 1. Set  $\mu_i = 1/M (i = 1, 2, \dots, M)$   
 2. Set  $iter = 1$   
 3. **Repeat**  
 4.     Use (2) to combine kernel function and (6) to compute  $\{\mathbf{a}, b\}$   
 5.     Fix  $\boldsymbol{\alpha}$ , update  $\boldsymbol{\mu}$  by equation (8)  
 6. **Until**  $(|1 - f_{iter}(\boldsymbol{\mu}, \boldsymbol{\alpha})/f_{iter}(\boldsymbol{\mu}, \boldsymbol{\alpha})| \leq \sigma)$   
 7. Uses  $\{\boldsymbol{\mu}, \boldsymbol{\alpha}\}$  to compute  $\mathbf{w}$

ALGORITHM 1.

where  $\mathbf{K} \in \mathbf{R}^{n \times n}$  is the kernel matrix. With (11), we can obtain the following two equations,

$$\begin{aligned} \boldsymbol{\alpha} &= \tilde{\mathbf{K}}^{-1} (\mathbf{Y} - b \mathbf{1}_n), \\ b &= \mathbf{I}_n^T \tilde{\mathbf{K}}^{-1} \mathbf{Y} (\mathbf{I}_n^T \tilde{\mathbf{K}}^{-1} \mathbf{I}_n)^{-1}, \end{aligned} \quad (6)$$

where  $\tilde{\mathbf{K}} = \mathbf{K} + \mathbf{I}/\lambda$  and  $\mathbf{Y} = (y_1, y_2, \dots, y_n)^T$ .

By integrating  $\mathbf{K}$  into (2) and (3), we can obtain multiple kernel least squares support vector machine (MK-LSSVM) as

$$\begin{aligned} \min_{\boldsymbol{\mu}} \max_{\boldsymbol{\alpha}} & -\frac{1}{2} \sum_{i=1}^M \mu_i \boldsymbol{\alpha}^T \mathbf{K}_i \boldsymbol{\alpha} - \frac{1}{2\lambda} \boldsymbol{\alpha}^T \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \mathbf{Y} \\ \text{s.t.} \quad & \sum_{i=1}^M \mu_i = 1, \quad 0 \leq \mu_i, \boldsymbol{\alpha}^T \mathbf{1} = 0. \end{aligned} \quad (7)$$

Let  $f(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \max_{\boldsymbol{\alpha}} - (1/2) \sum_{i=1}^M \mu_i \boldsymbol{\alpha}^T \mathbf{K}_i \boldsymbol{\alpha} - (1/2\lambda) \boldsymbol{\alpha}^T \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \mathbf{Y}$ , and replace  $f(\boldsymbol{\mu}, \boldsymbol{\alpha})$  with  $t$ ; we have

$$\begin{aligned} \min_{\boldsymbol{\mu}} \quad & t \\ \text{s.t.} \quad & f(\boldsymbol{\mu}, \boldsymbol{\alpha}) \leq t, \\ & \boldsymbol{\mu} \mathbf{1}_n = 1, \quad 0 \leq \mu_i, \boldsymbol{\alpha}^T \mathbf{1} = 0. \end{aligned} \quad (8)$$

It is obvious that (8) is a semi-infinite linear program (SILP) problem, which can be solved by many existing mature optimization toolkits. For an unseen sample  $\mathbf{x}$ , MK-LSSVM predicts it by using the following equation:

$$\begin{aligned} y &= \text{sign} (w^T \Phi(x) + b) = \text{sign} \left( \sum_{j=1}^n \alpha_j \sum_{i=1}^M \mu_i k_i(x_j, x) + b \right) \\ &= \text{sign} \left( \sum_{i=1}^M \mu_i \sum_{j=1}^n \alpha_j k_i(x_j, x) + b \right) = \\ &= \text{sign} \left( \sum_{i=1}^M \sum_{j=1}^n \mu_i w_i^T \Phi_i(x) + b \right). \end{aligned} \quad (9)$$

2.3. *MK-LSSVM Algorithm Process.* The algorithm steps of MK-LSSVM is shown in Algorithm 1.

### 3. SR-MKL-SVM

3.1. *Objective Function.* Let  $\mathbf{D} = \{(\mathbf{x}_1^1, y_1^1), \dots, (\mathbf{x}_1^{t_1}, y_1^{t_1}), \dots, (\mathbf{x}_N^1, y_N^1), \dots, (\mathbf{x}_N^{t_N}, y_N^{t_N})\}$  be a training set, where the set can be divided into  $N$  groups according to the style. The samples in each group have the same style, and the superscript  $t_j$  is the number of samples in group  $j$ .  $\mathbf{x}_j^k \in \mathbf{R}^d (j = 1, 2, \dots, N, k = 1, 2, \dots, t_j)$  is the  $k$ th sample in group  $j$ . Under the above definition, the objective function of **SR-MKL-SVM** can be formulated as

$$\begin{aligned} \min_{w_i, b, \mu_i, A_j} J_{\text{SR-MKL-SVM}} &= \frac{1}{2} \sum_{i=1}^M \mu_i \mathbf{w}_i^T \mathbf{w}_i + \frac{\lambda}{2} \sum_{j=1}^N \sum_{k=1}^{t_j} (e_j^k)^2 \\ &\quad + \lambda \gamma \sum_{j=1}^N \|\mathbf{A}_j^T - \mathbf{I}\|_F^2 \\ \text{s.t.} \quad y_j^k &= \sum_{i=1}^M \mu_i \mathbf{w}_i^T \mathbf{A}_j^T \Phi_i(\mathbf{x}_j^k) + b + e_j^k, \\ \sum_{i=1}^M \mu_i &= 1, \quad \mu_i \geq 0, i = 1, 2, \dots, M, j = 1, 2, \dots, N, k = 1, 2, \dots, t_j, \end{aligned} \quad (10)$$

where  $\mu_i (i = 1, 2, \dots, M)$  is the weight coefficient of the kernel matrix, where  $M$  is the number of predefined kernel matrices,  $\{\mathbf{A}_j \in \mathbf{R}^{d \times d}\}$  is the style conversion matrix of the sample of style  $j$ , and  $\mathbf{I} \in \mathbf{R}^{d \times d}$  is the identity matrix.

The first two subformulas in  $J_{\text{SR-MKL-SVM}}$  are standard MK-LSSVM expressions, and the third subformula is a penalty term using the Frobenius norm, which is used to control the degree of style conversion of the style conversion matrix to the sample, where the parameter  $\gamma \in \mathbf{R} (\gamma > 0)$  is used. Obviously, when  $\gamma$  is larger, the deviation of the sample  $\mathbf{A}_j^T \Phi(\mathbf{x}_j^k)$  is smaller after style conversion from its original style; otherwise, it is larger; especially when  $\gamma \rightarrow +\infty$  is set, there is  $\mathbf{A}_j^T \Phi(\mathbf{x}_j^k) \rightarrow \Phi(\mathbf{x}_j^k)$ .

3.2. *Optimization.* The goal of the algorithm is to minimize the value of  $J_{\text{SR-MKL-SVM}}$ . It is very difficult to directly

optimize the objective function. We can use the alternating optimization method to obtain a sufficiently available local optimal solution. When  $\mathbf{A}_j$  and  $\{\mathbf{w}_i, \mathbf{b}, \mu_i\}$  are given separately, the objective functions are optimization problems about  $\{\mathbf{w}_i, \mathbf{b}, \mu_i\}$  and  $\mathbf{A}_j$ , and the above two processes are repeated until convergence or the maximum number of iterations is exceeded. To be specific,

- (1) When fixing  $\mathbf{A}_j (j = 1, 2, \dots, N)$ , the optimization problem of formula (10) is transformed into

$$\begin{aligned} \min_{\mathbf{w}_i, \mathbf{b}, \mu_i} & \frac{1}{2} \sum_{i=1}^M \mu_i \mathbf{w}_i^T \mathbf{w}_i + \frac{\lambda}{2} \sum_{j=1}^N \sum_{k=1}^{t_j} (e^k)^2 \\ \text{s.t.} & y_j^k = \sum_{i=1}^M \mu_i \mathbf{w}_i^T \mathbf{A}_j^T \Phi_i(\mathbf{x}_j^k) + b + e_j^k, \\ & \sum_{i=1}^M \mu_i = 1, \quad \mu_i \geq 0, i = 1, 2, \dots, M, j = 1, 2, \dots, N, k = 1, 2, \dots, t_j. \end{aligned} \quad (11)$$

The above formula is about the standard MK-LSSVM problem of sample  $\mathbf{A}_j \Phi_i(\mathbf{x}_j^k)$  after style conversion, and  $\{\mu_i, \mathbf{w}_i, \mathbf{b}\}$  can be determined by Algorithm 1 in Section 2.2 of the article. At this time, the sample  $\mathbf{A}_j \Phi_i(\mathbf{x}_j^k)$  mapped to the high dimension cannot be directly calculated, but the synthetic kernel matrix formed by the style-converted sample  $\mathbf{A}_j \Phi_i(\mathbf{x}_j^k)$  can be updated by the kernel method to obtain the style-converted kernel matrix. The specific method of using the kernel method to obtain the style-converted synthetic kernel matrix will be introduced in Section 3.3 of the article

- (2) When  $\{\mu_i, \mathbf{w}_i, \mathbf{b}\}$  is fixed, then the optimization problem of Equation (10) is transformed into

$$\begin{aligned} \min_{\mathbf{A}_j \in \mathbb{R}^{d \times d}} & \frac{\lambda}{2} \sum_{k=1}^{t_j} (e_j^k)^2 + \lambda \gamma \|\mathbf{A}_j^T - \mathbf{I}\|_F^2 \\ \text{s.t.} & y_j^k = \sum_{i=1}^M \mu_i \mathbf{w}_i^T \mathbf{A}_j^T \Phi_i(\mathbf{x}_j^k) + b + e_j^k, \\ & \sum_{i=1}^M \mu_i = 1, \quad \mu_i \geq 0, i = 1, 2, \dots, M, j = 1, 2, \dots, N, k = 1, 2, \dots, t_j. \end{aligned} \quad (12)$$

The above formula is a linear constrained quadratic programming problem for  $\mathbf{A}_j$ , which can be transformed into  $N$  independent problems for each  $\mathbf{A}_j$  to be solved. At this time, the parameters of the synthetic kernel matrix and the classifier have been fixed, similar to the original LSSVM, and the dual form can be obtained after introducing the Lagrange

multiplier to Equation (12):

$$\begin{aligned} \max_{\alpha_j^k} L = & \frac{\lambda}{2} \sum_{k=1}^{t_j} (e_j^k)^2 + \lambda \gamma \|\mathbf{A}_j^T - \mathbf{I}\|_F^2 \\ & - \sum_{k=1}^{t_j} \alpha_j^k \left( \sum_{i=1}^M \mu_i \mathbf{w}_i^T \mathbf{A}_j^T \Phi_i(\mathbf{x}_j^k) + b + e_j^k - y_j^k \right) \\ & - \sum_{i=1}^M \left[ \beta_i \left( 1 - \sum_{i=1}^M \mu_i \right) - \rho_i \mu_i \right]. \end{aligned} \quad (13)$$

Let  $\partial L / \partial \mathbf{A}_j = 0$ ; we have

$$\mathbf{A}_j^T = \frac{1}{2\lambda\gamma} \sum_{k=1}^{t_j} \sum_{i=1}^M \alpha_j^k \mu_i \mathbf{w}_i \Phi_i^T(\mathbf{x}_j^k) + \mathbf{I}. \quad (14)$$

Let  $\partial L / \partial e_j^k = 0$  get  $\alpha_j^k = \lambda e_j^k$ . It can be seen that this formula has the same KKT [18] condition as LSSVM.

Through the process of alternating optimization, it can be known that in the process of training classifier parameters, the samples converted by the style conversion matrix are used as training data. In the first iteration, the style conversion matrix is initialized to the identity matrix. At this time, the samples after the style conversion are the same as the original samples, and no style conversion is generated. Therefore, the classifier parameters obtained by the first round of SR-MKL-SVM training are the same as the original MK-LSSVM. In the subsequent iteration process, due to the optimization of the style conversion matrix, the samples in each style group undergo the transformation of the style conversion matrix and gradually approach the standard style. The classifier parameters trained at this time fully consider the style information contained in the sample as a whole. At the same time, the process of solving the style conversion matrix from Equation (14) not only uses the physical characteristics of the samples obtained by training but also effectively uses the style information in the data. The style conversion matrix trained at this time contains each style group style information. According to the above analysis, the processes of training the classifier parameters and the style conversion matrix make full use of the style information contained in the sample, and the two processes promote each other.

**3.3. Style Transformation.** Since the dimension after the sample is mapped to the high-dimensional space may be infinite, the sample  $\mathbf{A}_j \Phi_i(\mathbf{x}_j^k)$  value after the style transformation cannot be obtained directly. At this point, each element in the synthetic kernel matrix can be updated with the help of the kernel method to obtain the synthetic kernel matrix after the style transformation.

Because the synthesis kernel function still has to satisfy the allowed kernel of the theorem, as  $k(\mathbf{x}_{j1}^{k1}, \mathbf{x}_{j2}^{k2}) = \langle \Phi(\mathbf{x}_{j1}^{k1}), \Phi(\mathbf{x}_{j2}^{k2}) \rangle = \sum_{i=1}^M \mu_i k_i(\mathbf{x}_{j1}^{k1}, \mathbf{x}_{j2}^{k2})$ , let  $\phi$  be for the synthesis of the combined map of the core matrix; by formula (9), you can make the synthesis of the core matrix  $k(\mathbf{x}_{j1}^{k1}, \mathbf{x}_{j2}^{k2}) =$

$\langle \Phi(\mathbf{x}_{j1}^{k1}), \Phi(\mathbf{x}_{j2}^{k2}) \rangle = \sum_{i=1}^M \mu_i \langle \Phi_i(\mathbf{x}_{j1}^{k1}), \Phi_i(\mathbf{x}_{j2}^{k2}) \rangle$ ; let  $\hat{\Phi}(\mathbf{x}_j^k) = \mathbf{A}_j^T \Phi(\mathbf{x}_j^k)$ ; you can get after the style conversion of the core matrix elements as

$$\begin{aligned} \hat{k}(\mathbf{x}_{j1}^{k1}, \mathbf{x}_{j2}^{k2}) &= \Phi^T(\mathbf{x}_{j1}^{k1}) \hat{\Phi}(\mathbf{x}_{j2}^{k2}) \\ &= \langle \hat{\Phi}(\mathbf{x}_{j1}^{k1}), \hat{\Phi}(\mathbf{x}_{j2}^{k2}) \rangle \\ &= \Phi^T(\mathbf{x}_{j1}^{k1}) \mathbf{A}_{j1} \mathbf{A}_{j2}^T \Phi(\mathbf{x}_{j2}^{k2}), \end{aligned} \quad (15)$$

where  $k^\wedge(\mathbf{x}_{j1}^{k1}, \mathbf{x}_{j2}^{k2})$  is the core matrix element after style transformation and  $\mathbf{w} = \sum_{j=1}^N \sum_{k=1}^{t_j} \alpha_j^k \Phi(\mathbf{x}_j^k)$ ; formula (15) can be updated to

$$\begin{aligned} \hat{k}(\mathbf{x}_{j1}^{k1}, \mathbf{x}_{j2}^{k2}) &= \Phi^T(\mathbf{x}_{j1}^{k1}) \mathbf{A}_{j1} \mathbf{A}_{j2}^T \Phi(\mathbf{x}_{j2}^{k2}) = \Phi^T(\mathbf{x}_{j1}^{k1}) \\ &\cdot \left[ \frac{1}{2\lambda\gamma} \sum_{m1=1}^N \sum_{n1=1}^{t_{m1}} \alpha_{m1}^{n1} \Phi(\mathbf{x}_{m1}^{n1}) \Phi^T(\mathbf{x}_{m1}^{n1}) + \mathbf{I} \right] \\ &\cdot \left[ \frac{1}{2\lambda\gamma} \sum_{m2=1}^N \sum_{n2=1}^{t_{m2}} \alpha_{m2}^{n2} \Phi(\mathbf{x}_{m2}^{n2}) \Phi^T(\mathbf{x}_{m2}^{n2}) + \mathbf{I} \right] \\ &\cdot \Phi(\mathbf{x}_{j2}^{k2}) \\ &= \frac{1}{4\lambda^2\gamma^2} \cdot \left[ \sum_{m1=1}^N \sum_{m2=1}^N \sum_{n1=1}^{t_{m1}} \sum_{n2=1}^{t_{m2}} \alpha_{m1}^{n1} \alpha_{m2}^{n2} \alpha_{j1}^{k1} \alpha_{j2}^{k2} \right. \\ &\cdot \langle \Phi(\mathbf{x}_{j1}^{k1}), \Phi(\mathbf{x}_{j1}^{k1}) \rangle \cdot \langle \Phi(\mathbf{x}_{j2}^{k2}), \Phi(\mathbf{x}_{j2}^{k2}) \rangle \\ &\cdot \langle \Phi(\mathbf{x}_{m1}^{n1}), \Phi(\mathbf{x}_{m2}^{n2}) \rangle \left. \right] + \frac{1}{2\lambda\gamma} \sum_{m1=1}^N \sum_{n1=1}^{t_{m1}} \alpha_{j1}^{k1} \alpha_{m1}^{n1} \\ &\cdot \langle \Phi(\mathbf{x}_{j1}^{k1}), \Phi(\mathbf{x}_{m1}^{n1}) \rangle \cdot \langle \Phi(\mathbf{x}_{m1}^{n1}), \Phi(\mathbf{x}_{j2}^{k2}) \rangle \\ &+ \frac{1}{2\lambda\gamma} \sum_{m2=1}^N \sum_{n2=1}^{t_{m2}} \alpha_{j2}^{k2} \alpha_{m2}^{n2} \cdot \langle \Phi(\mathbf{x}_{j1}^{k1}), \Phi(\mathbf{x}_{m2}^{n2}) \rangle \\ &\cdot \langle \Phi(\mathbf{x}_{m2}^{n2}), \Phi(\mathbf{x}_{j2}^{k2}) \rangle + \langle \Phi(\mathbf{x}_{j1}^{k1}), \Phi(\mathbf{x}_{j2}^{k2}) \rangle. \end{aligned} \quad (16)$$

Because of  $k(x_{j1}^{k1}, x_{j2}^{k2}) = \langle \Phi(x_{j1}^{k1}), \Phi(x_{j2}^{k2}) \rangle$  formula (16) can be updated to:

$$\begin{aligned} \hat{k}(x_{j1}^{k1}, x_{j2}^{k2}) &= \frac{1}{4\lambda^2\gamma^2} \cdot \left[ \sum_{m1=1}^N \sum_{m2=1}^N \sum_{n1=1}^{t_{m1}} \sum_{n2=1}^{t_{m2}} \alpha_{m1}^{n1} \alpha_{m2}^{n2} \alpha_{j1}^{k1} \alpha_{j2}^{k2} \right] \\ &\cdot k(x_{j1}^{k1}, x_{j1}^{k1}) \cdot k(x_{j2}^{k2}, x_{j2}^{k2}) \cdot k(x_{m1}^{n1}, x_{m2}^{n2}) \\ &+ \frac{1}{2\lambda\gamma} \sum_{m1=1}^N \sum_{n1=1}^{t_{m1}} \alpha_{j1}^{k1} \alpha_{m1}^{n1} \cdot k(x_{j1}^{k1}, x_{j1}^{k1}) \\ &\cdot k(x_{m1}^{n1}, x_{j2}^{k2}) + \frac{1}{2\lambda\gamma} \sum_{m2=1}^N \sum_{n2=1}^{t_{m2}} \alpha_{j2}^{k2} \alpha_{m2}^{n2} \\ &\cdot k(x_{j1}^{k1}, x_{m2}^{n2}) \cdot k(x_{j2}^{k2}, x_{j2}^{k2}) + k(x_{j1}^{k1}, x_{j2}^{k2}) \end{aligned} \quad (17)$$

SR-MKL-SVM.

Input: Dataset  $\mathbf{D}$ , parameters  $\lambda$  and  $\gamma$

Output: Weight  $\mu_i$ , classifier parameter  $\{\mathbf{w}, b\}$ , style transformation matrix  $\{\mathbf{A}_j\}$

1. Set  $\mu_i = 1/M (i = 1, 2, \dots, M)$

2. Set  $iter = 1, \mathbf{A}_j = \mathbf{I} (j = 1, 2, \dots, N)$

3. **Repeat**

4. Calculate the value of the objective function  $V_{iter}$

5. Update  $\mu_i$  and  $\{\mathbf{w}, b\}$  by (11)

6. Update  $\mathbf{A}_j$  by (14) and the synthetic kernel matrix by (11)

6. **Until** (Iteration count reaches its maximum)

ALGORITHM 2.

3.4. *Algorithm.* The training algorithm of SR-MKL-SVM is listed as follows.

SR-MKL-SVM uses alternate optimization method to solve the problem, which can be divided into two steps. The first step is kernel matrix weight coefficient and classifier parameter optimized steps can be divided into two subproblems, respectively, i.e., solving the kernel weight SILP problems and solving the linear programming problem of the classifier parameters for the synthesis of kernel matrix at the same time, the time complexity  $O(M^2n^2)$  and  $O(n)$ , respectively. Due to  $M \geq 1$ , the total time complexity can be treated as  $O(M^2n^2)$ . The second step is to optimize the style-standardization matrix, and the time complexity of this step is  $O(N^2n^2)$ . Therefore, the total time complexity of the algorithm training process is  $O(iter \cdot (M^2n^3 + N^2n^2))$ , where  $M$  is the number of predefined basic kernel matrices,  $N$  is the total number of samples,  $n$  is the number of styles in the data set, and  $iter$  is the number of iterations of the algorithm.

Compared with typical MKL-SVM, the MK-SRLSSVM algorithm is in the process of training in style transformation matrix to the regularization processing style samples, but the multikernel support vector machine (SVM) algorithm in solving the basic kernel function in the process of the weight coefficient is applied to solve the need to invoke the original SVM algorithm in this paper; using the original LSSVM subspaces, the SVM training process is essential in solving quadratic programming problem and the nature of the training process of LSSVM for solving linear programming problems. Therefore, the computational complexity of SR-MKL-SVM in this step is far less than that of the typical MKL-SVM algorithm. The algorithm presented in this paper optimizes the weight coefficient by solving SILP problems, which is superior to the support vector machine algorithm that optimizes the weight coefficient by solving SDP problems or QCQP problems and is comparable to the multikernel support vector machine algorithm that uses SILP and other problems to solve the weight coefficient. Therefore, SR-MKL-SVM has the same complexity as typical support vector machine algorithms.

3.5. *Prediction Rules.* Two new prediction rules were defined based on MK-LSSVM in order to use the weight, classifier parameter  $\{\mu_i, \mathbf{w}, b\}$  and the style transformation matrix



$A_j(j=1, 2, \dots, N)$ . Since the style of the sample may or may not appear in the training process in the practical application, two new prediction rules Rule 2 and Rule 3 are added into the traditional prediction method to deal with the two cases, respectively.

Let  $\mathbf{X}_0 = \{\mathbf{x}_0^1, \mathbf{x}_0^2, \dots, \mathbf{x}_0^{t_0}\}$  be a subset of the entire testing data set in which each element has the same style, and  $\mathbf{x}_0^k \in \mathbb{R}^d (k=1, 2, \dots, t_0)$  is a sample.

*Rule 1. Traditional prediction method.*

Traditional prediction methods only use weight  $\mu_i$  and classifier parameters  $w$  and  $b$  to predict the sample  $\mathbf{x}_0^k$  in the testing data set and obtain the corresponding label  $y_0^k$ :

$$\begin{aligned} y_0^k &= \text{sign} \left[ \mathbf{w}^T \Phi(\mathbf{x}_0^k) + b \right] \\ &= \text{sign} \left[ \sum_{j=1}^N \sum_{k=1}^{L_j} \alpha_j^k \sum_{i=1}^M \mu_i k_i(\mathbf{x}_j^k, \mathbf{x}_0^k) + b \right]. \end{aligned} \quad (18)$$

*Rule 2. Test sample style is known.*

If the style of the test sample already exists in the training data set, the corresponding style transformation matrix acquired during the training process can be directly used to process the style transformation of the sample, so that the sample is close to the standard style. Then, the predicted label  $y_0^k$  was obtained by using traditional prediction rules for the processed sample  $A_0^T \Phi(\mathbf{x}_0^k)$ .

$$\begin{aligned} y_0^k &= \text{sign} \left[ \mathbf{w}^T A_0^T \Phi(\mathbf{x}_0^k) + b \right] \\ &= \text{sign} \left[ \sum_{j=1}^N \sum_{k=1}^{t_j} \alpha_j^k \sum_{i=1}^M \mu_i \hat{k}_i(\mathbf{x}_j^k, \mathbf{x}_0^k) + b \right]. \end{aligned} \quad (19)$$

$\hat{k}_i(\mathbf{x}_j^k, \mathbf{x}_0^k)$  can be obtained from Section 3.3.

*Rule 3. Test sample style is unknown.*

If the sample group  $\mathbf{X}_0$ 's style does not exist in the training data set, to effectively make use of the style information obtained by training; based on the direct extrapolation idea, we consider the same style of the information contained in the sample group as a new style. The detailed steps are as follows:

*Step 1.* Obtain the temporary label  $Y_{\text{temp}} = \{y_0^1, y_0^2, \dots, y_0^{t_0}\}$  of testing data set  $\mathbf{X}_0$  by using Rule 1.

*Step 2.* Train  $\mathbf{X}_0$  and its temporary label  $Y_{\text{temp}}$  with the training data set to obtain the new weight  $\hat{\mu}_i$ , classifier parameter  $\{\hat{w}, \hat{b}\}$ , and style transformation matrix  $A_0^T$ .

TABLE 1: Description of epileptic EEG data set.

Group	Type
A	Healthy
B	
C	
D	Patient
E	

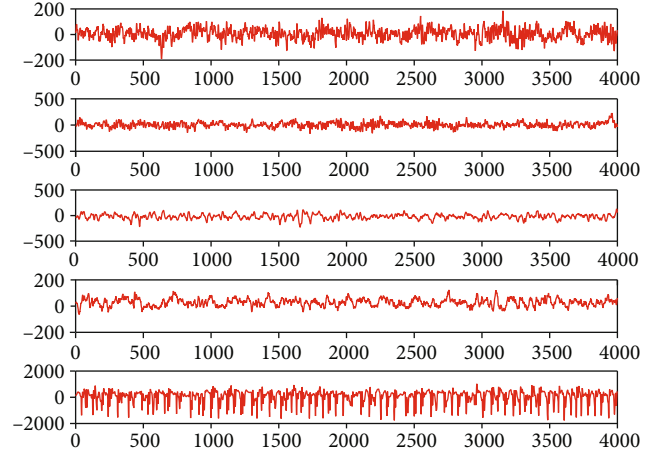


FIGURE 2: EEG data visualization.

TABLE 2: Detail of experimental data sets.

Num	Training data	Testing data
DS.1	Each 50% z (A, B, E)	Other 50% (A, B, E)
DS.2	Each 50% (B, D, E)	Other 50% (B, D, E)
DS.3	Each 50% (A, C, E)	Other 50% (A, C, E)
DS.4	Each 50% (A, C, E)	Other 50% (A, C, E)

*Step 3.* Use  $\{\hat{\mu}_i, \hat{w}, \hat{b}, A_0^T\}$  to predict test set  $\mathbf{X}_0$  and get the formal prediction label  $Y_0 = \{y_0^1, y_0^2, \dots, y_0^{t_0}\}$ .

Since that most of the data in real scenes contain implicit or obvious style characteristics, the new prediction method added in SR-MKL-SVM takes into account the situation of known style and unknown style. The style information corresponding to the predicted samples is directly used to predict the samples with known styles. The direct extrapolation method is used to predict the unknown style samples, and the trained style information is used effectively, so the algorithm has good universality.

**3.6. Analysis of SR-MKL-SVM.** Different from SVM, which only searches for the optimal classification hyperplane according to the physical distribution of the original data, SR-MKL-SVM not only considers the physical characteristics contained in the data but also mines the style characteristics of the data. In this paper, the whole training samples are used to optimize the classifier parameters and the data sets with

TABLE 3: Experimental results of epileptic EEG.

Algorithm	Precision			
	DS.1	DS.2	DS.3	DS.4
simpleMKL	0.9273 ( $C = e^4$ )	0.7920 ( $C = e^1$ )	0.8133 ( $C = e^5$ )	0.8013 ( $C = e^1$ )
easyMKL	0.9520 ( $\lambda_e = 0.1$ )	0.8333 ( $\lambda_e = 0.8$ )	0.5333 ( $\lambda_e = 0.6$ )	0.7767 ( $\lambda_e = 0.3$ )
GMKL	0.9260 ( $C = e^0$ )	0.7867 ( $C = e^5$ )	0.7507 ( $C = e^{-1}$ )	0.7973 ( $C = e^2$ )
LMKL	0.9600 ( $C = e^0$ )	0.8120 ( $C = e^3$ )	0.8200 ( $C = e^3$ )	0.8133 ( $C = e^{-1}$ )
NLMKL	0.9507 ( $C = e^1$ )	0.9493 ( $C = e^5$ )	0.8293 ( $C = e^1$ )	0.8480 ( $C = e^3$ )
RBMKL	0.9413 ( $C = e^0$ )	0.8440 ( $C = e^4$ )	0.7787 ( $C = e^{-1}$ )	0.8067 ( $C = e^{-1}$ )
GLMKL	0.9413 ( $C = e^1$ )	0.8333 ( $C = e^3$ )	0.7627 ( $C = e^0$ )	0.8227 ( $C = e^{-1}$ )
CABMKL	0.9373 ( $C = e^2$ )	0.8453 ( $C = e^2$ )	0.7560 ( $C = e^0$ )	0.8027 ( $C = e^1$ )
NB	0.9520	0.8947	0.8067	0.7087
DT	0.9747	0.8467	0.7533	0.8120
SR-MKL-SVM	0.9503 ( $\lambda = e^{-1}, \gamma = e^{-3}$ )	0.9353 ( $\lambda = e^3, \gamma = e^{-5}a$ )	0.9153 ( $\lambda = e^{-1}, \gamma = e^{-2}$ )	0.9120 ( $\lambda = e^1, \gamma = e^{-1}$ )

different styles are processed, respectively. With the advantage of multikernel learning for data mapping, the algorithm in this paper can represent and process the data containing more complex styles and make full use of the trained style information to conduct style regularization processing on the original samples in both training and testing methods, so that the data distribution after style transformation can be more easily divided. Compared with traditional SVM and SR-MKL-SVM, we find that SR-MKL-SVM can make full use of the information contained in the stylized data to improve the classification performance.

## 4. Experimental Results

**4.1. Data.** In this section, we introduce the EEG data provided by Bonn University to evaluate our proposed method. The EEG data set consists of 5 groups of samples from 2 groups, with detailed information as shown in Table 1, and randomly selected samples from each group as shown in Figure 2. As can be seen from Figure 2, the fluctuation of samples from different groups is very different. For example, the signal fluctuation of patients in group A and healthy people in group E is significantly different. The signal fluctuation of patients in group C and group E also differed greatly under different conditions.

Studies [19] showed that feature extraction of original EEG data in advance could effectively improve classification performance. In this paper, kernel principal component analysis (KPCA) [5, 20] was used to extract features from original data. In this section, the data after dimension reduction is used for experiments. As can be seen above, the number of samples in the data set is 500, the number of categories is 2, and the sample dimension is 70. Samples from the same group are considered to have the same style.

In order to verify the validity of this algorithm, different groups of data are selected to form two types of data sets. The first type of data is all styles contained in the test set exist in the training set at the same time. The second type of data is

the test set has a style not found in the training set, and the details of the construction data set are shown in Table 2.

**4.1.1. Epileptic EEG Data Set.** Data sets DS.1 and DS.2 are the first type of data; DS.3 and DS.4 are the second type of data. All data were random, and 10 experiments were conducted under the same set of parameters, averaging the results. Rule 2 and Rule 3 are used to predict the two types of data. The experimental results and parameters of all algorithms [21–32] are shown in Table 3.

From the experimental results in Table 3, it can be concluded that the decision tree algorithm in data set DS.1 has the best wave signal recognition effect, and the NLMKL algorithm in data set DS.2 has the best classification accuracy, leading all other algorithms including this algorithm. The results of this algorithm in the first two data sets are not as good as DT and NLMKL, but the difference is small.

From the above results, we can see the effectiveness and stability of the proposed algorithm in improving the accuracy of EEG signal recognition by mining and utilizing different fluctuation features contained in each group of samples.

## 5. Conclusion

In order to use the style information contained in the sample, this paper proposes a style regularization least squares support vector machine (SR-MKL-SVM) based on multicore learning. In addition to the advantage of multicore learning for the expression of physical similarity between samples, the algorithm also mines and uses the style information contained in the samples to improve the classification accuracy of the algorithm. SR-MKL-SVM takes the style information contained in the sample into the objective function, uses the style conversion matrix to standardize the sample, uses the regularization method to limit the degree of style conversion, and optimizes both the classifier parameters and the style standard during the training process conversion matrix. In addition to the traditional prediction methods, new

prediction rules that can use the trained style information are added. Experiments in stylized data sets show the effectiveness and certain practicality of the algorithm.

## Data Availability

The original EEG data are available and can be downloaded from <http://www.meb.unibonn.de/epileptologie/science/physik/eeldata.html>.

## 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 under grant 61876217.

## References

- [1] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
- [2] H. Avron, K. L. Clarkson, and D. P. Woodruff, "Faster kernel ridge regression using sketching and preconditioning," *SIAM Journal on Matrix Analysis and Applications*, vol. 38, no. 4, pp. 1116–1138, 2017.
- [3] J. W. Tao and S. T. Wang, "Kernel support vector machine for domain adaptation," *Acta Automatica Sinica*, vol. 38, no. 5, pp. 797–811, 2012.
- [4] C. Feng and S. Z. Liao, "Large-scale kernel methods via random hypothesis spaces," *Journal of Frontiers of Computer Science & Technology*, vol. 12, no. 5, pp. 785–793, 2018.
- [5] B. Schölkopf, S. Mika, A. Smola, G. Rätsch, and K.-R. Müller, "Kernel PCA pattern reconstruction via approximation preimages," in *Proceedings of the 8th international conference on Artificial neural networks*, pp. 147–152, Skövde, Sweden, Piscataway, September 2–4, 1998.
- [6] Y. Zhang, H. Ishibuchi, and S. Wang, "Deep Takagi–Sugeno–Kang fuzzy classifier with shared linguistic fuzzy rules," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1535–1549, 2018.
- [7] R. Alain and C. Stéphane, "More efficiency in multiple kernel learning," in *Proceedings of the 24th international conference on Machine learning*, pp. 775–782, Corvallis, Oregon, New York, 2007.
- [8] Y. Zhang, F. Chung, and S. Wang, "Fast exemplar-based clustering by gravity enrichment between data objects," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 8, pp. 2996–3009, 2020.
- [9] Y. Zhang, F. Chung, and S. Wang, "A multiview and multiexemplar fuzzy clustering approach: theoretical analysis and experimental studies," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 8, pp. 1543–1557, 2019.
- [10] S. S. Bucak, Rong Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1354–1369, 2014.
- [11] S. Veeramachaneni and G. Nagy, "Style context with second-order statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 14–22, 2005.
- [12] M. S. Cheema, A. Eweiwi, and C. Bauckhage, "Human activity recognition by separating style and content," *Pattern Recognition Letters*, vol. 50, pp. 130–138, 2014.
- [13] X. Y. Zhang, K. Huang, and C. L. Liu, "Pattern field classification with style normalized transformation," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pp. 1621–1626, Catalonia, Spain, 2011.
- [14] H. C. Jiang, K. Z. Huang, and R. Zhang, "Field support vector regression," in *Neural Information Processing: 24th International Conference*, pp. 699–708, Guangzhou, China, 2017.
- [15] K. Z. Huang, H. C. Jiang, and X. Y. Zhang, "Field support vector machines," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 6, pp. 454–463, 2017.
- [16] Y. Zhang, F. Chung, and S. Wang, "Fast reduced set-based exemplar finding and cluster assignment," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 5, pp. 917–931, 2019.
- [17] G. Marzinotto, J. C. Rosales, M. A. El-Yacoubi, and S. Garcia-Salicetti, "Age and gender characterization through a two layer clustering of online handwriting," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 428–439, Catania, Italy, 2015.
- [18] J. Suykens, T. van Gestel, J. de Brabanter, B. de Moor, and J. Vandewalle, "Least squares support vector machines," *International Journal of Circuit Theory and Applications*, vol. 27, no. 6, pp. 605–615, 2002.
- [19] Y. Jiang, Z. H. Deng, F. L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [20] M. E. Wall, A. Rechtsteiner, and L. M. Rocha, "Singular value decomposition and principal component analysis," in *A Practical Approach to Microarray Data Analysis*, D. P. Berrar, W. Dubitzky, and M. Granzow, Eds., pp. 91–109, Springer, Boston, MA, 2003.
- [21] M. Lu, L. H. Liu, and L. H. Wu, "Research on multi-kernel support vector data description method of classification," *Computer Engineering and Applications*, vol. 52, no. 18, pp. 68–73, 2016.
- [22] H.-Q. Wang, F.-C. Sun, Y.-N. Cai, N. Chen, and L.-G. Ding, "On multiple kernel learning methods," *Acta Automatica Sinica*, vol. 36, no. 8, pp. 1037–1050, 2010.
- [23] X. Chen, N. Guo, Y. Ma, and G. Chen, "More efficient sparse multi-kernel based least square support vector machine," *Communications and Information Processing*, vol. 289, pp. 70–78, 2012.
- [24] M. Kloft, " $\ell_p$ -norm multiple kernel learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 348–353, 2007.
- [25] V. Manic and R. Bodla, "More generality in efficient multiple kernel learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 1065–1072, Quebec, Canada, 2009.
- [26] C. Cortes and M. Mohri, "Learning non-linear combinations of kernels Advances in Neural Information Processing Systems 22," in *23rd Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, 2009.
- [27] G. Mehmet and E. Alpaydin, "Localized multiple kernel learning," in *Proceedings of the 25th international conference on Machine learning*, pp. 352–359, Helsinki, Finland, 2008.

- [28] C. Nello and S. John, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
- [29] Z. L. Xu, R. Jin, C. Stephane, H. Yang, I. King, and M. R. Lyu, "Simple and efficient multiple kernel learning by group lasso," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 1175–1182, Haifa, Israel, 2010.
- [30] C. Corinna, M. Mehryrar, and R. Afshin, "Two-stage learning kernel algorithms," in *Proceedings of the 27th Annual International Conference on Machine Learning (ICML 2010)*, pp. 239–246, Haifa, Israel, June 21-24, 2010.
- [31] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simple MKL," *The Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2491–2521, 2008.
- [32] F. Aiolli and M. Donini, "EasyMKL: a scalable multiple kernel learning algorithm," *Neurocomputing*, vol. 169, pp. 215–224, 2015.

## Research Article

# Automatic Detection of Coronary Metallic Stent Struts Based on YOLOv3 and R-FCN

Xiaolu Jiang<sup>1</sup>, Yanqiu Zeng<sup>1</sup>, Shixiao Xiao<sup>1</sup>, Shaojie He<sup>2</sup>, Caizhi Ye<sup>3</sup>, Yu Qi<sup>4</sup>,  
Jiangsheng Zhao<sup>2</sup>, Dezhi Wei<sup>1</sup>, Muhua Hu<sup>1</sup>, and Fei Chen<sup>5</sup>

<sup>1</sup>Chengyi University College, Jimei University, Xiamen 361021, China

<sup>2</sup>School of Informatics, Xiamen University, Xiamen 361021, China

<sup>3</sup>School of Management, Xiamen University, Xiamen 361021, China

<sup>4</sup>School of Mathematical Sciences, Xiamen University, Xiamen 361021, China

<sup>5</sup>Department of Cardiology, Tongji Hospital of Tongji University, Shanghai 200065, China

Correspondence should be addressed to Shixiao Xiao; [xiaoshixiao@jmu.edu.cn](mailto:xiaoshixiao@jmu.edu.cn) and Fei Chen; [chenfei1500843@163.com](mailto:chenfei1500843@163.com)

Received 11 May 2020; Accepted 29 July 2020; Published 1 September 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Xiaolu Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An artificial stent implantation is one of the most effective ways to treat coronary artery diseases. It is vital in vascular medical imaging, such as intravascular optical coherence tomography (IVOCT), to be able to track the position of stents in blood vessels effectively. We trained two models, the “You Only Look Once” version 3 (YOLOv3) and the Region-based Fully Convolutional Network (R-FCN), to detect metal support struts in IVOCT, respectively. After rotating the original images in the training set for data augmentation, and modifying the scale of the conventional anchor box in both two algorithms to fit the size of the target strut, YOLOv3 and R-FCN achieved precision, recall, and AP all above 95% in 0.4 IoU threshold. And R-FCN performs better than YOLOv3 in all relevant indicators.

## 1. Introduction

Coronary artery disease (CAD) is one of the most frequent causes of death despite being treatable. For treating the obstructive plaques, stenting is commonly used of the bare metal stent (BMS), the drug-eluting stent (DES), or bioresorbable vascular scaffolds (BVS). After implantation, the stents have to be assessed to detect malposition or endothelialisation. Intravascular optical coherence tomography (IVOCT) is one of imaging modality with the resolution and contrast necessary to enable accurate measurements of luminal architecture and neointima stent coverage. Figure 1 shows an IVOCT image frame after metallic stent implantation. However, since a pullback of the IVOCT image sequence for a single patient often contains hundreds of images and thousands of struts, it is labour-intensive and time-consuming to conduct a quantitative evaluation for every patient manually. Therefore, a

fully automatic method for metallic strut analysis is highly desired. Until now, several different strategies [1–19] have been proposed for the detection of stent strut candidates in IVOCT and the removal of false positives.

Since metallic struts appear as high-reflecting spots followed by trailing shadows in IVOCT images, as shown in Figure 1, most algorithms are searching for these features to detect stent struts [1]. Lu et al. [2] trained a bagged decision tree classifier, using specific features extracted from the images to classify the candidate stent struts. Han et al. [3] applied the Laplacian filter to the image in the polar coordinates map to extract corners and edges and then used the intensity threshold to identify the stent struts. Nam et al. [4] detected the candidate struts by IVOCT intensity image and gradient image, and then by using a hidden layer and a ten-node artificial neural network determines the candidate struts. Migliori et al. [5] classified pixels associated with high



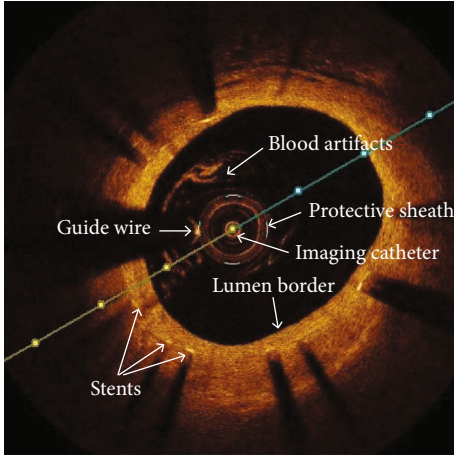


FIGURE 1: IVOCT image after metallic stent implantation.

slopes as candidate struts and applied a penalty function away from the lumen contour structure.

Alternative approaches for stent strut detection as follow. A controllable filter is designed by Xu et al. [6] to calculate the local ridge strength and direction to locate the deeply buried struts. Wang et al. [7] used the Bayesian network and the stent mesh information of the adjacent frame to determine the location of the struts in the A-scan. They used the graph cut algorithm to simultaneously locate the exact struts depth positions in the IVOCT pullback.

In recent years, a deep learning framework has achieved excellent results in the computer visual object detection and recognition domain, and it has attracted increasing attention and led to more research based on this framework. Traditional machine learning methods depend on manually designed features. Unlike that, novel representation patterns or models are automatically learned from low-level features to high-level semantics in deep learning, which often makes the detection performance more correct and robust. BVS detection in IVOCT images based on deep learning has been reported recently. Cao et al. [8] constructed a region-based fully convolutional network (R-FCN) detector for BVS detection in IVOCT images. Zhou et al. [9] proposed an automatic detection method for BVS based on a U-shaped convolutional neural network. Gessert et al. [10] can predict whether image slices contain metal supports, BVS, or do not contain any equipment only using image-level tags by a trained convolutional neural network, achieving 99.0% classification accuracy. However, there are few methods for detecting metallic struts based on deep learning. Given this, in this paper, we attempt to use two deep learning object detection models to detect metallic struts and compare the performance.

Conventional deep learning models for object detection fall into two types: one-stage and two-stage. YOLOv3 and R-FCN are, respectively, typical algorithms of these two types, and also are frequently used in the medical field. Wu et al. [11] developed a deep learning model (BMSNet) with the YOLOv3 architecture for assisting haematologists in the interpretation of bone marrow smears for faster diagnosis and disease monitoring. Park et al. [12] compared the performance of various state-of-the-art deep-learning architec-

tures, including YOLOv3, for detecting the optic nerve head and vertical cup-to-disc ratio in fundus images. Safdar et al. [13] highlighted the most suitable Data Augmentation technique for medical imaging by using YOLOv3. Wu et al. [14] investigated the potential for using Principal Component Analysis (PCA) and Adaptive Median Filter (AMF) to improve four algorithms, including R-FCN and YOLOv3. Zhang et al. [15] proposed a novel abnormal region detection approach for cervical screening based on R-FCN. Morrell et al. [16] presented a neural net architecture based on R-FCN to suit mammograms.

Since YOLOv3 and R-FCN perform well in medical fields, we used them in this paper for metallic stent struts detection and tried to compare the performance of these two models systematically. We also realised the data augmentation of the existing training set through images rotation to enhance the advantage of big data in feature extraction. To explore the use of anchor box in specialized fields, we also adjusted its size to suit the detection of metallic stent struts: *k*-means clustering in YOLOv3, manually fixed in R-FCN.

## 2. Material and Methods

**2.1. Dataset.** For validating the algorithm, ten pull-back runs were acquired with an IVOCT imaging system from a baseline study. The pull-back speed was 15 mm/s. All of the struts were metallic struts. The total stent length was 21 2.17 mm. The different patients who participated in the study were independent of each other. As shown in Figure 1, the IVOCT image contains the stent, guidewire, imaging catheter, protective sheath, blood artefacts, and lumen border. To assist medical personnel in judging the location and performance of the stent, we need to identify the metallic stent in these complex backgrounds automatically. There are 165 IVOCT images, and each image has about 3~22 metallic stent struts, which has manually marked all the stent struts as the ground truth by rectangular frames.

**2.2. Deep Learning Object Detection Model.** There are two types of deep learning models for object detection: one-stage and two-stage. Two-stage object detection strategy consists of: (i) region proposal, and (ii) region classification. Typical two-stage model includes R-CNN [20], Fast R-CNN [21], Faster R-CNN [22], and R-FCN [23]. The one-stage model is an end-to-end algorithm. It does not need to generate candidate frames and directly transform the problem of object frames positioning into a regression problem. The typical 1-stage model includes the YOLO series [24–26] and SSD [27]. Generally speaking, the method based on candidate regions has higher accuracy, but the end-to-end way has distinct advantages in speed. In this paper, R-FCN and YOLOv3 are compared, and they are used to detect the metallic stent struts in the IVOCT image.

**2.3. YOLOv3.** Given the input image, YOLOv1 directly returns the object's bounding box and its category at multiple locations in the image. YOLOv2 and YOLO9000 introduced anchor boxes to predict the offset and confidence of the anchor boxes instead of directly predicting the coordinate

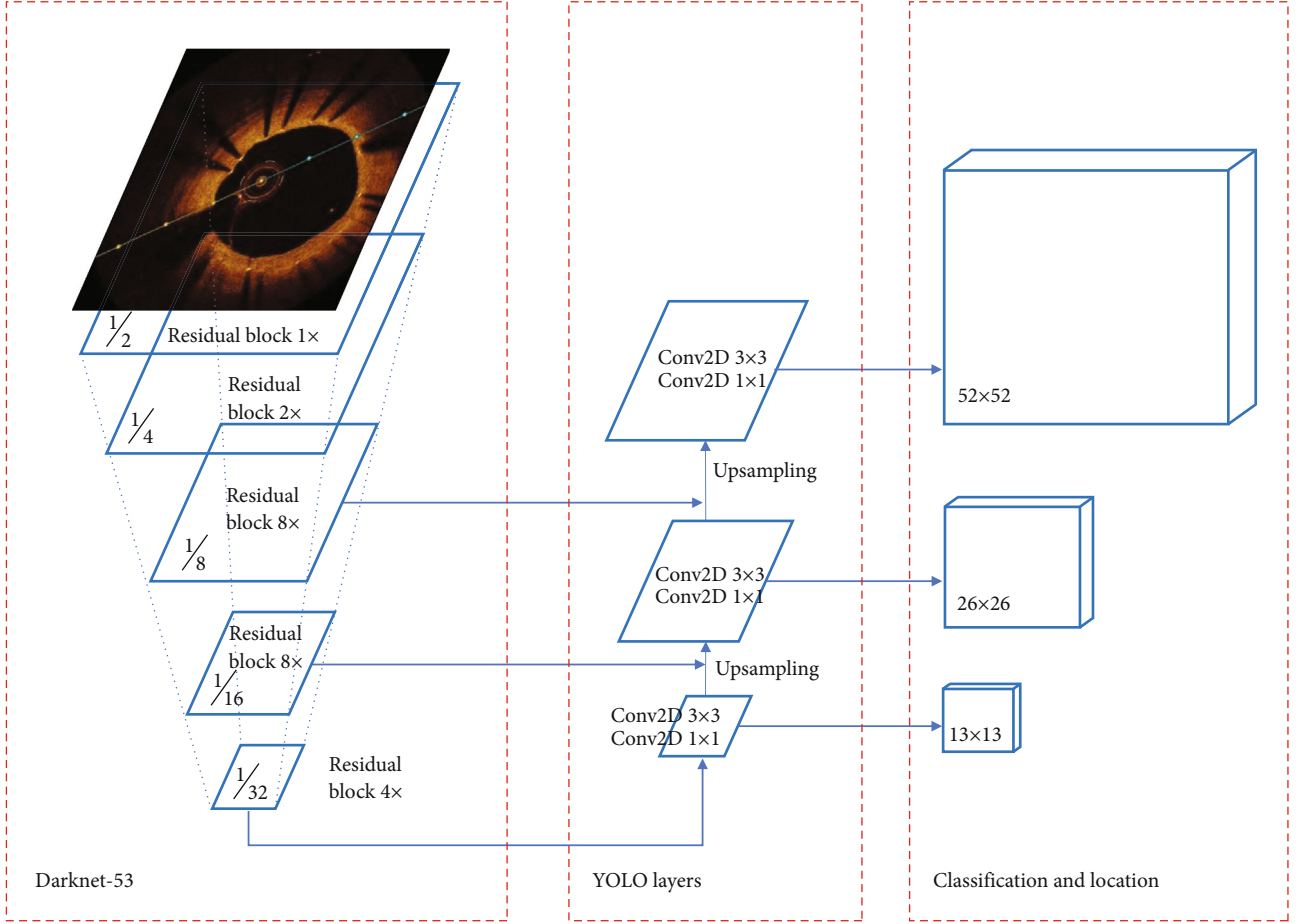


FIGURE 2: Architecture of metallic stent detection based on YOLOv3.

values. By adding a pass-through layer, the high-resolution shallow features are connected to the low-resolution features for fusion and detection. YOLOv3 detects objects on multiple fusion feature maps separately, which improves efficiency in the detection of smaller objects. At the same time, the classification uses multiple logistic classifiers instead of a softmax classifier, which is used to solve the multilabel classification problem in YOLOv2.

**2.3.1. Overall Architecture of YOLOv3.** The network architecture of YOLOv3 (Figure 2) is divided into three parts: darknet53 for feature extraction, YOLO layers for feature fusion, and classification and location. Darknet53 has a total of 53 convolutional layers, and the rest are residual layers. The YOLO layers are used for feature fusion to generate three scale feature maps. It takes feature maps from earlier in the network and merges it with the upsampled features using concatenation. Object classification and locating are carried out on the feature fusion maps of three scales ( $13 \times 13$ ,  $26 \times 26$ , or  $52 \times 52$ ), respectively, to the different size objects for detection.

**2.3.2. Unified Detection of YOLOv3.** Taking the  $13 \times 13$  fusion feature map as an example, YOLOv3 divides the map into  $13 \times 13$  grids. If the center of an object falls into a grid cell, the grid cell is responsible for detecting the object.

Each grid cell predicts three bounding boxes, thus, returning  $3 \times (4 + 1 + C)$  tensors, of which four bounding box offsets, one confidence score, and  $C$  conditional class probabilities. Four bounding box offsets refer to the offsets from the given anchor box. Each scale needs three anchor boxes as bounding boxes prior, so a total of 9 anchor boxes are clustered from our data set before. Including all cells, the scale feature map outputs  $13 \times 13 \times 3 \times (5 + C)$  tensors. Adding the output of  $26 \times 26$  and  $52 \times 52$  scale feature maps, we get a total of  $(13 \times 13 + 26 \times 26 + 52 \times 52) \times 3 \times (5 + C)$  tensor.

As shown in Figure 3, the four bounding box offsets  $t_x$ ,  $t_y$ ,  $t_w$ ,  $t_h$  can be converted into the center coordinates  $b_x$ ,  $b_y$  and the width  $b_w$  and the height  $b_h$  of the bounding box by formula:

$$b_x = \sigma(t_x) + c_x, \quad (1)$$

$$b_y = \sigma(t_y) + c_y, \quad (2)$$

$$b_w = p_w e^{t_w}, \quad (3)$$

$$b_h = p_h e^{t_h}, \quad (4)$$

where  $P_w$  and  $P_h$  are the width and height of the prior box,  $C_x$  and  $C_y$  are the offsets of the responsible grid from the upper left corner of the image, and  $\sigma$  is the sigmoid function.

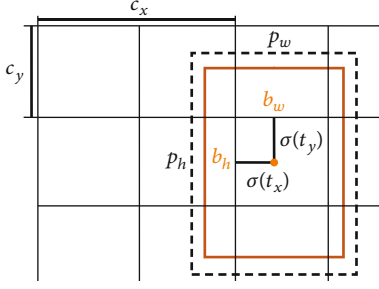


FIGURE 3: Bounding boxes with dimension priors and location prediction.

The objectness score reflects the confidence that the grid cell contains objects and the accuracy of predicting that the cell contains objects,

$$\text{objectness score} = \Pr(\text{object}) \times \text{IoU}_{\text{pred}}^{\text{truth}}. \quad (5)$$

When there are objects in the cell, the objectness score will be equal to the intersection over union (IoU) between the bounding box and the ground truth:

$$\text{IoU}_{\text{pred}}^{\text{truth}} = \frac{\text{ground truth box} \cap \text{predicted bounding box}}{\text{ground truth box} \cup \text{predicted bounding box}}. \quad (6)$$

C conditional class probabilities  $\Pr(\text{class}_i|\text{object})$  are conditioned on the grid cell containing an object. The final category of confidence is

$$\begin{aligned} & \Pr(\text{class}_i|\text{object}) \times \Pr(\text{object}) \times \text{IoU}_{\text{pred}}^{\text{truth}} \\ &= \Pr(\text{class}_i) \times \text{IoU}_{\text{pred}}^{\text{truth}}. \end{aligned} \quad (7)$$

**2.3.3. Training YOLOv3.** The final loss function will summarize the losses of the three scales. During training, the error function of each scale includes a localization error, a confidence error, and a classification error. Using the formula (1)–(4) to inverse the four coordinates  $\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i$  corresponding to the ground truth in cell  $i$ , we can calculate SSE of the corresponding predicted coordinates  $x_i, y_i, w_i, h_i$  as the localization error. YOLOv3 uses logistic regression to predict the confidence score  $c_i$ , and the actual score  $\hat{c}_i$  is depending on the IoU of the bounding box prior and ground truth. Then, the binary cross-entropy of the predicted and actual confidence score is the confidence loss. YOLOv3 uses independent logistics instead of softmax as the classifier. For each category, binary cross-entropy is also used as the loss function. Two parameters  $\lambda_{\text{coord}}$  and  $\lambda_{\text{noobj}}$  can adjust the balance of the loss from bounding box coordinate predictions and the loss from confidence predictions for boxes that do not contain objects. The final loss a function is

$$\text{Loss} = \text{Error}_{\text{localization}} + \text{Error}_{\text{confidence}} + \text{Error}_{\text{class}},$$

$$\begin{aligned} \text{Error}_{\text{localization}} &= \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ &\quad + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (2 - w_i \times h_i) \\ &\quad \cdot \left[ (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right], \\ \text{Error}_{\text{confidence}} &= - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \times [\hat{c}_i \times \log c_i + (1 - \hat{c}_i) \\ &\quad \times \log(1 - c_i)] - \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} \\ &\quad \times [\hat{c}_i \times \log c_i + (1 - \hat{c}_i) \times \log(1 - c_i)], \\ \text{Error}_{\text{class}} &= - \sum_{i=0}^{S^2} I_i^{\text{obj}} \sum_{c \in \text{classes}} [\hat{p}_i(c) \times \log p_i(c) \\ &\quad + (1 - \hat{p}_i(c)) \times \log(1 - p_i(c))], \end{aligned} \quad (8)$$

where  $S^2$  is the number of grid cells,  $B$  is the number of anchor boxes. By minimizing the loss function to learn the weights, we can obtain the location of the bounding box and the category prediction.

**2.4. Region-Based Fully Convolutional Networks (R-FCN).** R-FCN is a typical two-stage object detection method. In the first stage, the Regional Proposal Network (RPN) is used for regional proposals to generate candidate RoI. In the second stage, R-FCN uses position-sensitive score maps to synthesize the features of different positions of ROIs so that the network can solve the dilemma between the translation invariance in classification and the translation variance in object detection. At the same time, all the learnable weight layers are convolutional and can be calculated in the whole image. Finally, the entire network reaches the structure of full convolution, which significantly improves efficiency.

**2.4.1. Overall Architecture of R-FCN.** The overall architecture of the metallic stent strut detection based on R-FCN is shown in Figure 4. After extracting features through a series of convolutions in Resnet-50, a Region Proposal Network (RPN) uses a small sliding window and anchor boxes to generate candidate regions on a whole feature map. For the metallic stent strut and the background, the feature map of the entire image is, respectively, connected with  $3 \times 3$  position-sensitive score maps by convolution. Combining the RoI pooling of 9 position-sensitive scores, the category probability corresponding to each RoI can be voted. The four localization parameters that represent the offset from the anchor boxes are also obtained by voting similarly. After training the network, R-FCN outputs the adjusted new position and score of the metallic stent strut RoIs as “R-FCN output.” If the category score of each RoI is less than the score threshold, we

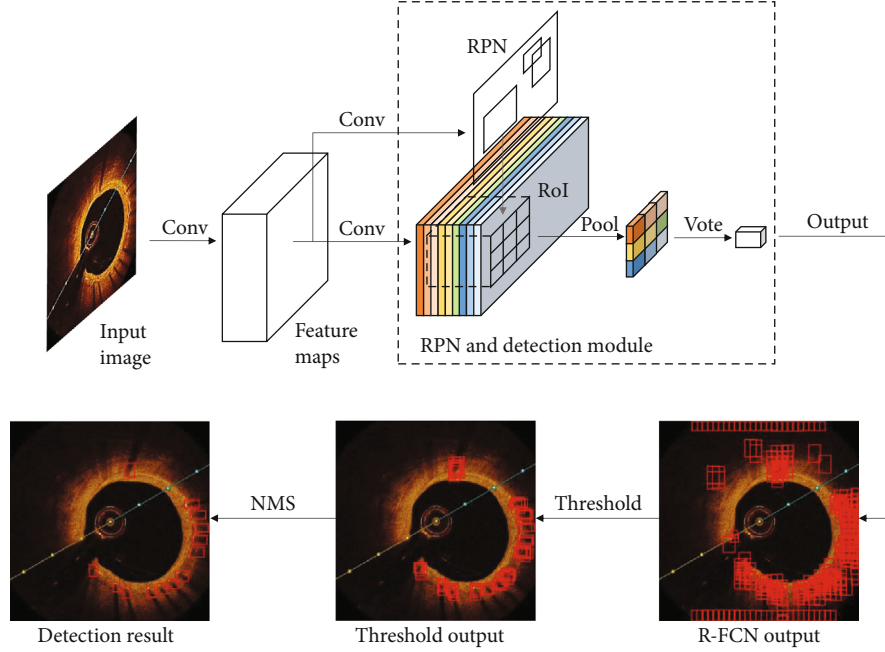


FIGURE 4: Architecture of metallic stent detection based on R-FCN.

remove the bounding box to get a “Threshold output.” The remaining bounding boxes still have a lot of overlap. Run a nonmaximum suppression (NMS), and only the bounding box with the highest score is kept where the IoU exceeds a certain threshold. The remaining bounding box is the final “Detection result.”

**2.4.2. Region Proposal Network (RPN).** RPN uses a fully convolutional network to output a set of rectangular region proposals at once on the entire feature map. Slide a small sliding window on the feature map, and use each area located by it as input. If  $k$  ( $k = 9$ ) anchor boxes are used as the regression reference, each sliding window will output  $4k$  coordinate regression  $t_x, t_y, t_w, t_h$  and  $2k$  bounding box classification to estimate the probability that each proposal is the object or not.

The RPN loss function consists of two parts, the log classification loss, and the smooth regression loss:

$$\begin{aligned}
 L(\{p_i\}, \{t_i\}) &= \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*), \\
 L_{cls}(p_i, p_i^*) &= -p_i^* \times \log p_i - (1 - p_i^*) \times \log (1 - p_i), \\
 L_{reg}(t_i, t_i^*) &= \text{smooth}_{L_1}(t_x - t_x^*) + \text{smooth}_{L_1}(t_y - t_y^*) \\
 &\quad + \text{smooth}_{L_1}(t_w - t_w^*) + \text{smooth}_{L_1}(t_h - t_h^*),
 \end{aligned} \tag{9}$$

where the smooth  $L_1$  is defined by

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1, \\ |x| - 0.5 & \text{otherwise,} \end{cases} \tag{10}$$

$\{p_i\}, \{t_i\}$  are the outputs of the anchor in the classification layer and regression layer. During training, we assign labels to the anchor based on the IoU of the anchor  $i$  and the ground truth box. A positive label is 1, and a negative label is 0.  $t_i^*$  is the vector about the ground truth box location associated with the positive anchor.

RPN only relies on a single-scale image and feature mapping, uses a single-size filter, and thus generates a region proposal that is translation-invariant. Shared features require no additional cost to process the scale of the object.

**2.4.3. Position-Sensitive Score Maps.** The innovation of R-FCN is the position-sensitive score map. Object classification and location all need  $3 \times 3$  score maps. We take the position-sensitive score maps of the stent strut classification as an example. 9 position-sensitive score maps correspond to features of nine positions of the strut. Each position-sensitive map in the RoI area is divided into  $3 \times 3$  bins, and a position-sensitive RoI pooling operated only over the appropriate bin of each score map:

$$r_c(i, j | \Theta) = \sum_{(x, y) \in \text{bin}(i, j)} z_{i, j, c}(x + x_0, y + y_0 | \Theta) / n. \tag{11}$$

Nine pool responses vote on the RoI by averaging; then, the classification probability of RoI is output by the softmax function.

$$\begin{aligned}
 r_c(\Theta) &= \sum_{i, j} r_c(i, j | \Theta), \\
 s_c(\Theta) &= e^{r_c(\Theta)} / \sum_{c'=0}^C e^{r_{c'}(\Theta)}.
 \end{aligned} \tag{12}$$

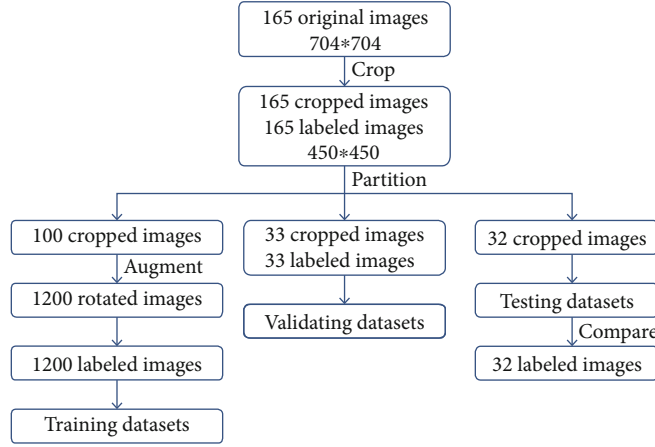


FIGURE 5: Data preprocessing.

TABLE 1: Comparisons between R-FCN and YOLOv3 algorithms corresponding to various IoU threshold. The amount of stents for testing is 425.

IoU	TP		FP		Precision		Recall		AP	
	R-FCN	YOLOv3	R-FCN	YOLOv3	R-FCN	YOLOv3	R-FCN	YOLOv3	R-FCN	YOLOv3
0.30	409	410	1	12	99.8%	97.2%	96.2%	96.5%	96.2%	96.0%
0.35	408	409	2	13	99.5%	96.9%	96.0%	96.2%	96.0%	95.5%
0.40	408	407	2	15	99.5%	96.4%	96.0%	95.8%	96.0%	95.0%
0.45	407	402	3	20	99.3%	95.3%	95.8%	94.6%	95.7%	92.7%
0.50	403	391	7	31	98.3%	92.7%	94.8%	92.0%	94.2%	88.7%
0.55	386	376	24	46	94.1%	89.1%	90.8%	88.5%	88.4%	81.9%
0.60	353	347	57	75	86.1%	82.2%	83.1%	81.6%	76.5%	69.6%

Bounding box regression is similar, except that the output after voting is the 4 d vector  $(t_x, t_y, t_w, t_h)$ .

The loss function for each RoI includes cross-entropy loss for classification and regression loss for the location of the positive sample:

$$L(s, t_{x,y,w,h}) = L_{cls}(s_{c^*}) + \lambda[C^* > 0]L_{reg}(t, t^*), \quad (13)$$

$$L_{cls}(s_{c^*}) = -s_{c^*} \times \log s - (1 - s_{c^*}) \times \log(1 - s).$$

Regression loss is the same as RPN's.  $C^*$  represents the label of the RoI.  $[C^* > 0]$  means that if the label is positive, it is equal to 1; otherwise, it is 0.

**2.5. Performance Measures.** precision (P), recall (R), and AP are principal quantitative indicators for algorithm performance evaluation in deep learning, which are employed in this experiment.

Denote by  $TP$ ,  $FP$ , and  $FN$  the numbers of true positives, false positives, and false negatives, respectively. Then, precision and recall are computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (14)$$

$$\text{Recall} = \frac{TP}{TP + FN}.$$

Here, whether a bounding box belongs to TP or FP depends on the IoU threshold of the ground truth and bounding box.

Here, AP refers to the average precision, the area under the P-R curve by numerical integration. The computation of it is shown as follows:

$$AP = \sum_n (R_n - R_{n-1})P_n, \quad (15)$$

where  $P_n$  and  $R_n$  are the precision and recall at the  $n$ th threshold.

### 3. Results and Discussion

**3.1. Data Preprocessing.** To effectively detect the metallic stent strut, we cropped the extraneous edges in all the IVOCT images, so that the image size changes from  $704 * 704$  to  $450 * 450$ . Of all 165 IVOCT images, we used 100 images as the training set, 33 images as the verification set for adjusting hyperparameters, and 32 images as the test set. To augment the training of samples, we rotated the training set images. Along the catheter centre, a new training set image is generated every 30 degrees of rotation, and finally, 1200 images are obtained as the training set (Figure 5).

**3.2. Parameters Setting.** Only one type of metallic stent strut is to be detected. We take  $C$  the number of categories in



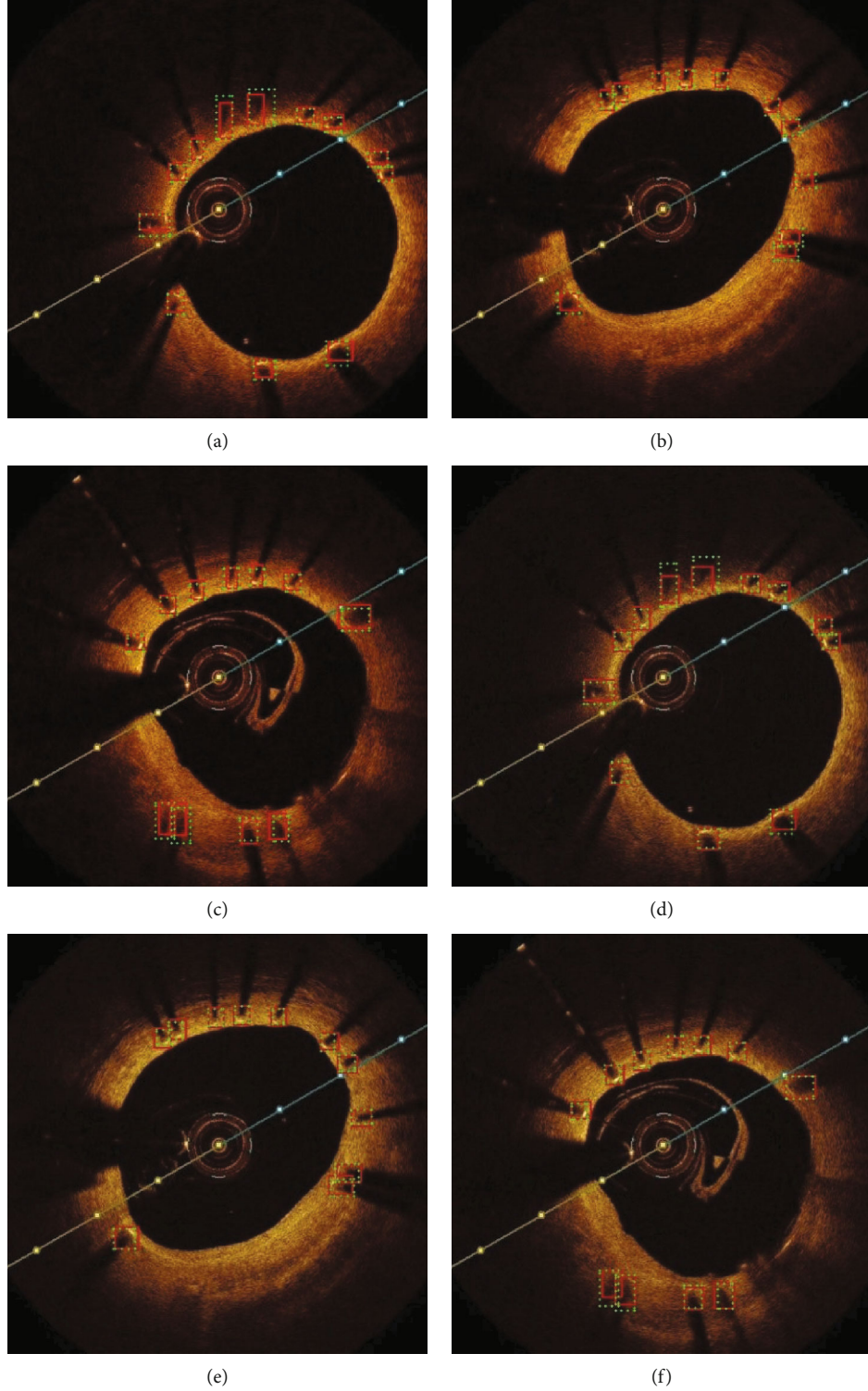


FIGURE 6: Examples of metallic stents detection results by YOLOv3 (a–c) or by R-FCN (d–f). The green dashed boxes refer to the ground truth, and those in red refer to bounding boxes (when IoU threshold = 0.4).

YOLOv3 and R-FCN as 1. Due to the relatively small size of the stent struts, the anchor box should be different from the usual. Through the *K*-means algorithm, nine anchor boxes were clustered in YOLOv3 with the data set, which size results in  $12 \times 14$ ,  $14 \times 18$ ,  $15 \times 15$ ,  $18 \times 18$ ,  $19 \times 26$ ,  $19 \times 15$ ,

$24 \times 19$ ,  $29 \times 26$ ,  $30 \times 16$ . As a comparison, the anchor boxes in R-FCN is manually fixed to the length of  $\{8, 16, 32\}$  and the ratio of  $\{0.85, 1, 1.85\}$ .

In YOLOv3, we set 0.1 as the IoU threshold to mark positive labels, and the threshold in the objectiveness score is also

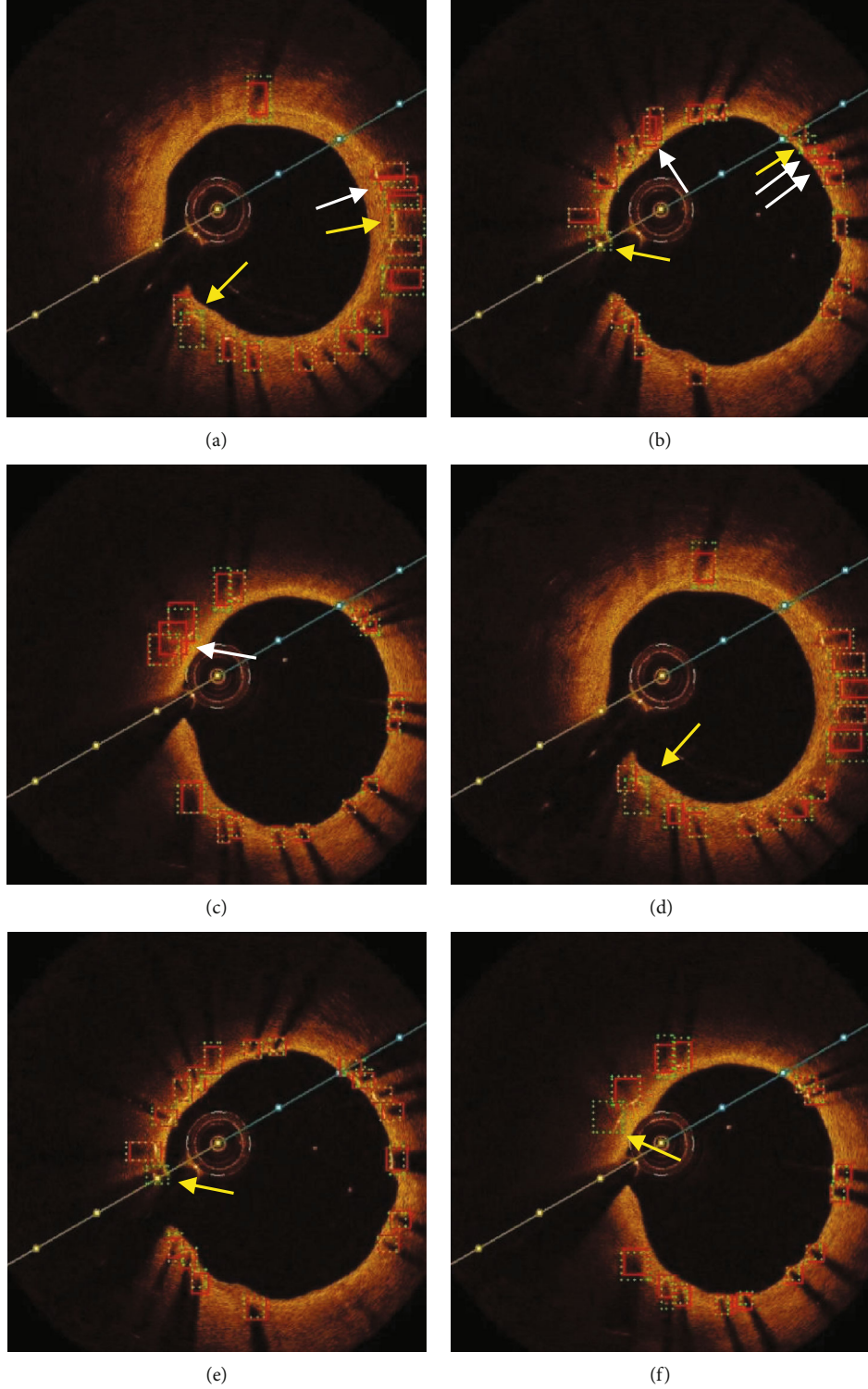


FIGURE 7: Examples of metallic stents detection result by YOLOv3 (a–c) or by R-FCN (d–f). The boxes which are pointed at by white arrow and yellow arrow refer to false positives and false negatives, respectively.

set to 0.1. The coordinate weight  $\lambda_{\text{coord}}$  and the no object weight  $\lambda_{\text{noobj}}$  in the loss function adopt the default values of 5 and 0.5.

In R-FCN, the positive overlap in RPN has a threshold of 0.7, while the threshold in “R-FCN output” is 0.1, and in NMS, it is 0.3.

**3.3. Results and Discussion.** The test results are shown in Table 1. We compared the performance of YOLOv3 and R-FCN corresponding to different IoU between the bounding box and the ground truth. As the IoU threshold gradually increases, the precision, recall, and AP decrease slowly in both algorithms. When the IoU threshold is less than 0.45,

all the indicators are above 92.7%. When 0.4 IoU threshold, they even all reach above 95%. And it is not hard to find that the R-FCN is superior to the YOLOv3 for any of the IoU thresholds.

Table 1 shows that the difference between YOLOv3 and R-FCN in precision is higher than that in the recall. It indicates that false positives (FP) are more likely to occur in YOLOv3 than false negatives (FN). For example, when the IoU threshold is 0.4, the number of false positives based on R-FCN is only 2, but yolov3 reaches 15. The difference between the two methods in the recall is only 0.2%, but in precision is 3.1%.

Examples of metallic stents detecting results got by YOLOv3 and R-FCN in the same image sets show more comparison in Figures 6 and 7 (when IoU = 0.4). The green dashed boxes refer to the ground truth, and those in red refer to bounding boxes in both figures. The boxes which are pointed at by the white arrow in Figure 7 refer to false positives, while those by yellow arrow refer to false negatives. Figure 6 shows that both algorithms perform quite well in metallic stents detection. But it is easy to find that YOLOv3 has some false positives while R-FCN does not have in the same image sets in Figure 7. R-FCN has better performance in samples with unobvious characteristics, most of which are located in the areas where the color changes or the stent struts are denser.

In general, both of YOLOv3 and R-FCN algorithms performed pretty well in metallic stents detection (Figure 6(a)–6(c) and Figures 6(d)–6(f)). However, R-FCN has better performance in obscure samples, such as images with intimal hyperplasia or noise interference (Figures 7(a)–7(c) and Figures 7(d)–7(f)).

## 4. Conclusion

In this paper, we presented two automatic methods for metallic stents detection based on YOLOv3 (one-stage) and R-FCN (two-stage), respectively. To augment the data, we rotated the images of the training data set. And we adjusted the size of the anchor box to adapt to the detection of small objects. The experiments demonstrate that both algorithms perform fairly well whether the characteristic of metallic stents is clear or blurred (on account of intimal hyperplasia and noise interference). When the IoU threshold of the ground truth and bounding box is set to 0.4, precision, recall, and AP all reach above 95%. Nevertheless, R-FCN performs better than YOLOv3 in all relevant indicators, as shown in Table 1. The precision of R-FCN reaches more than 99.3% when the IoU threshold is less than or equal to 0.45. The future work will mainly focus on adding the complexity of the network, combining multiple algorithms for reinforcement learning to improve the performance further.

## 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 Talent Training Program for Distinguished Young Scholars of Fujian Province (Grant No. ZX17033), the Education and Scientific Research Project for Young and Middle-aged Teachers of Fujian Province (Grant Nos. JT180874 and JAT191153), and the Startup Project for Doctor scientific research of Chengyi University College, Jimei University (Grant No. CK18013). The authors are very grateful to all anonymous reviews for their valuable comments that help improve the quality of this paper.

## References

- [1] C. Chiastra, S. Migliori, F. Burzotta, G. Dubini, and F. Migliavacca, "Patient-Specific Modeling of Stented Coronary Arteries Reconstructed from Optical Coherence Tomography: Towards a Widespread Clinical Use of Fluid Dynamics Analyses," *Journal of Cardiovascular Translational Research*, vol. 11, no. 2, pp. 156–172, 2018.
- [2] H. Lu, M. Gargsha, Z. Wang et al., "Automatic stent detection in intravascular OCT images using bagged decision trees," *Bio-medical Optics Express*, vol. 3, no. 11, pp. 2809–2824, 2012.
- [3] M. Han, D. Kim, W. Y. Oh, and S. Ryu, "High-speed automatic segmentation of intravascular stent struts in optical coherence tomography images," in , Article ID 85653Z *Proceedings of the Conference on Photonic Therapeutics and Diagnostics IX, Photonic Therapeutics and Diagnostics IX*, vol. 8565, San Francisco, CA, February 2013.
- [4] H. S. Nam, C.-S. Kim, J. J. Lee, J. W. Song, J. W. Kim, and H. Yoo, "Automated detection of vessel lumen and stent struts in intravascular optical coherence tomography to evaluate stent apposition and neointimal coverage," *Medical Physics*, vol. 43, no. 4, pp. 1662–1675, 2016.
- [5] S. Migliori, C. Chiastra, M. Bologna et al., "A framework for computational fluid dynamic analyses of patient-specific stented coronary arteries from optical coherence tomography images," *Medical Engineering & Physics*, vol. 47, pp. 105–116, 2017.
- [6] C. Xu, J. M. Schmitt, T. Akasaka, T. Kubo, and K. Huang, "Automatic detection of stent struts with thick neointimal growth in intravascular optical coherence tomography image sequences," *Physics in Medicine and Biology*, vol. 56, no. 20, pp. 6665–6675, 2011.
- [7] Z. Wang, M. W. Jenkins, G. C. Linderman et al., "3-D stent detection in intravascular OCT using a Bayesian network and graph search," *IEEE Transactions on Medical Imaging*, vol. 34, no. 7, pp. 1549–1561, 2015.
- [8] Y. Cao, Y. Lu, J. Li et al., "Deep learning based bioresorbable vascular scaffolds detection in IVOCT images," in *Proceedings of the 24th International Conference on Pattern Recognition (ICPR)*, pp. 3778–3783, Beijing, China, August 2018.
- [9] W. Zhou, F. Chen, Y. Zong et al., "Automatic detection approach for bioresorbable vascular scaffolds using a U-shaped convolutional neural network," *IEEE Access*, vol. 7, pp. 94424–94430, 2019.



- [10] N. Gessert, S. Latus, Y. S. Abdelwahed, D. M. Leistner, M. Lutz, and A. Schlaefer, "Bioresorbable scaffold visualization in IVOCT images using CNNs and weakly supervised localization," in , Article ID 109492C*Proceedings of the Conference on Medical Imaging*, vol. 10949, San Diego, CA, February 2019.
- [11] Y. Y. Wu, T. C. Huang, R. H. Ye et al., "A hematologist-level deep learning algorithm (BMSNet) for assessing the morphologies of single nuclear balls in bone marrow smears: algorithm development," *JMIR medical informatics*, vol. 8, no. 4, article e15963, 2020.
- [12] K. Park, J. Kim, and J. Lee, "Automatic optic nerve head localization and cup-to-disc ratio detection using state-of-the-art deep-learning architectures," *Scientific reports*, vol. 10, no. 1, p. 5025, 2020.
- [13] M. F. Safdar, S. S. Alkobaisi, and F. T. Zahra, "A comparative analysis of data augmentation approaches for magnetic resonance imaging (MRI) scan images of brain tumor," *Acta informatica medica*, vol. 28, no. 1, pp. 29–36, 2020.
- [14] S. X. Wu, C. C. Guo, and X. H. Wang, "Application of principal component analysis and adaptive median filter to improve real-time prostate capsula detection," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 2, pp. 336–347, 2020.
- [15] J. Zhang, J. He, T. Chen, Z. Liu, and D. Chen, "Abnormal region detection in cervical smear images based on fully convolutional network," *IET Image Processing*, vol. 13, no. 4, pp. 583–590, 2019.
- [16] S. Morrell, Z. Wojna, C. S. Khoo, S. Ourselin, and J. E. Iglesias, "Large-scale mammography CAD with deformable convnets," in *Proceeding of the 3rd International Workshop on Reconstruction and Analysis of Moving Body Organs (RAMBO) / 4th International Workshop on Breast Image Analysis (BIA) / 1st International Workshop on Thoracic Image Analysis (TIA)*, vol. 11040, pp. 64–72, Granada, Spain, September 2018.
- [17] C. Huang, C. Wang, J. Tong, L. Zhang, F. Chen, and Y. Hao, "Automatic quantitative analysis of bioresorbable vascular scaffold struts in optical coherence tomography images using region growing," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 1, pp. 98–104, 2018.
- [18] C. Huang, Y. Xie, Y. Lan et al., "A new framework for the integrative analytics of intravascular ultrasound and optical coherence tomography images," *IEEE Access*, vol. 6, pp. 36408–36419, 2018.
- [19] C. Huang, Y. Peng, F. Chen et al., "A deep segmentation network of multi-scale feature fusion based on attention mechanism for IVOCT lumen contour," in *IEEE/ACM transactions on computational biology and bioinformatics*, p. 1, 2020.
- [20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587, Columbus, OH, June 2014.
- [21] R. Girshick, "Fast R-CNN," in *Proceeding of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, December 2015.
- [22] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS)*, Montreal, Canada, December 2015.
- [23] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: object detection via region-based fully convolutional networks," in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS)*, vol. 29, Barcelona, Spain, 2016.
- [24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceeding of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 779–788, Seattle, WA, June 2016.
- [25] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, Honolulu, HI, July 2017.
- [26] J. Redmon and A. Farhadi, *YOLOv3: an incremental improvement*, 2018.
- [27] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multi-Box detector," in *Proceedings of the 14th European Conference on Computer Vision (ECCV)*, vol. 9905, pp. 21–37, Amsterdam, Netherlands, October 2016.

## Research Article

# Instance Transfer Subject-Dependent Strategy for Motor Imagery Signal Classification Using Deep Convolutional Neural Networks

Kai Zhang,<sup>1,2</sup> Guanghua Xu<sup>1,2</sup>, Longtin Chen,<sup>1,2</sup> Peiyuan Tian,<sup>1,2</sup> ChengCheng Han,<sup>1,2</sup> Sicong Zhang,<sup>1</sup> and Nan Duan<sup>1</sup>

<sup>1</sup>School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

<sup>2</sup>State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an, China

Correspondence should be addressed to Guanghua Xu; [ghxu@xjtu.edu.cn](mailto:ghxu@xjtu.edu.cn)

Received 14 May 2020; Revised 16 July 2020; Accepted 29 July 2020; Published 28 August 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Kai 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 the process of brain-computer interface (BCI), variations across sessions/subjects result in differences in the properties of potential of the brain. This issue may lead to variations in feature distribution of electroencephalogram (EEG) across subjects, which greatly reduces the generalization ability of a classifier. Although subject-dependent (SD) strategy provides a promising way to solve the problem of personalized classification, it cannot achieve expected performance due to the limitation of the amount of data especially for a deep neural network (DNN) classification model. Herein, we propose an instance transfer subject-independent (ITSD) framework combined with a convolutional neural network (CNN) to improve the classification accuracy of the model during motor imagery (MI) task. The proposed framework consists of the following steps. Firstly, an instance transfer learning based on the perceptive Hash algorithm is proposed to measure similarity of spectrogram EEG signals between different subjects. Then, we develop a CNN to decode these signals after instance transfer learning. Next, the performance of classifications by different training strategies (subject-independent- (SI-) CNN, SD-CNN, and ITSD-CNN) are compared. To verify the effectiveness of the algorithm, we evaluate it on the dataset of BCI competition IV-2b. Experiments show that the instance transfer learning can achieve positive instance transfer using a CNN classification model. Among the three different training strategies, the average classification accuracy of ITSD-CNN can achieve  $94.7 \pm 2.6$  and obtain obvious improvement compared with a contrast model ( $p < 0.01$ ). Compared with other methods proposed in previous research, the framework of ITSD-CNN outperforms the state-of-the-art classification methods with a mean kappa value of 0.664.

## 1. Introduction

A brain-computer interface (BCI) is a communication method between a user and a computer that does not rely on the normal neural pathways of the brain and muscles. Motor imagery (MI), one of the paradigms of BCI, is a way of thinking that imitates the motor intention without real motion output; that is, the brain imagines the entire movement without contracting the muscles [1]. Research has shown that motor imagery (MI) can produce the same change of sensory motor rhythms as a real movement [2]. This phenomenon will cause energy increase or decrease in specific frequency bands of EEG, which are called event-related desynchronization (ERD) and event-related synchrony (ERS) [3]. The differences of ERD/ERS are always

used to decode mental intentions, control a robot, and execute rehabilitation training for stroke patients [4]. During this process, the accurate decoding of MI is the essential factor that determines the effectiveness and quality of the rehabilitation. However, due to the differences in physiological structure and physiological condition across subjects/trials, there will be obvious variations in feature distribution for EEG signals. Especially, as a spontaneous potential activity, the signal of MI is extremely weak and always accompanied with nonlinearity and nonstationary. It brings a huge challenge for the decoding model for MI.

With the development of machine learning (ML) and deep learning (DL) technology, more classification models are widely used for EEG decoding [5]. During the training stage of the classification model, strategy can be divided into



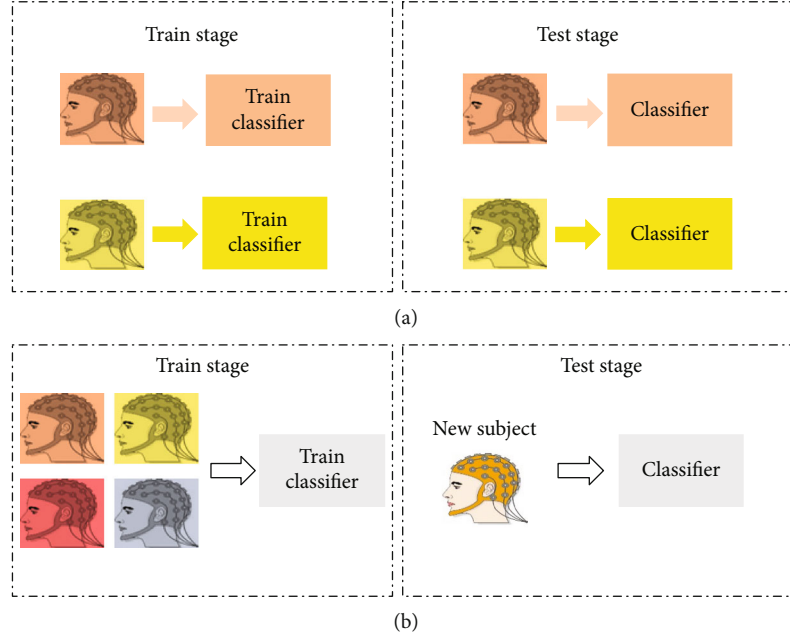


FIGURE 1: A diagram representing the (a) subject-dependent (SD) and (b) subject-independent (SI) training strategy.

two ways: subject-dependent (SD) and subject-independent (SI). As shown in Figure 1, SD strategy is aimed at training a subject-specific model using their own data. In contrast, SI strategy utilizes data from other subjects to train a generalized decoding model for a new subject [6]. One of the main assumptions of ML and DL is that training data and test data belong to the same feature space and subject to the same probability distribution. But it is often violated in the field of EEG signal processing. In other words, the SI strategy cannot satisfy performance of accuracy and generalization due to the individual differences across subject/sessions. SD strategy provides an effective way to optimize this issue; however, it requires long calibration sessions to collect the high-quality and large amounts of EEG datasets. All these restrictions greatly affect the application of BCI in practice.

One effective approach to solve this problem is instance transfer learning (ITL) [7], which combined the advantages of training strategy of SI and SD, i.e., training personalization classification model with enough data. The definition of transfer learning is that given a source domain  $D_s$  and learning task  $T_s$  and a target domain  $D_T$  and learning task  $T_T$  transfer learning are aimed at helping improve the performance of target predictive function  $f_T(\cdot)$  using the knowledge from  $D_s$ . ITL is one of the typical TL methods, which transfer instance knowledge by reweight the data from  $D_s$  to improve generalization ability for  $f_T(\cdot)$ .

The essence of ITL does not change the feature space or property of signals in MI task, but it finds the optimal transfer weighting coefficient for source data by similarity measurement [8–10]. The transfer weighting coefficient is then weighted with the number of corresponding data from  $D_s$ . As shown in Figure 2,  $w_{s,i}^k$  represents the transfer weighting coefficient for  $D_s$  data.  $k$  represents the serial number of the subject, and  $i$  is the  $i$ th trials. During the training stage,

weighted data from  $D_s$  were combined with  $D_T$  data to train classifier. Based on this, we could utilize similar EEG data from other subjects or sessions to help reduce system calibration time and train decoding model for target subject [11]. For example, Azab et al. proposed a weighted transfer learning for improving MI task that they use Kullback–Leibler divergence to measure the similarity between two feature spaces of signal. According to the results of similarity, the weight coefficient is assigned to the source data to optimize the small sample problem in classification model training [12]. A Jensen-Shannon ratio method is used to measure similarity between target data with source data in Giles et al.’s work [13]. Based on this method, they propose a target subject identification framework based on rule adaptation transfer learning, which can reduce the calibration time of the online BCI system by reusing the data with the highest similarity between  $D_s$  and  $D_T$ .

However, due to the obvious individual differences across subjects, the direct instance transfer method may bring negative transfer effects. In addition, traditional measurement does not concentrate on the specific feature of EEG data, which will affect the performance of the transfer learning. Especially for motor imagery signals, the traditional time-series signals cannot effectively reflect the feature of motor intention, but the energy feature of signal can represent the distribution of feature well. Therefore, choose the translocatable objects and assign transfer weights reasonably to the core research for instance transfer learning [14]. In the computer vision field, content-based image retrieval (CBIR) is an important research topic [15]. The goal of CBIR is to find images from the source domain that belongs to the same category. MI spectrogram image contains abundant information of frequency and energy feature, which is suitable for extracting feature of motor intention. Therefore, we

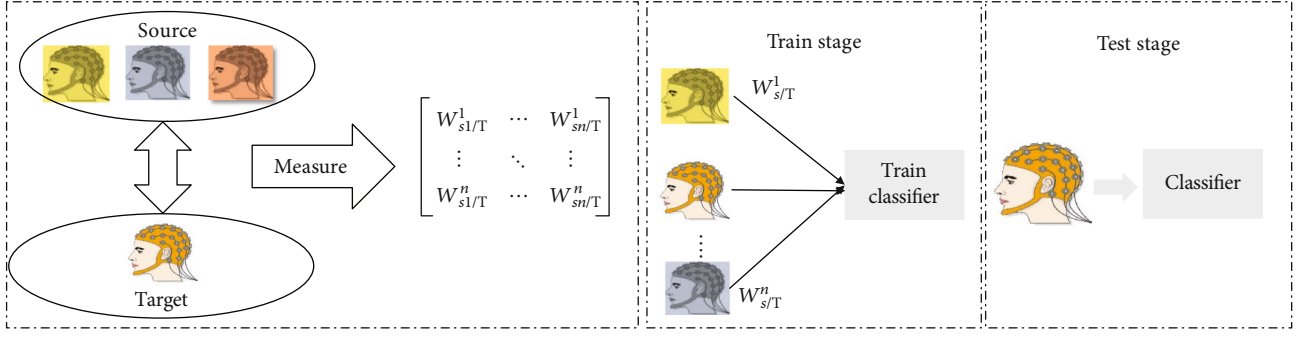


FIGURE 2: A diagram representing the instance transfer subject-dependent (ITSD) training strategy.

assume that technology of CBIR may implement effective data matching across subject and achieve the effective instance transfer from  $D_s$  to  $D_T$ . The perceptive Hash (p-Hash) algorithm is one of the typical CBIR methods, which is used to judge the similarity between different images by transforming these images to perceptual hash code and measure its distance [16].

With the development of deep neural network (DNN) technology, EEG decoding based on DNN has attracted wide attention. Due to the excellent ability of fitting and automatic feature extraction, DNNs achieve outperformed results for EEG classification. In Reference [17], a convolutional neural network (CNN) and variational autoencoder (VAE) were used for two-class MI classification task. The CNN utilized multiple hidden layers to extract the features, and the VAE was used for feature classification. The CNN-VAE method achieved a 3% improvement in classification accuracy than the best methods in their referred literature. Lu et al. [18] proposed a novel method based on restricted Boltzmann machines (RBMS) for EEG classification. Fast Fourier transform (FFT) and wavelet packet decomposition (WPD) were used to extract the frequency-domain features of signals, which were used as inputs of the network. Three RBMs were stacked with an additional output layer to train the classification network. The authors verified that the classification performance of this network was better than state-of-art methods evaluated by public datasets ( $p < 0.01$ ). In a recent study [19], the researchers compared the classification performance of a CNN and long short-term memory (LSTM) network for classifying the time-frequency domain signals of MI. The authors evaluated the adaptability between different network structures to individual differences and showed that CNN provided better performance for detecting the differences across subjects, and its classification rate was significantly higher than that of the LSTM. In summary, CNN shows satisfactory classification performance in MI-BCI task compared to traditional machine learning methods or other networks. However, the limitation of the amount of dataset hinders the practical applications of DNN. Especially for SD training strategy, it is difficult for a subject to collect enough high-quality EEG data. Therefore, we propose a novel instance transfer learning based on p-Hash to improve the utilization efficiency of data and build a CNN for MI classification.

Based on the problems mentioned above, we propose a novel instance transfer learning strategy combined with CNN for subject-dependent MI classification. The main contributions of this paper are as follows:

- (1) To address the issue of individual differences across subjects/sessions in MI classification, we creatively apply the methodology of content-based image retrieval to EEG classification. Based on this, we proposed a novel instance transfer learning (ITL) strategy using the p-Hash algorithm, which is aimed at calculating the transfer weight coefficient between the trails from different subjects/sessions
- (2) There are two main limitations of subject-dependent and subject-independent training strategies: small-scale dataset and large difference of signal across subjects. Therefore, we apply instance transfer learning to optimize the traditional training strategies. Similarity measurement in feature space is executed to calculate the transfer weight coefficient across subjects/sessions, which is aimed at exploring the correlation between different trials. And then we expand the training set for target subject based on instance transfer by weighted
- (3) To improve the classification performance in MI-BCI task, we combine CNN with transfer learning strategy using SD training strategy (ITSD-CNN) to classify MI signal. Experiments evaluate that the ITSD-CNN can achieve outperformed results than state-of-art methods

The step of ITSD-CNN can divide into these steps: firstly, we preprocess the raw MI-EEG signals and adopt short-time Fourier transform (STFT) to transform the raw MI signal into a 2-D spectrogram signal. Then, an ITL based on the perceptive Hash algorithm is proposed to measure the similarity of MI signals between  $D_s$  and  $D_T$ . Next, we build a convolutional neural network to classify the MI data after transfer learning. The BCI competition IV-2b dataset is used to verify the effectiveness of this framework. Our results show that the proposed approach can significantly improve the classification performance. Meanwhile, the ITSD provides a new training strategy to optimize the performance of SD

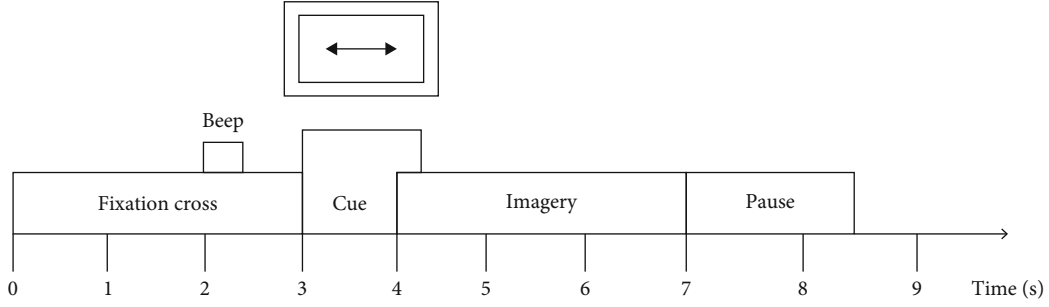


FIGURE 3: Diagram of a trial and timings.

training. The rest of the sections is organized as follows. Section 2 explains the materials and methods for ITSD and CNN. Section 3 introduces the experimental results and discussion. Discussion is described in Section 4, and Section 5 is the conclusion of the paper.

## 2. Materials and Methods

**2.1. Description of Datasets.** In this paper, we utilize BCI competition IV dataset 2b [20]. This dataset was provided by the BCI Research Institute in Berlin and contained two parts: the standard set and the evaluation set. Nine subjects participated this experiment, and three channels (C3, C4, and CZ) were used to record EEG with a 250 Hz sampling rate. Each subject is required to imagine the movement of left and right hands according to the cue. And all of them underwent 5-session experiments. The experimental process is shown in Figure 3.

**2.2. Preprocessing of EEG.** MI signals are extremely weak and accompanied with irrelevant component. And the feature of MI tends to appear in specific frequency band: mu band (8-14 Hz) and beta band (14-30 Hz). To reduce the effect of the artifact of signals, we uniformly filter the signal to 8-30 Hz through the Butterworth filter with 6 orders.

The potential activity of MI always causes the variation of energy in the contralateral cortex and ipsilateral cortex during MI, which is recorded by C3, Cz, C4, and surrounding channels [3]. However, this phenomenon cannot reflect in time domain clearly. To describe features in a better form, we transform time-series signals to spectrogram signals after filtering. As shown in Figure 4, the three channels are converted into a two-dimensional form and are mosaicked into an image using vertical stacking. The variations along X-axis and Y-axis represent the trend of time series and frequency, respectively. And the depth of color indicates the energy feature. For one trial, we chose the data from 3 to 7s (period of imagery) and set the window size of the STFT as 256. After transformation, all spectrogram images were resized to  $64 \times 64$  for convenience and consistency in the subsequent calculation.

**2.3. Instance Transfer Learning Based on the Perceptive Hash Algorithm.** The spectrogram signal of MI can vividly reflect the feature variations especially for the energy of frequency band. The perceptive hash (p-Hash) algorithm can obtain

the most sensitive information by discrete cosine transform (DCT) in the human and machine vision system [16]. This transformation concentrates the energy on the main diagonal of the image matrix and has effectively removed redundant and irrelevant components. Under specific EEG task, the feature distribution of signals across subjects may exist difference but the form of feature is consistent. Therefore, we assume that change between different modes for MI can be effectively perceived and recognized by p-Hash.

This paper uses p-Hash to measure the similarity of spectrogram data across subject. Then, the obtained similarity is transformed into the ITL coefficient that inputs into a classifier combined with corresponding data. The implementation of ITL based on p-Hash are as follows.

Firstly, some denotation of symbols should be explained. We define  $D_T$  representing the target subject's data and  $D_S$  is other subjects' data. Let us define  $G_n^i = \{g_t^i\}_{t=1}^n \in D_S^{l \times l}$  which is a set of single-trial EEGs represented by spectrogram from  $D_S$ ,  $Q_n^i = \{q_t^i\}_{t=1}^n \in D_T^{l \times l}$  for target subject, where  $t$  is the number of EEG trials,  $l$  is the dimensions of a square matrix,  $i$  represents the  $i$ -th subject.  $\tilde{q}_n^i(1/n) \sum_{t=1}^n q_t^i \tilde{q}_t^i$  represent the average spectrogram for the current target subject.

Before calculation, spectrogram images from  $D_S$  and  $D_T$  are separately resized to  $64 \times 64$  and converted to grayscale level. Then, discrete cosine transform (DCT) is utilized to compress an image:

$$\begin{aligned}
 G_{u,u} &= \frac{1}{4} \alpha(u)^2 \sum_{x=0}^{n-1} \sum_{x=0}^{n-1} g_{x,x} \cos \left[ \frac{(2x+1)u\pi}{2N} \right]^2, \\
 Q_{v,v} &= \frac{1}{4} \alpha(v)^2 \sum_{y=0}^{n-1} \sum_{y=0}^{n-1} \tilde{q}_{s_{iy},y} \cos \left[ \frac{(2y+1)v\pi}{2N} \right]^2, \\
 \alpha(u) &= \alpha(v) = \sqrt{\frac{1}{N}}, \quad u, v = 0, \\
 \alpha(u) &= \alpha(v) = \sqrt{\frac{2}{N}}, \quad u, v \neq 0,
 \end{aligned} \tag{1}$$

$\alpha(u)$  and  $\alpha(v)$  are coefficient matrixes after transformation.  $G_{u,u}$  and  $Q_{v,v}$  are results after transformation. The energy variations of the image after DCT are mainly concentrated in the low-frequency part [21]. Therefore, the  $8 \times 8$  matrix  $d$  located in left diagonal is extracted for subsequent

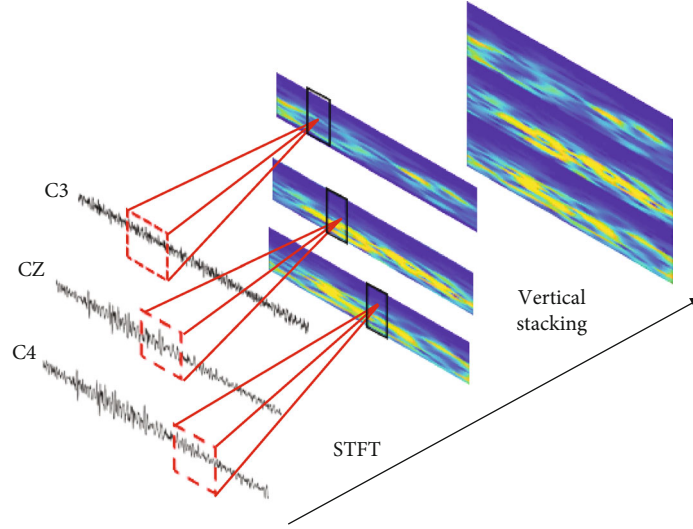


FIGURE 4: Spectrogram images with 3 electrodes after STFT.

calculations. Next, the mean value of DCT coefficients is calculated, respectively:

$$\begin{aligned} m_s &= \frac{1}{n \times n} \sum_{u=1}^n \sum_{u=1}^n d_{u,u}, \\ m_t &= \frac{1}{n \times n} \sum_{v=1}^n \sum_{v=1}^n d_{v,v}. \end{aligned} \quad (2)$$

In addition, the mean value of DCT coefficients is set as threshold standard to compare with each coefficient. By the rule of threshold, the two-dimensional matrix of  $n \times n$  can be compressed into one dimension of  $1 \times n$  matrix  $H_i$ .

$$\begin{aligned} h_i &= 0, & b_i < m, \\ h_i &= 1, & b_i > m, \end{aligned} \quad (3)$$

where  $h_i$  is the bit of the perceptual hash at position  $i$ ,  $m$  is the mean value of the DCT coefficients, and  $b_i (i = 0, 1, \dots, N-1)$  is DCT coefficient of the array. The obtained  $1 \times n$  matrix is  $H_i$  which represents perceptual hash code [22].

Finally, respectively, calculate the Hamming distance  $d_H$  of perceptual hash code from  $D_T$  and  $D_S$  and set the distance  $d_H(H_T, H_S)$  as the ITL weight coefficient from source domain to target domain. For each trial from source subject, weight  $w_{S_i/T_i}$  can be calculated:

$$w_{S_i/T_i} = d_H(H_T, H_S) = \sum_{i=1}^L [H_T(i) - H_S(i)]^2. \quad (4)$$

The calculation processing of transfer weight is shown in Figure 5.

**2.4. Convolutional Neural Network.** Researches show that the CNN has obvious advantages in processing MI signals [23]. CNN is a multilayer neural network composed of a sequence

of convolution, pooling, and fully connected layers. The convolution layer extracts different levels of feature of input image by kernel size, while the pooling layer reduces the complexity of the model by subsampling. With the increase of layers, the more advanced features can be extracted. The fully connected layer will transform the output matrix from the last layer to a  $n$ -dimensional vector ( $n$  is the number of classes). Backpropagation is utilized to decrease the classification error.

In the convolution layer, the input image can be convolved with a spatial filter to form the feature map and output function, which is expressed as

$$X_j^l = f \left( \sum_{i \in M_j} X_i^{l-1} \times w_{ij}^l + b_j^l \right). \quad (5)$$

This formula describes the  $j$ th feature map in layer  $l$ .  $X_j^l$  is calculated by the previous feature map  $X_i^{l-1}$  multiplied by the convolution kernel  $w_{ij}^l$  and bias parameter  $b_j^l$ . Finally, the mapping is completed by RELU function  $f()$ .

$$f(a) = \text{RELU}(a) = \ln(1 + e^a). \quad (6)$$

The pooling layer is sandwiched in the continuous convolution layer to compress the amount of data and parameters and reduce overfitting. The max pooling method in this work is chosen as follows:

$$X_{j,k}^l = \max_{0 \leq m, n \leq s} (X_{j+s+m, k+s+n}^{l-1}), \quad (7)$$

where  $j$  and  $k$  are the locations of the current feature map  $X_j^l$  and  $s$  stands for pooling size. The double fully connected layer structure can effectively translate the multiscale features of the image. Considering the multiple influencing factors of time, frequency, and channel, this paper uses double full-

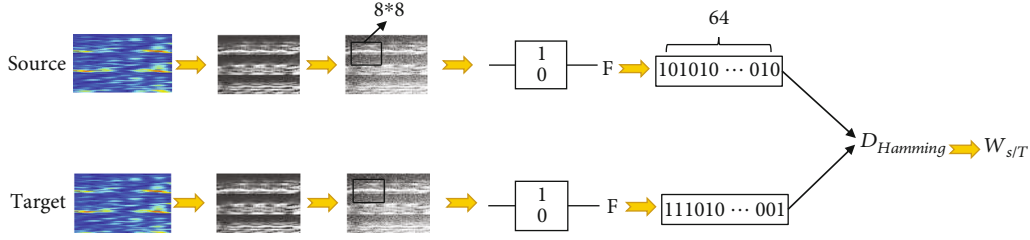


FIGURE 5: Transfer weight calculation using perceptive hash algorithm.

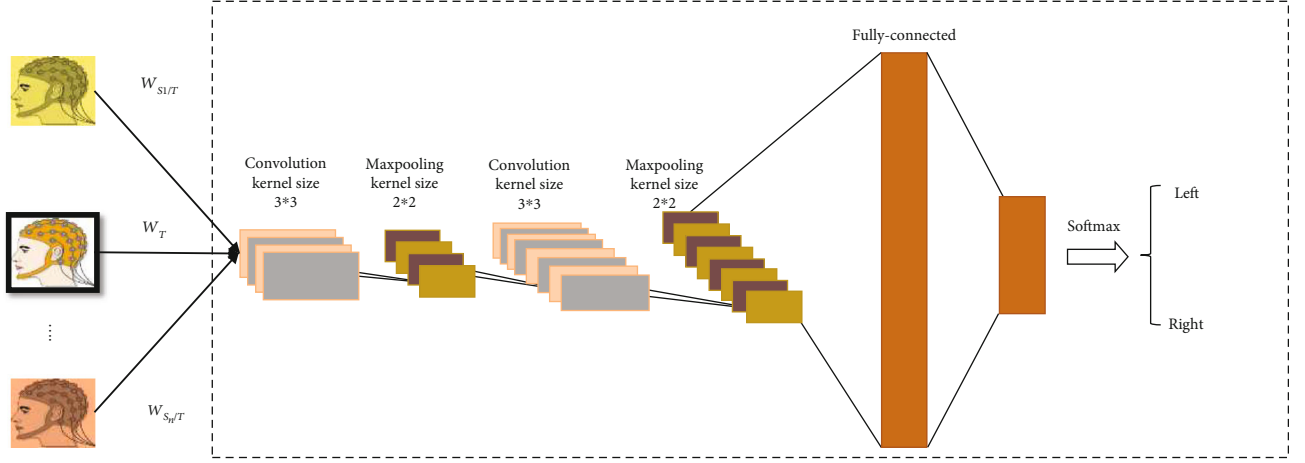


FIGURE 6: The structure of convolutional neural network for classification.

TABLE 1: Detailed architecture for the CNN.

Layers	Type	Size	Stride	Output dimension	Activation	Mode
Input	1			(64,64,3)		
Convolution	2	$3 \times 3$		(64,64,8)		
Max pooling	3	$2 \times 2$	(1, 1)	(32,32,8)		
Convolution	4	$3 \times 3$		(32,32,8)	RELU	Valid
Max pooling	5	$2 \times 2$		(16,16,8)		
Dense	6			(10, 1)		
Dense	7			(2, 1)	Softmax	

connection layers to improve the performance gain of the softmax layer. Two-way softmax in the last layer in the deep networks is used to predict the distribution of two motor imagery tasks.

$$y_i = \frac{\exp(\sum x_i \cdot w_{i,j} + b_j)}{\sum \exp(\sum x_i \cdot w_{i,j} + b_j)}. \quad (8)$$

$x_i$  is the  $i$ th feature map and  $y_i$  represents an output probability distribution. The gradient of back-propagation was calculated according to the cross-entropy loss function.

$$\text{Loss} = -[y \log \tilde{y} + (1 - y) \log (1 - \tilde{y})]. \quad (9)$$

And we used the stochastic gradient descent (SGD) optimizer with a learning rate of  $1e-4$  to improve the speed of network training.

$$\begin{aligned} W^k &= W^k - \mu \frac{\text{tial} E}{\text{tial} W^k}, \\ b_k &= b_k - \mu \frac{\text{tial} E}{\text{tial} b_k}, \end{aligned} \quad (10)$$

where  $\mu$  is the learning rate, while  $W^k$  represents the weight matrix for kernel  $k$  and  $b_k$  represents the bias value.  $E$  represents the difference between desired output and real output. There are eight layers in the proposed network (Figure 6).

The first layer is the input layer, and the second layer is a convolutional layer with kernel size  $3 \times 3$ ; the next layer is the max pooling layer with kernel size  $2 \times 2$ . The next two layers have the same kernel size and function. Two fully connected layers, respectively, consist of 10 and 2 neurons to compute the predicted labels. The gradient of backpropagation is calculated according to the cross-entropy loss function. The stochastic gradient descent with momentum (SGDM) optimizer is used for optimization with learn rate  $= 1e-4$ , momentum  $= 0.9$ , and decay  $= 1e-6$ . We set the minibatch size to 50 and the max epoch to 6. To reduce computation time and prevent overfitting, we adopt to the dropout operation. The proposed CNN model is summarized in Table 1.



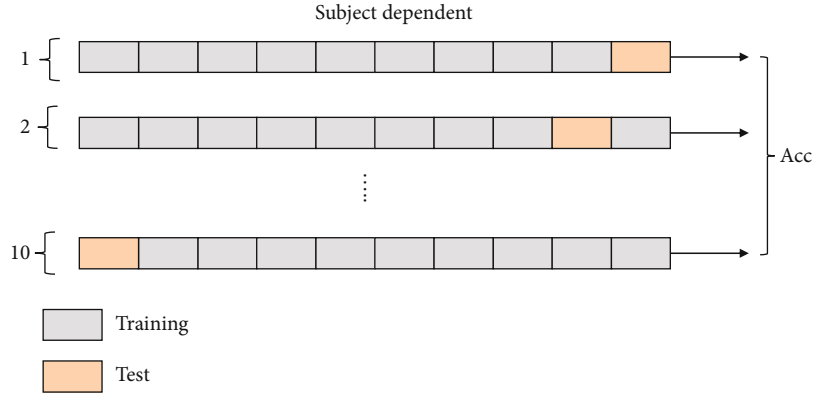


FIGURE 7: Subject-dependent training strategy.

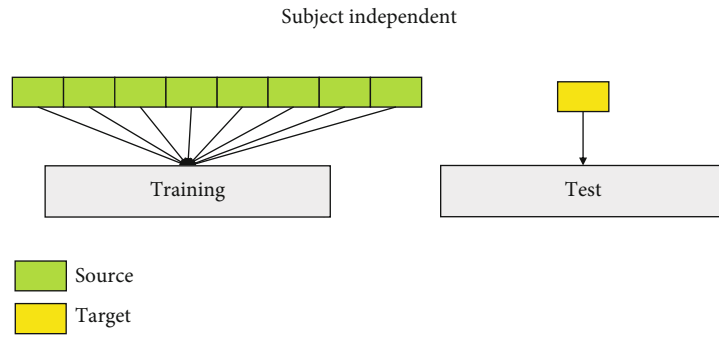


FIGURE 8: Subject-independent training strategy.

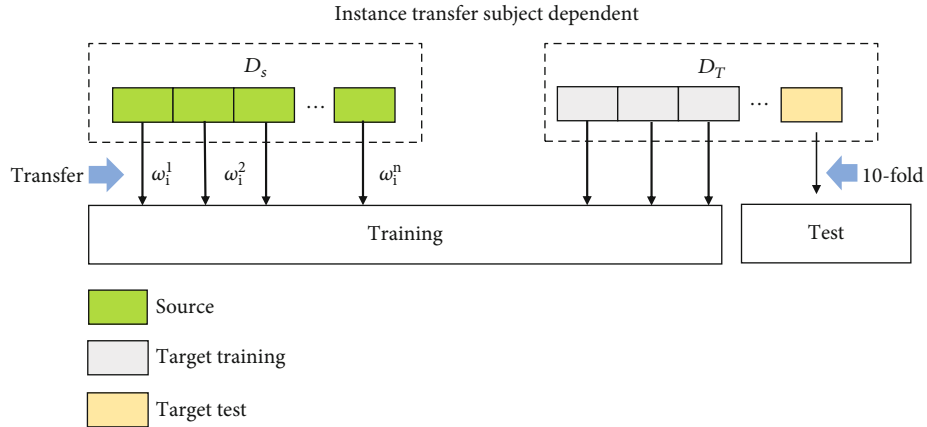


FIGURE 9: Instance transfer subject-dependent training strategy.

**2.5. Evaluation for Classification Performance.** In our study, the average classification accuracy and mean kappa value are utilized to test the performance of the proposed framework. The kappa value is a typical method to evaluate the EEG classification performance which can remove the effect of result of random classification. It can be calculated as follows:

$$\text{kappa} = \frac{\text{acc} - \text{rand}}{1 - \text{rand}} \quad (11)$$

To evaluate the effectiveness of instance transfer learning, three training strategies are compared. As for the subject

dependent method (Figure 7), a total of 720 trials for one subject are divided into the training data and test data using 10-fold cross-validation.

A generalized model is trained using data from other subjects ( $D_s$ ) in the subject-independent training stage (Figure 8).

In the ITSD method, weighted data from  $D_s$  together with target data are input into the training set. And data from  $D_T$  are used to test model performance; the method of data partition is shown in Figure 9.

To show the size of training and test data more clearly, we briefly summarize the number of data for three training strategy in Table 2.

TABLE 2: Size of dataset for three training strategy.

	SI-CNN	SD-CNN	ITSD-CNN
Training data	5760	648	648+transfer instance
Test data	720	72	72

### 3. Experimental Results and Analysis

**3.1. Performance of the Proposed Framework.** In this paper, we use BCI competition IV dataset 2b to verify the proposed methods. During the training stage for each subject in ITSD, we supply dataset from other subjects by instance transfer (Figure 9). After the mixture of source data and target data, we adjust the number of instances to keep the class balance in the training stage. To evaluate the performance of different training strategies, we compared the classification accuracy of different methods.

As depicted in Table 3, the SD training strategy is better than SI based on the CNN classifier even though SI obtains more training data. This indicates that MI-EEG from different subjects causes an obvious difference of feature under the same label. The average classification accuracy of ITSD-CNN is superior to that of SD-CNN, which obtains a 14.1% improvement. It is worth noting that subjects 2 and 3 can better adapt to model preference by efficiently data transfer to greatly improve the classification accuracy.

To verify the significance of results, analysis of variance was performed. As shown in Figure 10, there is no significant difference between SD-CNN and SI-CNN, while the strategy of ITSD-CNN performs satisfied convergence and high accuracy than the other two methods ( $p < 0.01$ ).

By observing the training process, the weakness of small sample can directly influence the results of network training. Effective data transfer can increase the number of samples to improve the generalization of network and prevent underfitting. Moreover, this method can validly reduce the influence of classifier result from individual differences.

**3.2. Comparisons with Previous Research.** Numbers methods have been proposed for MI classification using BCI competition IV dataset 2b. In this section, we further compared our method with that of a commonly used strategy by the metric of mean kappa value. Based on the analysis of Table 4 and Figure 11, we can observe that ITSD-CNN outperforms the baseline and the state-of-the-art methods. It is worth noting that the hybrid framework based on CNN obtains an ideal result among these methods. This indicates that CNN has strong robustness and high accuracy in MI classification. In addition, instance transfer effectively improves the classification performance of CNN using the same model and parameters.

### 4. Discussion

Compared with traditional methods, the application of deep learning for EEG classification has successfully improved the performance [28]. However, there are still some limitation hinder its application in practice. The feature distribution

TABLE 3: Classification accuracy of different training strategies.

Subjects	Accuracy % (mean $\pm$ std.dev.)		
	SI-CNN	SD-CNN	ITSD-CNN
1	82.3 $\pm$ 2.3	80.0 $\pm$ 2.9	93.2 $\pm$ 2.1
2	79.5 $\pm$ 5.1	76.7 $\pm$ 3.1	96.7 $\pm$ 3.2
3	63.8 $\pm$ 5.8	66.7 $\pm$ 3.2	94.1 $\pm$ 6.1
4	76.5 $\pm$ 3.2	83.3 $\pm$ 3.2	97.2 $\pm$ 1.0
5	79.8 $\pm$ 3.2	86.7 $\pm$ 2.9	92.7 $\pm$ 2.7
6	76.2 $\pm$ 4.7	79.0 $\pm$ 3.8	95.6 $\pm$ 2.5
7	77.6 $\pm$ 3.5	83.3 $\pm$ 2.1	94.9 $\pm$ 3.1
8	78.9 $\pm$ 2.8	83.3 $\pm$ 4.5	96.1 $\pm$ 0.8
9	81.3 $\pm$ 3.7	86.7 $\pm$ 3.5	92.2 $\pm$ 2.0
Ave	77.3 $\pm$ 3.8	80.6 $\pm$ 3.2	94.7 $\pm$ 2.6

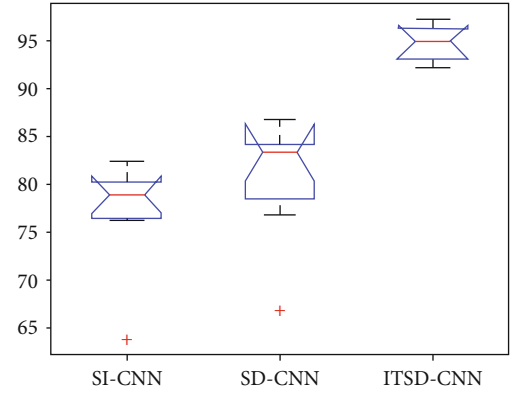


FIGURE 10: The ANOVA stats of classification accuracy for the compared model.

TABLE 4: The review of classifiers performance for BCI competition IV dataset 2b.

Method	Researcher	Mean kappa value
FBCSP	Ang et al. [24]	0.502
Twin SVM	Soman and Jayadeva [25]	0.526
CNN-SAE	Tabar and Halici [23]	0.547
CSCNN	Rong et al. [26]	0.663
NCA + DTCWT	Malan and Sharma [27]	0.615
CNN-VAE	Dai et al. [17]	0.564
ITSD-CNN	Our method	0.664

of EEGs always shows a difference in the same mental task across subject/session, which may cause the overfitting during the network training. Transfer learning turns out to be instrumental in subject/session classification performance. It can be used to initialize a BCI using knowledge transfer from other subjects for a naive subject. At the same time, this strategy may help a classifier to learn global features from all subjects without falling into the local optimal. Therefore, transfer learning combines the advantages of SI and SD strategy and outperforms them. In future studies, it is valuable to

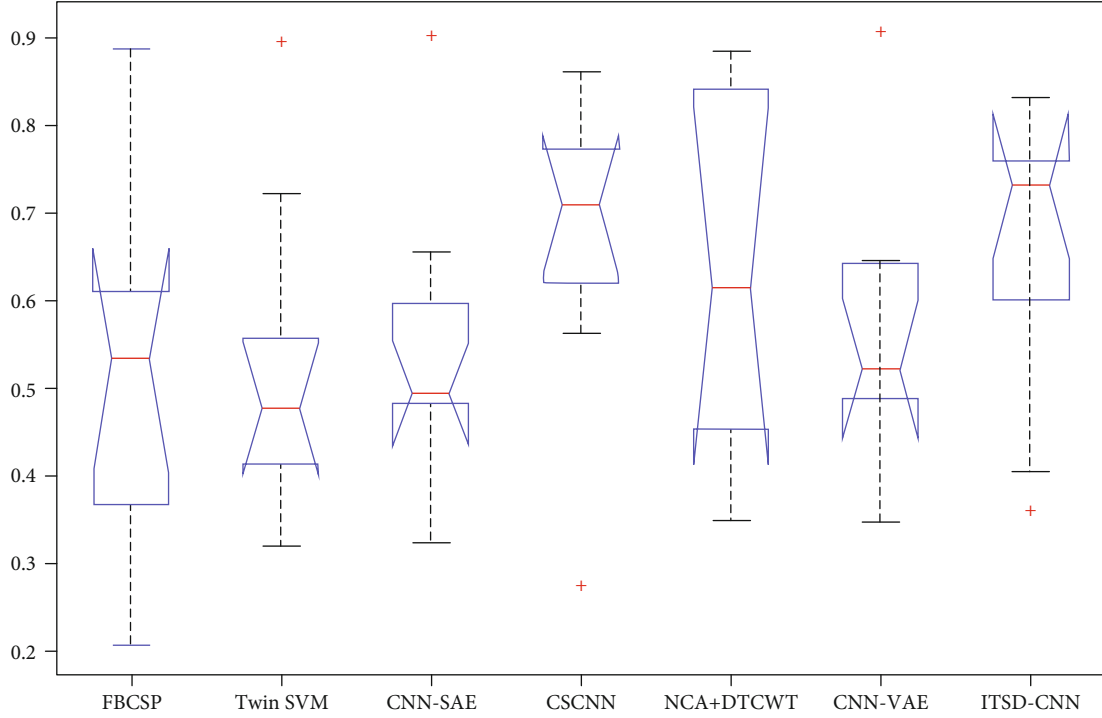


FIGURE 11: The ANOVA stats of the mean kappa for existing methods.

design a decoding model for EEGs based on transfer learning combined with the deep neural network.

Another limitation is small-scale sample for classifier training. Strict requirements of quality and collection of EEG data make it difficult to obtain large datasets in practice. The performance of EEG decoding based on DNN is directly related to the amount of training data. Data augmentation is a promising way to address this issue. As discussed in References [29, 30], artificially generated data can be used to training classification model and the augmentation method has been proved efficient in EEG decoding. The addition of generated dataset improves the complexity and robustness of models. Traditional augmentation methods contain geometric transformation and model generation, which requires a long time to prepare and select suitable generated data. It takes up a lot of computing resources in the BCI system. Therefore, data augmentation from an available database may provide a probable method. As proposed in this study, instance transfer learning can easily obtain data from other subjects and adaptively assign weights to the transfer data, which achieve the utilization maximization of data across subjects. Although the training process of this method is similar with subject-dependent strategy, i.e., it requires recomputing for a new participant; low-cost calculation would not burden the operation of the BCI system. In later research, we will explore the detail of variability across subjects and achieve more effective transfer learning.

## 5. Conclusion

In this paper, we propose a novel instance transfer learning method with a deep neural network applying for the subject-dependent classification of motor imagery in the

BCI system. In this work, we firstly transform the raw data to spectrogram image by STFT. Then, instance transfer learning based on the perceptive Hash algorithm is utilized to measure the similarity between the data of source domain and target domain. Next, we convert the similarity into a transfer weight coefficient to realize the data transfer of a single trial between different subjects. Finally, a convolutional neural network is built to verify the performance of proposed methods and some other methods are adopted to evaluate the results. Experiment verifies that instance transfer learning by the perceptive Hash algorithm can effectively provide data augmentation based on subject-dependent training strategy and improve the performance of the classifier, which demonstrates the superior performance and promising potential of proposed novel training strategy. Meanwhile, the proposed method provides a solution for the weakness of small samples in the deep neural network.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the Development Plan of Shaanxi Province (Grant No. 2018ZDCXL-GY-06-01) and National Key Research and Development Plan of China (Grant No. 2017YFC1308500) and special guidance funds

for the construction of world-class universities (disciplines) and characteristic development in Central Universities.

## References

- [1] G. Bonassi, M. Biggio, A. Bisio, P. Ruggeri, M. Bove, and L. Avanzino, "Provision of somatosensory inputs during motor imagery enhances learning-induced plasticity in human motor cortex," *Scientific Reports*, vol. 7, no. 1, p. 9300, 2017.
- [2] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 6, pp. 642–651, 1997.
- [3] M. Phothisonothai and M. Nakagawa, "EEG-based classification of motor imagery tasks using fractal dimension and neural network for brain-computer interface," *IEICE Transactions on Information and Systems*, vol. 91, no. 1, pp. 44–53, 2008.
- [4] K. Zhang, X. Chen, F. Liu, H. Tang, J. Wang, and W. Wen, "System framework of robotics in upper limb rehabilitation on poststroke motor recovery," *Behavioural Neurology*, vol. 2018, 14 pages, 2018.
- [5] F. Lotte, L. Bougrain, A. Cichocki et al., "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.
- [6] K. Vo, T. Pham, D. N. Nguyen, H. H. Kha, and E. Dutkiewicz, "Subject-independent ERP-based brain-computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 719–728, 2018.
- [7] M. N. A. Khan and D. R. Heisterkamp, "Adapting instance weights for unsupervised domain adaptation using quadratic mutual information and subspace learning," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 1560–1565, Cancun, Mexico, December 2016.
- [8] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, p. 114, New York, NY, USA, July 2004.
- [9] C. Cortes, M. Mohri, M. Riley, and A. Rostamizadeh, "Sample selection bias correction theory," in *Algorithmic Learning Theory. ALT 2008. Lecture Notes in Computer Science*, vol. 5254, Y. Freund, L. Györfi, G. Turán, and T. Zeugmann, Eds., pp. 38–53, Springer, Berlin, Heidelberg, 2008.
- [10] B. Tan, Y. Song, E. Zhong, and Q. Yang, "Transitive transfer learning," in *KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1155–1164, New York, NY, USA, August 2015.
- [11] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain-computer interfaces," *IEEE Transactions on BioMedical Engineering*, vol. 57, no. 6, pp. 1318–1324, 2010.
- [12] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.
- [13] J. Giles, K. K. Ang, L. S. Mihaylova, and M. Arvaneh, "A subject-to-subject transfer learning framework based on Jensen-Shannon divergence for improving brain-computer interface," in *CASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3087–3091, Brighton, UK, May 2019.
- [14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [15] J. Wan, D. Wang, S. C. H. Hoi et al., "Deep learning for content-based image retrieval: a comprehensive study," in *MM '14: Proceedings of the 22nd ACM international conference on Multimedia*, pp. 157–166, New York, NY, USA, November 2014.
- [16] V. Monga and B. L. Evans, "Perceptual image hashing via feature points: performance evaluation and tradeoffs," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3452–3465, 2006.
- [17] M. X. Dai, D. Z. Zheng, R. Na, S. Wang, and S. L. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, 2019.
- [18] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 6, pp. 566–576, 2017.
- [19] Z. Wang, L. Cao, Z. Zhang, X. Gong, Y. Sun, and H. Wang, "Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition," *Concurrency and Computation Practice and Experience*, vol. 30, no. 23, 2018.
- [20] *BCI Competition 2008- Graz data sets 2A and 2B* <http://www.bbci.de/competition/iv/>.
- [21] B. Coskun and B. Sankur, "Robust VideoHash extraction," in *2004 12th European Signal Processing Conference*, Vienna, Austria, September 2004.
- [22] M. Fei, Z. Ju, X. Zhen, and J. Li, "Real-time visual tracking based on improved perceptual hashing," *Multimedia Tools and Applications*, vol. 76, no. 3, pp. 4617–4634, 2017.
- [23] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, article 016003, 2017.
- [24] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.
- [25] S. Soman and Jayadeva, "High performance EEG signal classification using classifiability and the twin SVM," *Applied Soft Computing*, vol. 30, pp. 305–318, 2015.
- [26] Y. Rong, X. Wu, and Y. Zhang, "Classification of motor imagery electroencephalography signals using continuous small convolutional neural network," *International Journal of Imaging Systems and Technology*, 2020.
- [27] N. S. Malan and S. Sharma, "Feature selection using regularized neighbourhood component analysis to enhance the classification performance of motor imagery signals," *Computers in Biology and Medicine*, vol. 107, pp. 118–126, 2019.
- [28] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, no. 3, article 031001, 2019.
- [29] J. T. C. Schwabedal, J. C. Snyder, A. Cakmak, S. Nemati, and G. D. Clifford, *Addressing class imbalance in classification problems of noisy signals by using fourier transform surrogates*, arXiv preprint arXiv:1806.07108, 2018.
- [30] Q. Zhang and Y. Liu, *Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks*, arXiv preprint arXiv:1806.07108, 2018.

## Research Article

# Design and Development of Virtual Medical System Interface Based on VR-AR Hybrid Technology

**Xu Cong and Tingting Li** 

*School of Digital Art and Design, Dalian Neusoft University of Information, Dalian, 116023 Liaoning, China*

Correspondence should be addressed to Tingting Li; [lingting\\_ys@neusoft.edu.cn](mailto:lingting_ys@neusoft.edu.cn)

Received 24 April 2020; Revised 26 June 2020; Accepted 10 July 2020; Published 25 August 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Xu Cong and Tingting Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of information technology and digital medicine, computer-assisted virtual medicine has become the development trend of a new generation of clinical surgery, which aims to improve the accuracy of surgery, reduce the risk of surgery, and achieve precise and minimally invasive treatment. The interface design in the computer-aided virtual medical system is a medium for transmitting and exchanging information between humans and machines. This article uses virtual reality technology and augmented reality technology to develop a virtual medical system interface, which aims to solve the interaction problem between users and virtual medical systems and satisfy users. The multidemand psychology is an effective way of interaction. It provides users with a multichannel and comprehensive communication method, which truly meets the design goals that meet the user's psychological needs. It also expands applications for virtual reality technology and augmented reality technology.

## 1. Introduction

With the increase of China's medical level and the rapid development of mobile Internet, the interactive display method of augmented reality has been favored by more people. Many scholars at home and abroad have carried out related exploration and application research on augmented reality technology. Augmented reality technology can use virtual additional information to enhance the user's observation of the real world, so as to fully stimulate the enthusiasm of the visitors from the perspective of the visitors. However, in actual practice, some problems have been exposed. For example, the immersion is not strong. Users hope that they will have close contact with the models in the scene. A single augmented reality technology weakens the user's observation and perception. In order to overcome these problems, this article integrates virtual reality technology on the basis of augmented reality technology and combines the two. This paper proposes a virtual medical system interface design and development based on AR-VR hybrid technology.

The interface design in the virtual medical system is a medium for transmitting and exchanging information between humans and machines and refers to the overall design of software's human-computer interaction, operation

process, and beautiful interface [1]. The friendly interface design not only gives the user a sense of personality and taste but also makes the user's operation process simpler, more comfortable, and free, while fully reflecting the features and functions of the software. In the virtual medical system, users need to implement various operations through its interface, so the interface design is particularly important in virtual medical design [2]. At present, the interface of China's virtual medical system lacks human-system interaction, and the user's multiple needs cannot be met. The interface design of the virtual medical system is relatively single, which is inconvenient for users and increases the user's operating burden [3]. In response to the above problems, in 2005, Pioneer introduced a 3D user interface technology that allowed users to draw text and pictures on a PC without using a keyboard, mouse, or touch screen. This system realizes a completely new feeling atmosphere environment, which is very different from the traditional interface in the past [4]. Enea China Service and R & D Center has launched a three-dimensional stereo operation interface that can be used on smart mobile devices to facilitate users to complete multiple applications more quickly. The interface introduced by Enea mainly includes the main interface, tool interface, movie display interface, and video and audio interface. The picture interface



is a five-category interface. This functional interface for different needs can save users' time to use the application and add more entertainment [5].

This article proposes to apply virtual reality technology and augmented reality technology to the interface design of the virtual medical system, presenting a three-dimensional effect to the user, while enjoying different service methods. The discussion of the interface of the virtual medical system is a human-computer interaction method. A new attempt and innovation has theoretical and practical significance for realizing a human-friendly human-machine interface. At the same time, it also expands a wider development space for virtual reality.

## 2. Virtual Reality Technology and Augmented Reality Technology

**2.1. Virtual Reality Technology.** Virtual Reality (VR for short) was formally proposed by Jaron Lanier, founder of the American VPL company, in the early 1980s. As a collection of high technology, virtual reality technology integrates many key technologies, such as multimedia, computer graphics, sensor technology, artificial intelligence, and simulation technology to achieve three-dimensional spatial expression and a natural human-computer interactive operating environment, which has fundamentally changed the status quo of tedious interactions between humans and computers [6]. Virtual reality technology uses a computer to generate a simulation environment with multiple perceptions such as vision, hearing, and touch (such as aircraft cockpits and operation sites), allowing users to invest in the simulation environment through a variety of sensing devices, thereby generating body immediate feelings and experiences [7].

In the research and development of virtual reality technology, most areas are scrambling for research. Virtual reality technology has the advantages of low cost and strong sense of substitution, so it has broad development space. Using glasses connected to smart devices can support the operation of virtual reality experience halls. In the research of virtual reality technology, if it is applied to more complex environments, it will affect the accurate image display of virtual reality technology. Because of this, virtual reality technology needs further research.

Under the virtual reality technology, the glasses equipment must perform focusing, judgment, and other actions and operate according to the user's head. The direction of sound discrimination cannot be effectively applied in virtual reality technology. VR glasses cannot provide users with tactile sensations. Although they have been experienced visually, they cannot actually touch objects, so they are lacking in tactile experience. Information cannot be input in time in the virtual reality technology. The ordinary information input method does not match the virtual reality technology. In this case, the research of VR glasses is extended to data clothing and data gloves. At present, the application range of virtual reality technology is very wide, which has provided a lot of help for aerospace, medical research, and industrial research.

**2.2. Augmented Reality.** Augmented Reality (AR) is a new technology developed on the basis of virtual reality technology, which accurately superimposes computer-generated virtual objects or other auxiliary information into the real scene, thereby extending the human sensory experience. The visual perception of the real environment enhances the realism of user experience with the effect of combining reality and reality [8].

The industry first contacted by augmented reality technology is video recording devices, which helps video recording devices collect real-world data and then use virtual scene synthesis technology to synthesize and process them through computers. The application of augmented reality technology is different from the virtual reality technology, and the application is more convenient. You can operate the software by downloading the related software of augmented reality. Of course, this operation experience is not ideal. Augmented reality technology is constantly innovating, combining the principles of optical perspective, integrating the real world with virtual information, and enhancing the content of reality display. The application of this technology requires high investment in cost research, and the technical requirements are very strict.

In recent years, as a cutting-edge technology, augmented reality technology has been widely used in machinery manufacturing, virtual assembly, computer games, multimedia design, simulation training, and other fields [9]. Augmented reality technology continues to innovate, combining the principles of optical perspective, fusion display of the real world and virtual information, and enhanced display content of reality; the application of this technology requires higher investment in cost research; and the technical requirements are very strict. In reality, augmented reality technology mainly includes Google Glass. Of course, because augmented reality technology is in the development and research stage, it needs to be further upgraded in many aspects. Augmented reality technology is involved in military, medical, and industrial fields. Augmented reality technology provides users with accurate information and positioning and assists users to achieve research and operation purposes. At the same time, it has been gradually applied in the field of clinical medicine and has become a research hotspot in this field [10].

**2.3. Differences between Virtual Reality Technology and Augmented Reality Technology.** The visual presentation method of virtual reality technology is to block the connection between the human eye and the real world and render a brand new world in real time through the device, and the virtual reality technology presents all fake and fictional. The visual presentation method of augmented reality technology is to superimpose the holographic image and strengthen the visual presentation method when the human eye is connected with the real world. It is a combination of physical and virtual.

**2.4. What Can Virtual Reality and Augmented Reality Do.** Because virtual reality and augmented reality technology can establish a connection between the virtual world and the real world and form a new mode of human-computer interaction, thereby greatly expanding the human sensory experience and recognizing the world, with the gradual

maturity of VR/AR technology, certain applications have been realized, and there is a trend to further form the industry. According to the VR report released by Goldman Sachs, VR and AR not only have the potential to create new markets but will also disrupt some of the current markets. The technology can be applied to 9 major areas: video games, live events, video entertainment, healthcare, real estate, retail, education, engineering, and military.

**2.5. Opportunities and Challenges of Virtual Reality and Augmented Reality.** In the future, VR technology will have greater applications and breakthroughs in the medical field. Specifically, VR technology can be used to generate visual images to check the patient's specific conditions, such as using CAT scans or ultrasound diagnostic images to generate 3D models of patient anatomy. In addition, VR technology can also be used to recover patients. In Europe, immersive virtual display therapy is used in patients with stroke and brain damage in many places to restore their motor and cognitive abilities. Because of the related equipment of VR technology, it still has problems such as inconvenience and poor results, and the processing speed of its hardware is far from meeting the needs of processing large amounts of data in the virtual world in real time. In addition, the price of some equipment is very expensive, and it is difficult to make the equipment universal. Due to the limitations of hardware technology, the development of virtual reality software requires huge expenses, but it cannot achieve the expected application effect. At the same time, there are still many problems in the application of VR technology in high-speed graphics and image processing, artificial intelligence, and other fields, which need to be solved urgently.

At present, AR technology is developing rapidly, and there is much room for development in many fields. AR technology has three main characteristics: integrating real-world and virtual environment information, real-time interactivity, and adding positioning virtual objects in three-digit scale space. However, due to the diversity and complexity of the external environment, AR equipment is often subject to external interference, and there are still great shortcomings in application. Therefore, it is necessary to improve the synthetic tactile input technology of users in the real environment.

### 3. Analysis of the Characteristics of the Interactive Interface Design of the Virtual Medical System

**3.1. Interactive Interface Concepts.** [11]. Interactive interface refers to the medium for communication between people and machines, the medium for exchanging and transmitting information between users and computers, and also a comprehensive operating environment for users to conduct computer operation (i.e. information exchange and communication through this medium)[11]. In the design of interactive interface, it is especially emphasized to be user centered, so it is also called user interface (UI) or human-computer interface.

For individuals, cognition of the real world is actually an electronic signal transmitted from the senses to the brain, and enhanced perception can expand our cognitive world.

Just as you can see farther with a telescope and tiny creatures with a microscope, VR/AR technology can present things we could not previously perceive, simulate sounds we cannot hear, and make things unclear clear. And it can create an enhanced version of the real world. From this perspective, VR/AR technology helps us see the world more clearly, more thoroughly, and richer by enhancing our perception.

The invention is aimed at a problem that is not effective in the prior art and can be achieved through human-computer interaction so that patients can truly understand the causes, pathologies, schemes, procedures of surgery, and matters needing attention after surgery, etc., and provides a virtual reality-based interactive medical system.

**3.2. Basic Elements of Interactive Interface.** The basic elements of interactive interface design are divided into several parts, including text, color, graphics, and sound [12]. These basic elements are responsible for providing users with sufficient visual and auditory feelings, and these sensory stimuli are relevant to users. The information is also used as a source of information and auxiliary tools for users during operation. Without these basic elements, users will not be able to understand the meaning and information conveyed by the machine or the interface itself. The central processing unit includes a message receiving and sending module and a VR video storage module for various medical processes and knowledge and patient information identification module; VR glasses and doctors' terminals also set up information receiving and sending modules to achieve connection with the central processor, the doctor's terminal, and the central processing unit of the VR device connected through the information receiving and sending module; and the doctor's terminal selects the appropriate VR video for the patient according to the patient's situation and sends it to the VR device through the information receiving and sending module. The central processor of the VR device retrieves the corresponding VR video from the VR video storage module according to the information sent to the VR glasses, assisted with the screening for patients to learn from human-computer interaction.

**3.3. Feature Analysis of Interactive Interface of AR Medical System.** The basic characteristics of the traditional two-dimensional interactive interface are the interaction between the mouse and the desktop system and design elements based on the Windows operating system. When designing a two-dimensional interactive interface in the desktop era, you only need to consider the window system and two-dimensional visual design elements in the desktop environment. However, when designing an AR-based interactive interface, because its operating environment is a combination of reality and reality, this also determines the complexity of its interactive interface design. Therefore, in specific applications, attention should be paid to how to enable users to effectively interact with AR-based interaction interfaces in complex environmental relationships. Considering the above factors, when designing an AR-based interactive interface, it should pay attention to its three design characteristics: virtual-real superposition, three-dimensionality, and real-time interaction.

**3.3.1. Virtual and Real Superposition.** Compared with the interactive interface based on augmented reality and the interactive interface based on virtual reality, the interactive environment of the latter is completely virtual. In the former, the user is faced with an information environment in which virtual information and real environment are superimposed. On the other hand, compared with the latter's interactive interface, the former's user experience will be more real and immersive. Among them, the displayed virtual information can be two-dimensional information, such as text and images, and can also be three-dimensional information, such as audio, video, animation, and smell. In the AR-based interactive interface, the real physical environment is superimposed and enhanced by virtual information. Compared with a completely virtual interactive interface or a completely real interactive interface, the AR interactive interface has more room to play in the interactive experience. Sex makes the human-computer interaction experience even more extreme.

**3.3.2. Three-Dimensionality.** The three-dimensional characteristics are mainly reflected in the three-dimensionality of the interaction mode and the interactive environment. In the AR-based interactive interface, the virtual world and real physical world of the two required elements determine the true three-dimensional properties of the AR interactive interface. What the user feels is a three-dimensional world where the virtual world and the real physical world are superimposed. This immersive three-dimensional environment enables users to obtain a completely new user experience. At the same time, the interaction between the user and the computer also has three-dimensional properties. For example, interaction technologies such as gesture control will form an essential element of the three-dimensional interaction system.

**3.3.3. Real-Time Interaction.** In the AR interactive interface, users not only are exposed and experienced an interactive interface that mixes virtual information and real environment but also experience the real-time nature of interactive operations. Similar to computer operation, in the operation of AR interactive interface, users essentially obtain and exchange information with computer hardware through the interactive interface, but the difference is that the interaction of augmented reality is the interaction of virtual information and real environment. In order to give the user a realistic user experience, when the user performs interactive operations, it should interact with the augmented reality system in real time. In simple terms, when a user issues an instruction, the system must implement and feedback in time. The response time during this period should be handled as quickly as possible. Therefore, in the AR interactive interface, its real-time interaction is reflected in the two aspects of human-computer interaction and virtual-real interaction. The real environment seen by users through the device must be immediately and in real time superimposed on the virtual information that matches the user's needs.

**3.4. Analysis of Interactive Interface Features of VR Medical System.** Virtual reality technology is a man-machine interface technology that highly realistically simulates human behav-

iors such as sight, hearing, and touch in the natural environment. American scientist Burdy proposed the "virtual reality technology triangle" theory. The three outstanding features are immersion, interactivity, and imagination.

VR technology is a virtual reality technology, which is currently being applied to human-computer interaction technology. How to effectively apply the technology to the medical system and solve the technical problems of the existing technology, so as to realize the guidance of the medical system through the various procedures of VR technology; to explain the cause of disease, pathology, treatment plan, operation process, postoperative precautions, etc; The existing technologies need to be solved urgently, such as reducing patients' fear of disease, operation and unfamiliar environment in hospital, making operation plan, conducting surgical rehearsal, surgical teaching, surgical skill training, guiding operation during operation and postoperative rehabilitation.

**3.4.1. Immersion.** The ideal virtual environment should be to make the user's vision, hearing, touch, smell, and taste extremely realistic, that is, to allow users to immerse themselves in a part of the virtual environment, so that the user is fully placed in the virtual environment. The performance standard of a virtual reality system is the level of immersion. The key to the good and bad of a virtual environment setting is whether to make users feel a very strong immersion.

VR technology allows the real world to be transformed into a virtual world, and AR technology allows the virtual world to be superimposed on the real world, enabling the virtual world and the real world to interact. There is a convergence trend between the two worlds. In the future, you may not be able to tell which part is virtual and which part is real. The whole world becomes a mixture of virtual and real worlds.

**3.4.2. Interactivity.** Interactivity refers to the degree to which users can manipulate objects and the degree of emotional feedback experience. This interactive mode uses special input and output devices (such as data gloves, data suits and head-mounted displays) to use your own body language to interact naturally with objects in the virtual environment, and the computer can use the user's hands, eyes, head, language and body movements to adjust the virtual image and constantly update the virtual environment to display to the user. For example, users can grab objects directly with their hands in a virtual environment and can feel the weight and at the same time can move with their hands. Therefore, the existing technology needs a system that can realize human-computer interaction and enable patients to experience the real experience to complete the above tasks, in order to save medical resources to the greatest extent and minimize doctor-patient contradictions.

VR/AR technology can reconstruct the information of the physical characteristics of the physical world and build a virtual world as real as the real world. Just as a telephone can convert sound waves into electrical signals and convert them into sound, video phones can reconstruct information from sound and pictures, while "real" virtual worlds reconstruct information from the multidimensional characteristics of objects. If we reconstruct all the information of the physical world and transform it into a virtual space, then your



mobile phone can use the Internet of Things technology to fit the entire world, easily put it in your pocket, and see the world without leaving the house.

**3.4.3. Conceptual.** Imagination, also called creativity, refers to the fact that virtual reality technology has a very wide imaginable space. It not only provides creative ideas for designers to conceive and design virtual environments but also helps humans to obtain more information and knowledge through their own thinking. In short, the three characteristics, immersive, interactive, and imaginative features of the virtual reality system, visually present a virtual space of natural interaction. The virtual environment with these three characteristics allows people to not only feel the immersion of the body but also be spiritually satisfied.

The establishment of multidimensional memories and multidimensional thinking, combined with the assistance of artificial intelligence and the Internet of Things, is likely to trigger a revolution in human learning and break through the physiological limits of the human brain. From word of mouth to text recording to audio and video, each media promotion brings the leaping development of human civilization, and the three-dimensional and interactive learning method produced by VR/AR technology will also promote the progress of human civilization. VR is far more than “vivid” and “lively” in the teaching field, and its significance is more profound in medicine. For example, a beating heart is projected on the screen, and the operator can observe each part at any time and split them. Such teaching not only saves the biological cost of experiments but also improves the accuracy and understanding of students’ operations.

## 4. Design and Development of Interactive Interface of Virtual Medical System

**4.1. Virtual Medical System Development Platform.** This system uses Unity 2018.2.16 as the base platform and combines Vuforia to realize the virtual medical system function development. Vuforia is the industry’s leading and most widely used augmented reality platform. It can easily and quickly create the interaction between the virtual world and the real world. It uses computer vision technology to identify and track flat images and simple 3D objects in real time. Developers can bridge the real-world and digital experiences.

**4.2. Virtual Medical System Interface Design.** The interface of the virtual medical system is designed in two parts. One uses the Vuforia SDK to implement the function of the augmented reality module of the virtual medical system and scans the picture for tracking. When the picture is scanned by the camera, some set 3D objects appear above the picture. The other part uses the Unity3D platform to build virtual scenes to achieve 3D interactive virtual roaming effects.

### 4.3. Virtual Medical System Interface Development

**4.3.1. Augmented Reality Module Development.** There are two stages to process target pictures in the development of the augmented reality module of the virtual medical system. First, the target image needs to be designed and then

uploaded to the Vuforia platform for target processing and evaluation. In the recognition process, Vuforia compares the natural feature points of the input image with its own feature point database to determine the recognition process. In the Vuforia image recognition workflow, you need to use Vuforia Target Manager to generate a feature database of recognition maps.

**4.3.2. Virtual Reality Module Development.** The virtual reality module in the virtual medical system is developed based on the Unity3D platform. First of all, a virtual medical model needs to be built. The production of this virtual medical model can be done with 3d Max software, and the texture of the model can be designed with Photoshop. This is a very heavy workload. It takes a long time to make the model. Use 3d Max software to make a virtual medical scene, and then put the completed model into Unity3D to synthesize according to the actual proportion. Then, according to the predesigned function, the corresponding code is added to implement the roaming module function of the virtual medical scene.

**4.4. Virtual Medical System Interface Test.** The system compilation environment is Visual Studio 2017. The code is written in C# and is based on the Unity3D platform combined with the Vuforia SDK. After starting the system, enter the system login interface, click the instructions to view the system usage method, and enter the account and password to log in to the system, as shown in Figure 1. After entering the system, enter the scene selection mode. There are 2 scenes to choose from, as shown in Figures 2–4. After entering the scene, you can control the third-person position movement in the scene and change the perspective of roaming by touching with your finger, so as to realize individualized autonomous learning of medical knowledge.

**4.5. VR and AR Technology Utilization in Medical Industry.** At present, as AR and VR technologies are emerging technologies, their application in the medical field is still in an experimental stage. However, once these technologies are applied, they have gained great recognition from the medical community and have provided great impetus to the development of medical treatment. It can be seen from Figure 5 that the utilization rate of VR and AR technology in the medical industry is also increasing. VR technology has increased from 0.05% in 2016 to 0.11 in 2018, more than doubled; AR technology has increased from 0.04% in 2016 to 0.1% in 2018, more than doubled.

**4.6. Praise Rate of VR and AR Technologies.** Due to the rapid development of China’s technology in the past two years, the technical levels of AR and VR have also greatly improved, and people’s recognition of these two technologies has gradually increased. It can be seen from Figure 6 that AR technology has increased from 55% in 2015 to 79% in 2018; and because AR technology will be affected by environmental factors, the praise rate has grown slowly, from 50% in 2016 to 65% in 2018.

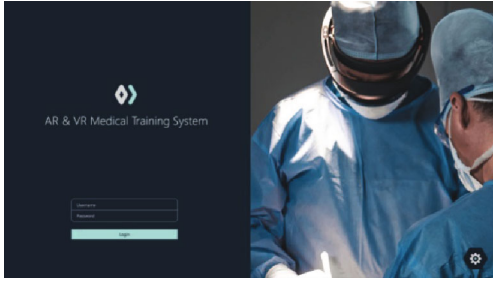


FIGURE 1: System login screen.

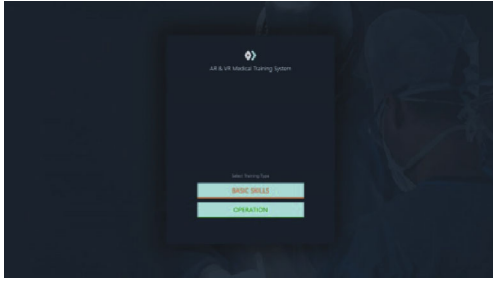


FIGURE 2: System function selection interface.

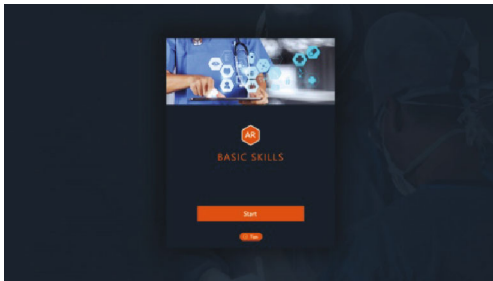


FIGURE 3: AR system module interface.

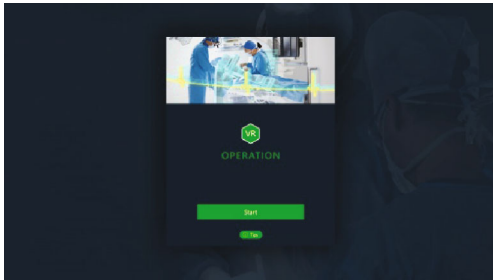


FIGURE 4: VR system module interface.

## 5. Conclusions and Outlook

In this paper, by studying the existing design theories and combining the related technology implementation backgrounds of augmented reality and virtual reality technology, this paper discusses and summarizes the theory and design schemes of interactive interface design of augmented and virtual reality for virtual medical systems. Technology and augmented reality technology are applied to home media

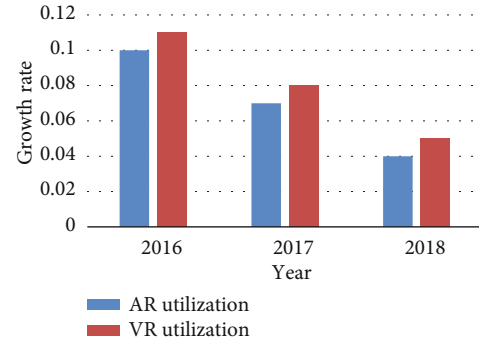


FIGURE 5: Utilization of VR and AR technologies in the medical community.

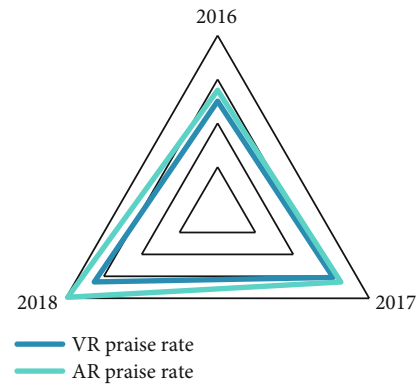


FIGURE 6: Praise rates for VR and AR technologies.

products, and the user's sense of immersion and experience is improved by constructing virtual scenes, and finally the virtual medical system interactive interface is developed.

At present, the AR/VR industry is still in its infancy, and there are still problems such as insufficient applications, insufficient technical reserves, and insufficient database construction. For example, AR/VR products still have problems in terms of convenience and application popularity, and developers need to make great efforts to find the combination of technology and applications. For AR/VR technology, what ordinary users expect is not a cool technical experience, but a real improvement and improvement in the quality of life and learning level. The current augmented reality technology is still in the development stage, and we have no way to completely rely on augmented reality applications and devices in our lives. For example, VR devices are prone to vertigo. To make matters worse, reliable research has confirmed that the dizziness caused by VR devices may still exist after you stop using the device for 8 hours. Until now, in addition to vision and hearing, virtual reality technology has not been able to simulate the other three senses (olfactory, touch, and taste). Unlike the hardware market that contends, software and content are the shortcomings of AR/VR. VR/AR can really enter the consumer and practical stage. The performance of the GPU needs to be greatly improved in order to meet the needs of our users. For AR/VR technology to be well applied, it is necessary to build a huge virtual reality database and development tools. This is beyond the capabilities of a single



enterprise. It requires the state to provide certain funding and policy support to promote the development of technology and the cultivation of the industry.

In the future, with the development of augmented reality technology and the popularization of corresponding display equipment, its display method and effect will certainly have great progress and development than now, the effect of design and production will be more realistic, more in line with the habits of users, so that people can get more diversified and rich information.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Scientific Research Funding Project of Liaoning Provincial Department of Education (No. JZR2019005), Dalian Science and Technology Innovation Fund Project (No. 2019J13SN112), and Guiding Plan of Liaoning Natural Science Foundation (No. 2019-ZD-0352).

## References

- [1] H. E. Zehao, S. U. I. Xiaomeng, Z. H. A. O. Yan, C. A. O. Liangcai, and J. I. N. Guofan, "The development trend of virtual reality and augmented reality technology based on holographic optics," *Science & Technology Review*, vol. 36, no. 9, pp. 8–17, 2018.
- [2] H.-H. Liou, S. J. H. Yang, S. Y. Chen, and W. Tarng, "The influences of the 2d image-based augmented reality and virtual reality on student learning," *Educational Technology & Society*, vol. 20, no. 3, pp. 110–121, 2017.
- [3] W. Bennie, G. T. S. Ho, and E. Tsui, "Development of an intelligent E-healthcare system for the domestic care industry," *Industrial Management & Data Systems*, vol. 117, no. 17, pp. 1426–1445, 2017.
- [4] A. Sharma, P. Bajpai, S. Singh, and K. Khatter, "Virtual reality: blessings and risk assessment," *Indian Journal of Science & Technology*, vol. 11, no. 20, pp. 1–20, 2018.
- [5] L. Hu, Y. Tian, J. Yang, T. Taleb, L. Xiang, and Y. Hao, "Ready player one: UAV-clustering-based multi-task offloading for vehicular VR/AR gaming," *IEEE Network*, vol. 33, no. 3, pp. 42–48, 2019.
- [6] V. T. Nguyen, R. Hite, and T. Dang, "Learners' technological acceptance of VR content development: a sequential 3-part use case study of diverse post-secondary students," *International Journal of Semantic Computing*, vol. 13, no. 3, pp. 343–366, 2019.
- [7] M. Zhang, M. Wen, X. Cheng, and L. Yang, "A dual-hop virtual MIMO architecture based on hybrid differential spatial modulation," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6356–6370, 2016.
- [8] E. Pegon-Machet, D. Faulks, K. A. Eaton, E. Widström, P. Hugues, and S. Tubert-Jeannin, "The healthcare system and the provision of oral healthcare in EU member states: France," *British Dental Journal*, vol. 220, no. 4, pp. 197–203, 2016.
- [9] K. Malkiewicz, E. Malkiewicz, K. A. Eaton, and E. Widström, "The healthcare system and the provision of oral healthcare in European Union member states. Part 6: Poland," *British Dental Journal*, vol. 221, no. 8, pp. 501–507, 2016.
- [10] E. Nilsson, A. Gasparini, J. Ärnlov et al., "Incidence and determinants of hyperkalemia and hypokalemia in a large healthcare system," *International Journal of Cardiology*, vol. 245, no. 3, pp. 277–284, 2017.
- [11] J. Hanen, Z. Kechaou, and M. B. Ayed, "An enhanced healthcare system in mobile cloud computing environment," *Vietnam Journal of Computer Science*, vol. 3, no. 4, pp. 267–277, 2016.
- [12] J. C. Chrisler, A. Barney, and B. Palatino, "Ageism can be hazardous to women's health: ageism, sexism, and stereotypes of older women in the healthcare system," *Journal of Social Issues*, vol. 72, no. 1, pp. 86–104, 2016.

## Research Article

# Patient-Specific CT-Based Fluid-Structure-Interaction Aorta Model to Quantify Mechanical Conditions for the Investigation of Ascending Aortic Dilation in TOF Patients

Heng Zuo<sup>1,2</sup>, Yunfei Ling,<sup>3</sup> Peng Li,<sup>1</sup> Qi An,<sup>3</sup> and Xiaobo Zhou<sup>4</sup>

<sup>1</sup>Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>2</sup>School of Mathematics, Sichuan Normal University, Chengdu 610068, China

<sup>3</sup>Department of Cardiovascular Surgery, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>4</sup>Center for Computational Systems Medicine and School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA

Correspondence should be addressed to Heng Zuo; zuoheng@sicnu.edu.cn

Received 23 April 2020; Revised 10 June 2020; Accepted 8 July 2020; Published 8 August 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Heng Zuo 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.** Some adult patients with Tetralogy of Fallot (TOF) were found to simultaneously develop ascending aortic dilation. Severe aortic dilation would lead to several aortic diseases, including aortic aneurysm and dissection, which seriously affect patients' living quality and even cause patients' death. Current practice guidelines of aortic-dilation-related diseases mainly focus on aortic diameter, which has been found not always a good indicator. Therefore, it may be clinically useful to identify some other factors that can potentially better predict aortic response to dilation. **Methods.** 20 TOF patients scheduled for TOF repair surgery were recruited in this study and were divided into dilated and nondilated groups according to the Z scores of ascending aorta diameters. Patient-specific aortic CT images, pressure, and flow rates were used in the construction of computational biomechanical models. **Results.** Simulation results demonstrated a good coincidence between numerical mean flow rate at inlet and the one obtained from color Doppler ultrasonography, which implied that computational models were able to simulate the movement of the aorta and blood inside accurately. Our results indicated that aortic stress can effectively differentiate patients of the dilated group from the ones of the nondilated group. Mean ascending aortic stress-P1 (maximal principal stress) from the dilated group was 54% higher than that from the nondilated group (97.97 kPa vs. 63.47 kPa,  $p$  value = 0.044) under systolic pressure. Velocity magnitude in the aorta and aortic wall displacement of the dilated group were also greater than those of the nondilated group with  $p$  value < 0.1. **Conclusion.** Computational modeling and ascending aortic biomechanical factors may be used as a potential tool to identify and analyze aortic response to dilation. Large-scale clinical studies are needed to validate these preliminary findings.

## 1. Introduction

Tetralogy of Fallot (TOF) is a congenital heart defect which involves four anatomical abnormalities of the heart: ventricular septal defect, pulmonary infundibular stenosis, overriding aorta, and right ventricular hypertrophy. In the United States, the prevalence of TOF is 3.9 per 10,000 live births and accounts for 7-10% of congenital heart diseases [1]. Some adult patients with congenital heart disease were found

to simultaneously develop ascending aortic dilation [2-4]. A dilatation of all wall layers of the aorta of more than 50% in comparison to the normal diameter is defined as the ascending aortic aneurysm [5], which is generally indolent and asymptomatic until presentation with catastrophic complications of rupture and dissection. When a rupture or dissection occurs, it is fatal in a large proportion of patients. During the acute phase of aortic dissection, more than 50% of patients die if not treated surgically, and the mortality of those treated

ranges from 5% to 30% despite the significant improved operative strategies [6–8].

The risk of aortic dissection is known to highly correlate with increasing aortic diameter. Current practice guidelines have recommended elective aortic repair at a diameter of 5.5 cm because the risk of rupture of an aneurysm is found to increase dramatically with ascending aortic diameter greater than 6 cm [9–13]. However, recent studies have found that many patients with acute ascending aortic dissection have aortic diameters of <5.5 cm at hospital presentation [14, 15]. Moreover, Ming et al. conducted a systematic review and meta-analysis about the natural history of ascending aortic aneurysm and found the conclusions about the conventional risk factors, such as aortic diameters, not only varied significantly but were often contradictory among studies [16]. For example, 4 studies [17–20] identified that aortic dilation had no association with baseline aortic diameter, while 2 studies [21, 22] reported higher annual aortic growth rate with lower baseline aortic diameter. Therefore, more and more researches focus on searching for more reliable risk factors of ascending aortic dilation rather than aortic diameter only. Geisbusch et al. proposed computed tomography volume measurements which provide an objective method for ascertaining aortic size and monitoring expansion [23]. Donato Aquaro et al. used the maximum rate of systolic distension (MRSD) as an index of aortic wall properties and found MRSD was a valuable predictor for progression in ascending aortic dilation [24]. Alreshidan et al. used speckle tracking echocardiography to estimate aortic stiffness in vivo and proved that the stiffness was helpful to determinate the risk of complications in patients with aortic diseases [25]. Pasta et al. used CT angiographic imaging to estimate patient-specific in vivo strain fields and found stiff behavior for the aneurysmal aorta compared with that of the healthy ascending aorta [26]. Farzaneh et al. proposed a noninvasive method by using gated CT scans to identify the patient-specific local extensional stiffness of aortic walls, which was found to be highly correlated with the rupture risk criterion [27, 28].

With the rapid improvement of computer and medical imaging technologies, image-based computational models have been widely used in the research of cardiovascular diseases. Youssefi et al. used computational fluid-only biomechanical models to investigate the impact of inflow velocity profiles on hemodynamics of the thoracic aorta [29]. Condemni et al. performed patient-specific computational fluid dynamics analysis to obtain the effect of hemodynamics on the ascending thoracic aortic aneurysm risk of rupture [30] and also investigated the hemodynamic alterations for a patient who was treated by only ascending aorta replacement preserving the BAV [31]. Underhill et al. constructed image-based computational models by integrating plaque morphology, components, and mechanical stress/strain conditions to assess the vulnerability of human atherosclerotic carotid plaques [32]. According to some studies associated with aortic diseases, biomechanics of aortic tissue has been found to be a possible good predictor of the histologic integrity of the aortic wall [17, 33–35]; therefore, biomechanics might be a better risk factor of aortic diseases than aortic diameter.

The objective of this study is to construct a computational biomechanical model by combining patient-specific CT scans, aortic pressure measurements, color Doppler ultrasonography, and aortic tissue properties for a better understanding of mechanical environment of the ascending aorta, which may provide a more useful insight about the ascending aortic dilation.

## 2. Data Acquisition, Models, and Methods

**2.1. Data Acquisition.** Twenty TOF patients scheduled for TOF repair surgery were recruited in this study, which was approved by the review board on human subject research (West China Hospital, West China School of Medicine), and informed consent was obtained.

ECG-gated cardiac computed tomography (CT) scans were performed in one cardiac cycle for each patient, and scans of approximately 10 time points were obtained. The CT scans of each time point contained around a total of 70 slices (slice thickness was 3 mm) covering the whole cardiac structure and ascending aorta, and scans under diastolic and systolic pressure were used in the model construction. 3D geometry of the aortic root and ascending aorta was obtained from CT scans using thresholding-segmentation technique implemented in the Mimics (Materialise, Belgium). The lumen of the aortic root and ascending aorta were acquired by using Mimics segmentation module “CT Heart Segmentation” with the recommended parameters. Due to the image qualities, patient-specific aortic wall is hard to obtain and a uniform wall thickness 2.0 mm was used for all the patients in this paper. Figure 1 shows selected CT scans, segmented results of the aorta, and the corresponding 3D reconstructed geometry for one patient in our dataset. The fluid-structure-interaction (FSI) finite element (FE) model of the ascending aorta would be constructed based on the CT-based 3D geometry and color Doppler ultrasonographic flow information to quantify the mechanical properties of aorta.

Patients’ ascending aortic diameters were calculated from the 3D reconstructed geometries and converted to their corresponding  $Z$  scores by using a web-based calculation tool (<http://zscore.chboston.org>) which had collected baseline data over the past 12 years [36]. On account of the inclusion of patient-specific gender, age, and body surface area (BSA),  $Z$  score has the advantage in determining whether an aorta is within normal size limits. Rather than raw aortic diameter, aortic  $Z$  score is more accurate in determining whether true pathology exists. We evenly divided all the patients into 2 groups according to their  $Z$  scores of ascending aortic diameters, where the group with larger  $Z$  scores was defined as the dilated group and the one with smaller  $Z$  scores was the non-dilated group. The patient-specific systolic and diastolic blood pressure was measured at the same time when performing CT scanning, and they were used as boundary conditions in the simulations. Patient-specific color Doppler ultrasonographic imaging of blood flow in the ascending aorta was obtained and used in the determination of patient-specific boundary condition.

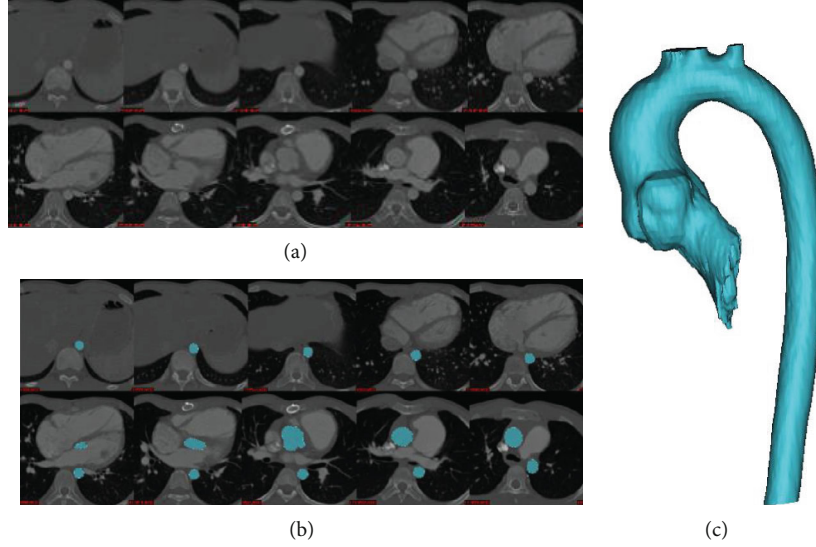


FIGURE 1: CT-based model construction process. (a) Selected CT slices from a patient, under diastolic pressure; (b) segmented results of the aorta; (c) reconstructed 3D geometry of the aorta.

Patients' demographic information, aorta diameters and corresponding  $Z$  scores, systolic and diastolic blood pressures, and blood properties are summarized in Table 1. Continuous variables (such as age, aorta diameter, blood viscosity, and pressure) were summarized as mean  $\pm$  SD and compared between the outcome groups using an unpaired Student  $t$ -test. Categorical variables (such as gender) were compared between different groups using a nonparametric test.

**2.2. The Anisotropic Ascending Aorta Model.** The aorta material was assumed to be hyperelastic, anisotropic, nearly incompressible, and homogeneous. The governing equations for the structure model are as follows:

$$\begin{aligned} \rho \frac{\partial^2 u_i}{\partial t^2} &= \frac{\partial \sigma_{ij}}{\partial x_j}, \quad i = 1, 2, 3, \\ \epsilon_{ij} &= \frac{1}{2} \left( \frac{\partial u_j}{\partial a_i} + \frac{\partial u_i}{\partial a_j} \right) + \sum_l \frac{\partial u_l}{\partial a_i} \frac{\partial u_l}{\partial a_j}, \quad i, j = 1, 2, 3. \end{aligned} \quad (1)$$

Here,  $\sigma$  is the stress tensor,  $\epsilon$  is Green's strain tensor,  $u$  is the displacement, and  $\rho$  is the material density.

The nonlinear Mooney-Rivlin model was used to describe the aorta properties, which has been widely used to model anisotropic hyperelastic organs including ventricular tissue [37] and vessel [38]. The strain energy function for the anisotropic modified Mooney-Rivlin model is as follows [39]:

$$\begin{aligned} W &= c_1(I_1 - 3) + c_2(I_2 - 3) + D_1[\exp(D_2(I_1 - 3)) - 1] \\ &\quad + K_1/K_2 \exp[K_2(I_4 - 1)^2 - 1], \\ I_1 &= \sum C_{ii}, \\ I_2 &= \frac{1}{2} [I_i^2 - C_{ij}C_{ij}], \end{aligned} \quad (2)$$

where  $I_1$  and  $I_2$  are the first and second strain invariants,  $C = [C_{ij}] = X^T X$  is the right Cauchy-Green deformation tensor,  $X = [X_{ij}] = [\partial x_i / \partial a_j]$  ( $x_i$  is the current position,  $a_i$  is the original position),  $I_4 = C_{ij}(\mathbf{n}_f)_i(\mathbf{n}_f)_j$ ,  $\mathbf{n}_f$  is the fiber direction which was set to circumferential in our study.  $c_i$ ,  $K_i$ , and  $D_i$  are material parameters. In our models, initial values of material parameters were obtained by fitting the experimental stress-stretch data of human aorta samples [40] with the goodness of fit ( $R^2$ ) 0.73:  $c_1 = -525.16$  kPa,  $c_2 = 165.9$  kPa,  $D_1 = 231.8$  kPa,  $D_2 = 3.5$ ,  $K_1 = 33.4$  kPa, and  $K_2 = 12.6$ . The patient-specific structure-only ascending aorta model was constructed using the geometry under diastolic pressure; then, the pressure difference between systolic and diastolic pressures was applied as boundary condition on the inner surface of the aorta to pressurize the aorta to its shape under systolic pressure. Patient-specific material parameters were acquired through adjusting initial material parameters by one ratio to match the numerical volume of the ascending aorta under systolic pressure with the one calculated from CT scans (relative error  $< 5\%$ ). The normal stress on the outer aorta surface was assumed to be zero. On the inner aorta surfaces, we applied fluid-structure-interaction boundary conditions.

**2.3. Blood Model.** Blood flow was assumed to be laminar, Newtonian, and incompressible. The Navier-Stokes equations with arbitrary Lagrangian-Eulerian formulation were used as the governing equations. No-slip conditions and natural traction equilibrium conditions were assumed at fluid-structure interfaces, pressure curves were adjusted according to patient-specific blood pressure, and flow rate information obtained from color Doppler ultrasonography was imposed at inlet (location of the aortic valve) and outlet (intersection between the ascending aorta and the aortic arch) of the aorta (see Figure 2).

TABLE 1: Patients' demographic information, ascending aorta diameters, blood pressures, and blood properties.

	Age (year)	Sex	Height	Weight	SBP (mmHg)	DBP (mmHg)	AAD (mm)	AAD Z-score	Viscosity (cPoise)
Nondilated group									
P1	6.9	F	118	19.0	108	62	17.59	0.64	4.28
P2	21.8	M	163	50.0	130	85	27.18	1.73	10.01
P3	8.7	M	126	22.0	107	61	20.69	1.76	4.08
P4	8.4	M	119	20.0	98	64	20.16	1.86	6.73
P5	6.1	M	110	18.0	105	59	20.12	2.34	6.48
P6	21.3	F	150	45.0	105	56	30.32	3.13	6.07
P7	15.2	M	162	40.0	112	62	25.29	3.92	6.08
P8	15.1	M	170	52.0	95	58	30.32	4.11	4.14
P9	20.3	M	171	65.0	140	96	35.43	4.12	7.18
P10	12.0	M	138	31.5	88	57	29.76	4.15	2.78
Mean $\pm$ SD	13.5 $\pm$ 6.1		143 $\pm$ 23	36.3 $\pm$ 16.6	109 $\pm$ 16	66 $\pm$ 13	25.68 $\pm$ 5.86	2.78 $\pm$ 1.27	5.78 $\pm$ 2.06
Dilated group									
P11	7.6	F	129	20.0	79	50	20.60	5.81	3.82
P12	7.8	M	113	16.0	109	74	25.70	5.88	7.22
P13	10.4	F	125	23.0	105	60	30.89	7.03	7.27
P14	8.8	M	115	20.0	105	70	27.79	7.35	8.83
P15	19.9	M	166	46.0	133	77	40.17	7.45	8.59
P16	13.0	M	157	37.0	118	54	35.96	7.82	3.69
P17	13.2	M	152	28.0	129	69	37.07	8.66	8.30
P18	19.3	M	170	45.0	99	63	39.26	9.72	7.64
P19	8.6	F	120	15.0	90	60	28.71	11.36	4.74
P20	6.9	F	106	15.0	99	40	29.17	11.90	4.10
Mean $\pm$ SD	11.6 $\pm$ 4.8		135 $\pm$ 24	26.5 $\pm$ 12.0	107 $\pm$ 17	62 $\pm$ 11	31.53 $\pm$ 6.38	8.30 $\pm$ 1.27	6.42 $\pm$ 2.09
<i>p</i> value	0.42		0.49	0.15	0.76	0.45	0.051	<0.01	0.50

SBP: systolic blood pressure; DBP: diastolic blood pressure; AAD: ascending aortic diameter.

$$\begin{aligned}
\rho \left( \frac{\partial \mathbf{u}}{\partial t} + ((\mathbf{u} - \mathbf{u}_g) \cdot \nabla) \mathbf{u} \right) &= -\nabla p + \mu \nabla^2 \mathbf{u}, \\
\nabla \cdot \mathbf{u} &= 0, \\
\mathbf{u}|_T &= \frac{\partial \mathbf{x}}{\partial t}, \\
p|_{\text{inlet}} &= p_{\text{in}}(t), \\
p|_{\text{outlet}} &= p_{\text{out}}(t), \\
\sigma_{ij}^f \cdot \mathbf{n}_j|_{\text{interface}} &= \sigma_{ij}^s \cdot \mathbf{n}_j|_{\text{interface}},
\end{aligned} \tag{3}$$

where  $\mathbf{u}$  and  $p$  are the fluid velocity and pressure,  $\mathbf{u}_g$  is the mesh velocity,  $\mu$  is the dynamic viscosity,  $\rho$  is the density (set to  $1 \text{ g}\cdot\text{cm}^{-3}$ ),  $T$  stands for the aorta inner boundary,  $\sigma$  is the stress tensor (superscript letters “f” and “s” indicate fluid and solid materials, respectively),  $\epsilon$  is the strain tensor, and  $\mathbf{v}$  is the solid displacement vector. Aortic pressure curve in the Wiggers diagram (Figure 3) was adjusted according to patient-specific diastolic/systolic blood pressure and used as  $p_{\text{in}}(t)$  prescribed at the inlet of the fluid domain, and the pressure condition at the outlet  $p_{\text{out}}(t)$  was obtained by adjusting  $p_{\text{in}}(t)$  to match the numerical mean velocity magnitude at

inlet with the one measured by color Doppler ultrasonography (relative error < 5%). Patient-specific blood viscosity  $\mu$  was calculated from the following equation [41]:

$$\begin{aligned}
\mu \text{ (cPoise)} &= 0.12 \text{ hematocrit (\%)} \\
&+ 0.17 \text{ total plasma protein (g/L)} - 2.07.
\end{aligned} \tag{4}$$

**2.4. Preshrink Process.** Numerical simulation needs to start from an initial condition where the initial aorta shape, blood pressure, and stress/strain distributions were provided. Unfortunately, it is hard to measure stress distributions of soft tissues in vivo; therefore, our numerical simulations would start from zero-load state when blood pressure in the aorta was zero so that initial stress and strain distributions in the aorta were approximately considered as zero. However, CT scans were performed under the in vivo condition where the aorta was pressurized; thus, the zero-load shape was not obtained directly from CT scans. In our model construction process, a preshrink process was applied to the in vivo aorta shape to generate the zero-load shape. The initial shrinkage rate for the aorta was set to 5%, and diastolic blood pressure was applied so that the zero-load aorta shape would regain its in vivo morphology. The shrinkage rate was adjusted until



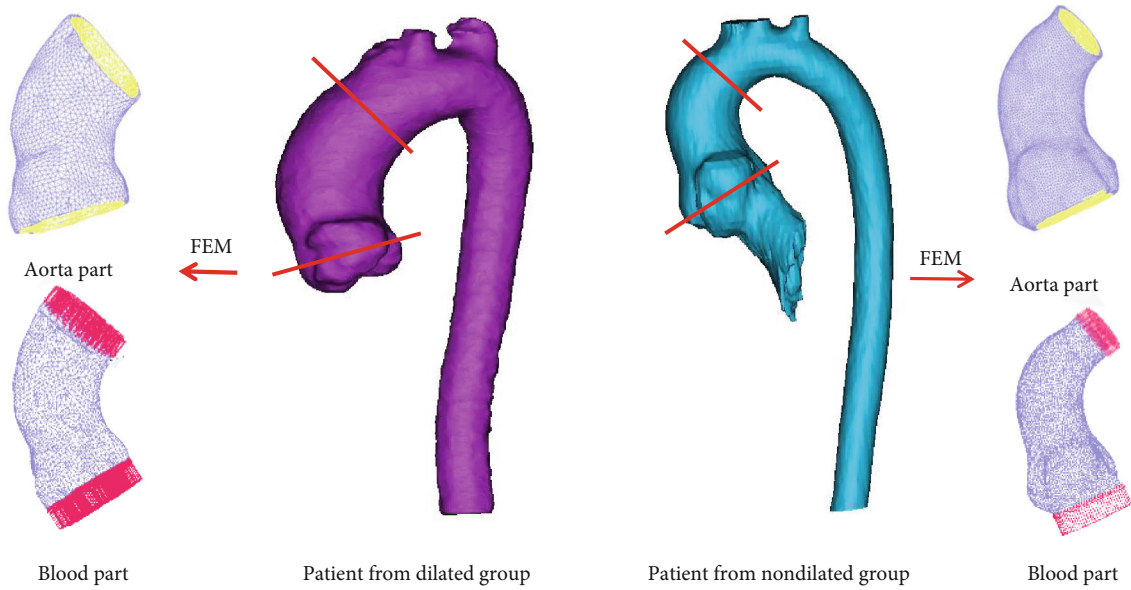


FIGURE 2: In the middle: 3D reconstructed aorta of two representatives from each group. On the left: structure mesh (up) and fluid mesh with imposed pressure conditions (down) of the dilated group representative. On the right: structure mesh (up) and fluid mesh with imposed pressure conditions (down) of the nondilated group representative.

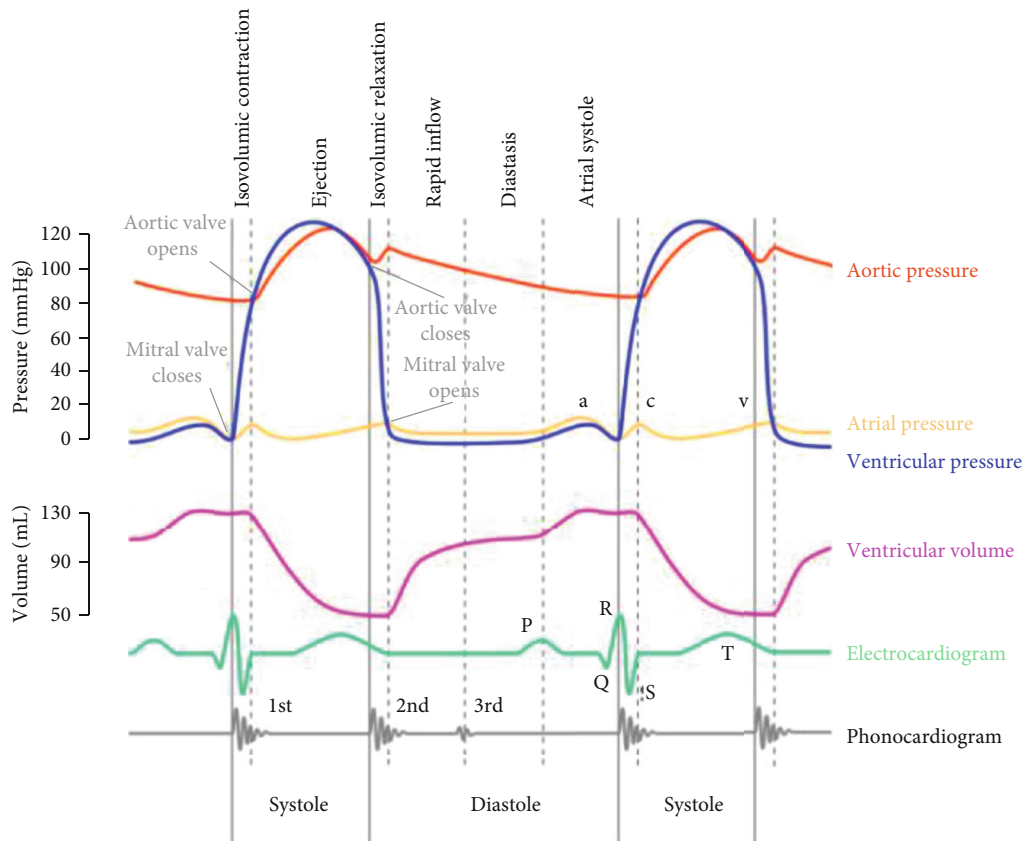


FIGURE 3: Prescribed pressure curve. Aortic pressure curve (the red curve) in the Wiggers diagram was adjusted according to patient-specific diastolic and systolic pressures and used as the pressure curve in the simulations.

pressurized numerical aorta shape highly agreed with CT scans (the relative error between the numerical volume of the ascending aorta section and the one calculated from CT

scans < 5%). Without this preshrink process, the actual computing domain would be greater than the actual aorta due to the initial expansion when pressure was applied.

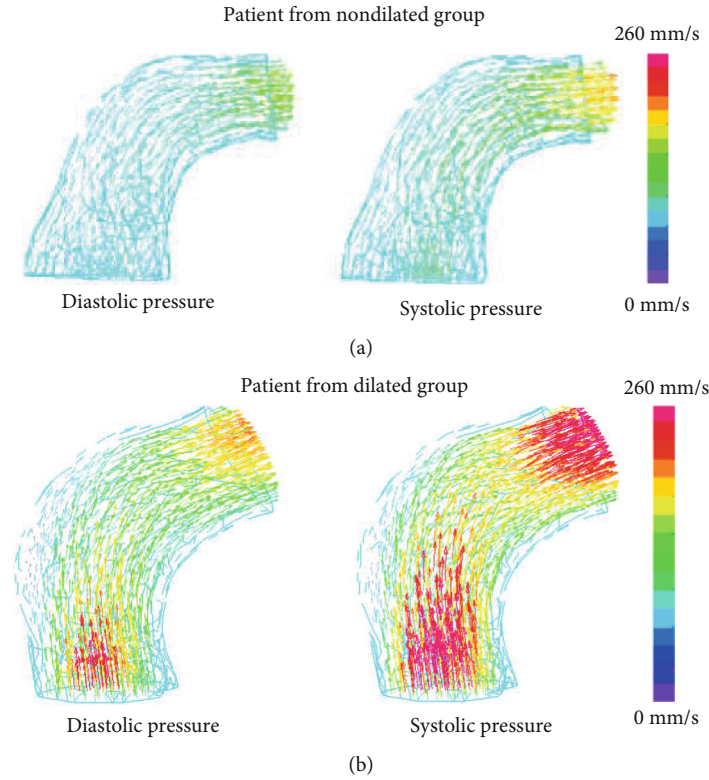


FIGURE 4: Flow distributions of the representative models: (a) flow distributions of a patient from the nondilated group at diastolic pressure (left) and systolic pressure (right); (b) flow distributions of a patient from the dilated group.

**2.5. Geometry-Fitting Mesh Generation and Solution Method.** Due to the complex irregular geometries of the ascending aorta, a geometry-fitting mesh generation technique [42] was adopted to generate meshes for our models. The fully coupled FSI model was solved by ADINA (ADINA R&D, Watertown, MA) using unstructured finite elements and the Newton-Raphson iteration method. Mesh analysis was performed by decreasing mesh size by 10% (in each dimension) until solution differences in maximal stress-P1/strain-P1 predictions were less than 2%, where stress-P1 and strain-P1 mean the maximal principal stress and strain, respectively. In our models, the optimal element size in each dimension was around 0.5 mm; the final number of tetrahedral meshes was about 3000 for the fluid domain and about 2000 for the solid domain.

**2.6. Statistical Analysis.** Due to the small sample size, the Shapiro-Wilk test was used to test if the data satisfied the normal distribution. Also, Levene's test was used to determine if variances for a variable calculated from two different groups were equal. Based on the results of Shapiro-Wilk and Levene's tests, the appropriate test for the comparison of means, such as Student's *t*-test or unequal variances *t*-test, was chosen to compare the differences in the numerical mechanical results between the nondilated and dilated groups.

### 3. Results

**3.1. Blood Flow Dynamics in the Ascending Aorta.** One patient was chosen from each group to represent flow distri-

butions in the aorta; the streamlines of blood velocity fields in the two patients under systolic/diastolic pressure are shown in Figure 4. Numerical mean and inlet velocity magnitude in the ascending aorta of all the patients were summarized in Table 2, where inlet velocity means the average velocity magnitude over the inlet surface. It is worth mentioning that numerical volumes of the ascending aortic segment were compared with the one measured by CT scans, and relative errors of volume for all patients were less than 8%. The results of Shapiro-Wilk and Levene's test showed that both mean and inlet velocity magnitudes in each group satisfied the normal distribution and no significant different variances were found between the groups. Therefore, Student's *t*-test was appropriate to compare the means of velocity magnitudes between two groups. The results indicated that the dilated group had larger velocity magnitudes than the nondilated group, but the differences were not significant due to 5% level of significance. Inlet velocity magnitudes in the dilated group were greater than those in the nondilated group at the time of both systolic and diastolic pressure (*p* value: 0.058 for systolic pressure time and 0.097 for diastolic pressure time).

#### 3.2. Mechanical Analysis of the Ascending Aorta

**3.2.1. Displacement Analysis.** Representative models' displacement distribution at the time of systolic pressure was presented in Figure 5. All patients' mean and maximal displacement values and corresponding test results were summarized in Table 3. Student's *t*-test results showed that no

TABLE 2: Patients’ numerical velocity magnitude in the ascending aorta.

Diastolic pressure				Systolic pressure			
Inlet velocity (mm/s)		Mean velocity (mm/s)		Inlet velocity (mm/s)		Mean velocity (mm/s)	
Nondilated	Dilated	Nondilated	Dilated	Nondilated	Dilated	Nondilated	Dilated
119.69	135.88	29.34	23.14	166.67	185.10	40.50	31.57
136.27	137.17	29.48	34.97	194.24	230.54	42.39	53.81
144.56	169.83	31.56	40.87	197.95	230.85	44.91	57.91
146.73	174.11	34.12	41.00	210.61	231.60	47.41	58.37
150.44	175.34	34.48	42.60	210.86	241.72	49.21	59.78
153.99	179.42	35.36	46.67	214.02	242.13	49.77	61.86
160.59	187.01	46.60	51.58	219.28	249.11	63.01	68.40
160.94	189.70	46.64	53.70	238.50	269.57	68.07	72.33
184.46	190.39	50.92	57.95	247.49	302.21	68.77	90.59
207.19	262.59	60.00	62.17	271.10	391.20	78.55	90.71
Mean ± SD							
156.49 ± 24.57	180.14 ± 34.98	39.85 ± 10.48	45.47 ± 11.54	217.07 ± 29.49	257.40 ± 55.70	55.26 ± 13.19	64.53 ± 17.49
Shapiro-Wilk test							
0.54	0.04	0.12	0.92	0.93	0.03	0.18	0.39
Levene’s test							
0.64		0.79		0.44		0.71	
<i>p</i> value							
0.097		0.270		0.058		0.197	

Inlet velocity means the average velocity over inlet surface. Mean velocity means the average velocity of all nodes in the ascending aorta. Value ordered from min to max in each group. *p* value stands for Students’ *t*-test between the nondilated and dilated groups.

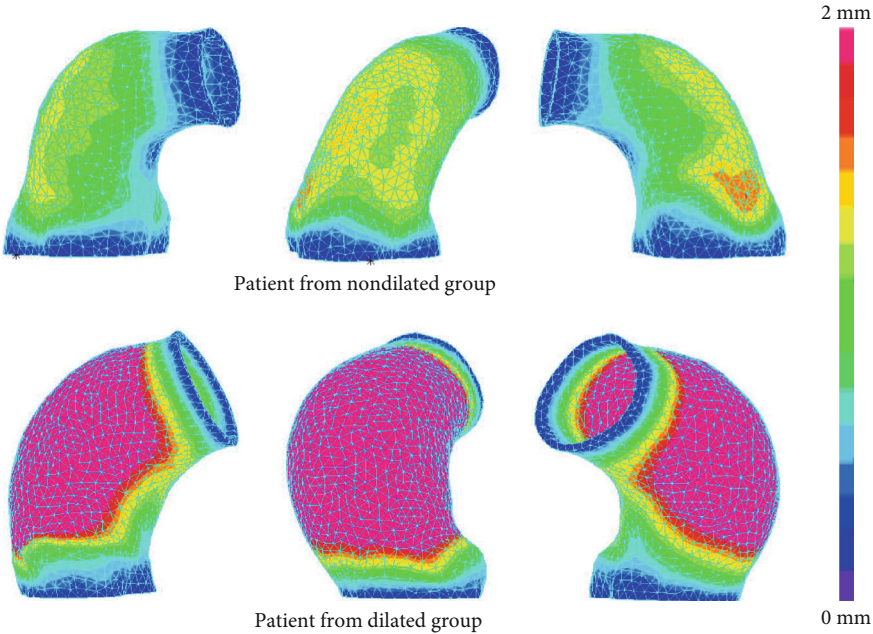


FIGURE 5: Displacement distributions of the representative models: the top row shows the displacement distributions of a patient from the nondilated group in three different views; the bottom row shows the displacement distributions of a patient from the dilated group. The same color band used to better present the differences between the two models.

significant differences were found in mean displacement at the time of systolic pressure between the two groups (*p* value: 0.056). However, maximal systolic pressure displacement of

the dilated group was found to be significantly larger than that of the nondilated group through Student’s *t*-test with *p* value 0.038.

TABLE 3: Patients' numerical displacement results in the ascending aorta.

Max displacement (mm)		Systolic pressure	
		Mean displacement (mm)	
Nondilated	Dilated	Nondilated	Dilated
0.107	0.113	0.049	0.042
0.113	0.205	0.054	0.088
0.126	0.220	0.057	0.101
0.167	0.231	0.076	0.108
0.189	0.243	0.089	0.111
0.194	0.250	0.091	0.115
0.197	0.397	0.092	0.193
0.249	0.445	0.117	0.208
0.257	0.454	0.127	0.221
0.372	0.514	0.175	0.236
Mean $\pm$ SD			
0.197 $\pm$ 0.081	0.307 $\pm$ 0.133	0.093 $\pm$ 0.039	0.142 $\pm$ 0.066
Shapiro-Wilk test			
0.27	0.25	0.31	0.22
Levene's test			
0.19		0.17	
<i>p</i> value			
0.038		0.056	

Value ordered from min to max in each group. *p* value stands for Student's *t*-test between the nondilated and dilated groups.

**3.2.2. Internal Stress Analysis in the Ascending Aorta.** Figure 6 shows stress-P1 distributions of representative models for two different groups, and stress-P1 means the maximal principal stress. Average stress-P1 values in the ascending aorta of the 20 patients were summarized in Table 4. Mean stress-P1 values of the dilated group were significantly larger than those of the nondilated group at the time of systolic pressure with *p* value 0.044. Diastolic pressure mean stress-P1 of the dilated group was highly greater than that of the nondilated group with *p* value 0.089.

#### 4. Discussion

In previous researches about TOF patients, while main attentions were paid on RV pathology including RV hypertrophy, pulmonary regurgitation, and RV outflow tract obstruction, few attentions were paid on LV. However, more and more TOF patients were found to have ascending aortic dilation of different levels in recent clinical observations. Severe aortic dilation would lead to some aortic diseases, such as aortic aneurysm and dissection, which seriously affect patients' living quality and even cause patients' death. Conventional clinical guidelines of aortic dilation only rely on the aortic diameter and ignore the tissue properties of the aorta which have been proved to own better abilities in prediction of aortic dilation progression especially in patients with small-to-moderate size ascending aortas [24]. Considering the good performances of computational biomechanical modeling in the investigations of cardiovascular diseases, we proposed

image-based computational models in this study to simulate the movement of the ascending aorta and blood inside to obtain a better understanding of ascending aortic mechanics. To our knowledge, this is the first study applying biomechanical computational modeling in the research of ascending aortic dilation for TOF patients.

This study provides proof of concept that computational modeling based on patient-specific CT images can be used to simulate ascending aorta response to aortic dilation. Computational biomechanical models integrate patient-specific CT images, aortic pressure measurements, tissue properties, and color Doppler images through well-established equations of motion, constitutive equations of materials, and finite element method. Simulation results were patient-specific by comparing the numerical results with the ones obtained from clinical measurements. These complex computational models have provided new insights into mechanical environment (stress, strain, and flow distributions) in the ascending aorta, which is able to provide more information than aortic diameter only. In our simulations, sometimes patients with larger ascending aortic diameters were found to present lower stress in the aorta. For instance, patient #2 (Table 1) with ascending aortic diameter 27.18 mm had mean stress-P1 of the whole ascending aorta 89.44 kPa at systolic pressure and 55.33 kPa at diastolic pressure, while patient #8 with ascending aortic diameter 30.32 mm had mean stress-P1 71.67 kPa at systolic pressure and 41.77 kPa at diastolic pressure. The phenomenon happening indicated that diameter-only criterion for aortic dilation was insufficient. In our study, we divided the patients into the dilated and nondilated groups according to *Z* scores rather than raw measurements of ascending aorta diameters, since *Z* scores have been proven more capable to reflect the dilation degree of the aorta than raw aorta diameter measurements. Differences in *Z* scores between the dilated and nondilated groups were highly significant (*p* value < 0.01), while raw diameter measurements were not significantly different between the two groups (*p* value 0.051). According to numerical results, the dilated group was more likely to expand than the nondilated group. Meanwhile, the dilated group also tended to have larger stress than the other group indicating larger probability of vessel rupture for the dilated group. The mean flow rates in the ascending aorta of the dilated group were found to be larger than the ones of the nondilated group. Larger flow rates indicated the larger acting force of blood on the vessels which may cause the aortic dilation. In a word, the biomechanical model is a complex combination of aortic geometry, tissue property, and pressure which could provide more intrinsic insights into aortic dilation, dissection, and rupture. Moreover, the biomechanical models are able to display the distribution of mechanical parameters (see Figures 4–6) which is helpful in determination of focal area due to dilation.

The current study adds computational modeling as a new investigative tool and stress as new potential predictors for ascending aortic dilation. To obtain the mechanics of aortas as correctly as possible, the model assumptions were set to as close to reality as possible and the data including aortic geometry, pressure, and flow rate were all patient-specific.



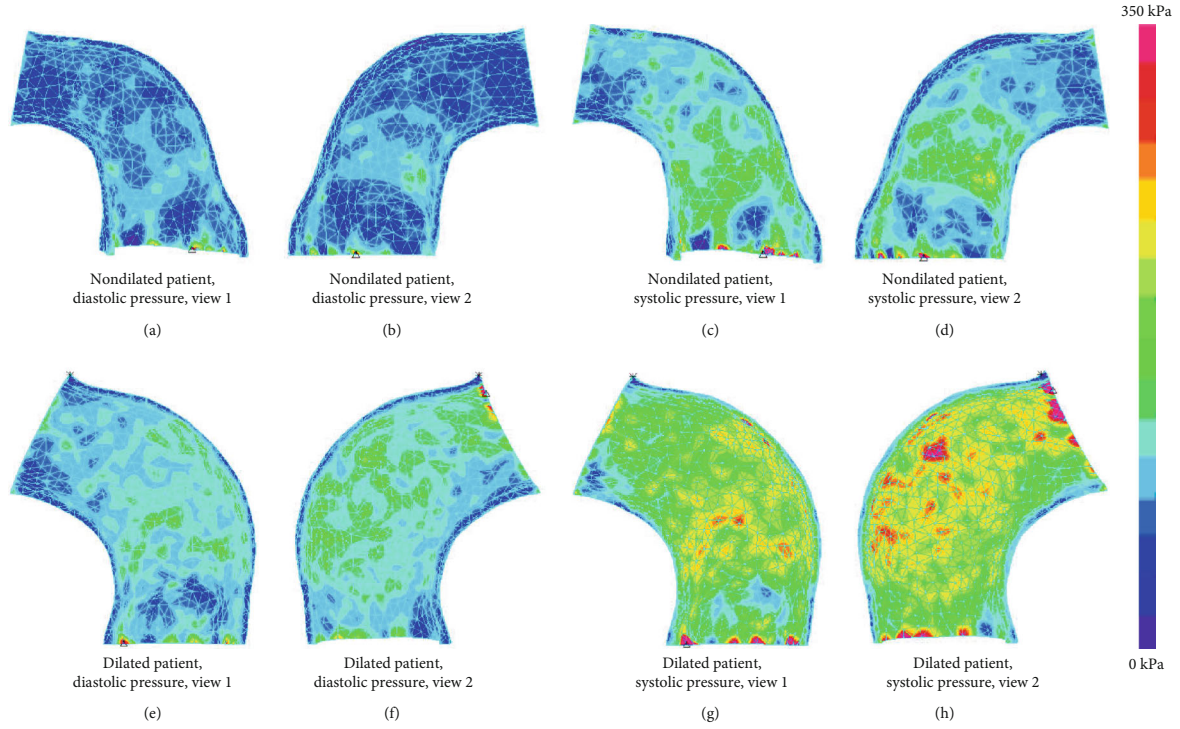


FIGURE 6: Stress distributions of the representative models: (a, b) the stress distributions of a patient from the nondilated group at the time of diastolic pressure, (c, d) the stress distributions of a patient from the nondilated group at the time of systolic pressure, (e, f) the stress distributions of a patient from the dilated group at the time of diastolic pressure, and (g, h) the stress distributions of a patient from the dilated group at the time of systolic pressure. The same color band used to better present the differences between the two models.

TABLE 4: Patients' numerical stress results in the ascending aorta.

Diastolic pressure				Systolic pressure			
Max stress-P1 (kPa)		Mean stress-P1 (kPa)		Max stress-P1 (kPa)		Mean stress-P1 (kPa)	
Nondilated	Dilated	Nondilated	Dilated	Nondilated	Dilated	Nondilated	Dilated
158.43	172.41	15.63	20.86	250.23	272.17	28.82	34.09
166.76	265.08	19.77	28.36	272.75	440.02	31.81	61.82
338.61	302.73	21.37	39.39	581.86	634.24	39.85	76.13
376.63	337.26	25.15	48.35	608.32	687.91	40.43	76.98
436.37	413.98	32.47	49.40	636.05	702.11	60.17	79.49
438.42	441.59	35.04	52.45	753.76	711.01	66.93	89.61
464.80	464.23	41.41	56.66	771.48	764.58	71.67	121.22
546.69	594.70	41.77	71.14	803.06	846.03	82.54	134.07
578.26	617.56	55.33	73.59	959.85	884.96	89.44	140.91
649.82	793.40	79.77	90.10	963.93	1112.47	123.10	165.40
Mean $\pm$ SD							
415.48 $\pm$ 162.40	440.29 $\pm$ 186.55	36.77 $\pm$ 19.35	53.03 $\pm$ 21.09	660.13 $\pm$ 247.47	705.55 $\pm$ 232.08	63.47 $\pm$ 29.77	97.97 $\pm$ 40.76
Shapiro-Wilk test							
0.57	0.90	0.20	0.95	0.27	0.83	0.49	0.75
Levene's test							
0.65		0.75		0.65		0.38	
<i>p</i> value							
0.755		0.089		0.677		0.044	

Value ordered from min to max in each group. *p* value stands for Student's *t*-test between the nondilated and dilated groups.



The results of this computational analysis study were intriguing. In particular, stress shows the most significant differences between the aortic dilated and nondilated groups, which also implies that aortic stress may be a good predictor of aorta response to aortic dilation. From a pathophysiologic perspective, these findings are plausible given that stress more accurately reflects the functional status of the aortas as compared with aortic diameter. From a clinical perspective, most published criteria for ascending aortic dilation have focused on aortic diameter. However, the results of this study suggest that aortic stress is also and may be more helpful in identifying the aortic tissue status, therefore informing the decision of clinical intervention than relying on aortic diameter alone.

This study is a preliminary work applying biomechanical modeling in the investigations of ascending aortic dilation. Several improvements can enhance our work in the future for better accuracy and applicability: (1) sample size needs to be enlarged to validate the findings in this study; (2) the addition of aortic valve mechanics to the model may improve accuracy with regard to timing of valve opening and closure and allow incorporation of aortic valve regurgitation; (3) in vivo measurements of tissue properties will be very desirable for improved accuracy of our models; (4) the uniqueness of tissue parameters is difficult to be guaranteed given the limited experimental data; (5) residual stress/strain which is defined as the initial stress/strain in the solid under zero-load state should be considered as nonzero for more accurate simulation results; (6) the addition of patient-specific aortic wall thickness will improve the accuracy of our simulations.

## 5. Conclusion

In this study, patient-specific biomechanical computational models of 20 TOF patients undergoing ascending aortic dilation of different levels were constructed by comparing numerical volume of the ascending aortic segment with the ones obtained from CT scans. Simulation results demonstrated that aortic stress, flow rates, and displacement were able to differentiate the patients with dilated ascending aorta from the nondilated ones, which implied that these factors may be highly correlated with aortic dilation and even the cause of dilation. These findings provide a basis for future studies aimed at validating these results in larger groups of patients and further refinements of the computational modeling to improve its accuracy.

## Data Availability

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

## Conflicts of Interest

No competing interests to declare.

## Authors' Contributions

These authors take responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation. Yunfei Ling shares the same authorship with the first author.

## Acknowledgments

This work was partially supported by the 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (Grant ZYJC18010) and the Sichuan Science and Technology Program (Grant 2019YFS0147).

## References

- [1] Centers for Disease Control and Prevention, "Improved national prevalence estimates for 18 selected major birth defects—United States, 1999–2001," *Morbidity and Mortality Weekly Report*, vol. 54, no. 51, pp. 1301–1305, 2006.
- [2] A. A. Alghamdi and G. S. Van Arsdell, "Replacement of aortic root and ascending aorta in adult congenital heart disease," *Expert Review of Cardiovascular Therapy*, vol. 5, no. 6, pp. 1087–1094, 2014.
- [3] S. Chauvaud, A. Serraf, S. Mihaileanu et al., "Ventricular septal defect associated with aortic valve incompetence: results of two surgical managements," *The Annals of Thoracic Surgery*, vol. 49, no. 6, pp. 875–880, 1990.
- [4] M. G. Kaya, E. Mavili, O. Dogdu, A. Dogan, T. Inanc, and A. Baykan, "Acute type A aortic dissection in a patient with ventricular septal defect," *Cardiovascular Revascularization Medicine*, vol. 11, no. 3, pp. 186–188, 2010.
- [5] S. Leontyev, M. Misfeld, and F. W. Mohr, "Aneurysms of the ascending aorta and aortic arch," *Chirurg*, vol. 85, no. 9, pp. 758–766, 2014.
- [6] L. A. Pape, M. Awais, E. M. Woznicki et al., "Presentation, diagnosis, and outcomes of acute aortic dissection: 17-year trends from the International Registry of Acute Aortic Dissection," *Journal of the American College of Cardiology*, vol. 66, no. 4, pp. 350–358, 2015.
- [7] W. Wang, W. Duan, Y. Xue et al., "Clinical features of acute aortic dissection from the Registry of Aortic Dissection in China," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 148, no. 6, pp. 2995–3000, 2014.
- [8] T. Krüger, E. Weigang, I. Hoffmann, M. Blettner, and H. Aebert, "Cerebral protection during surgery for acute aortic dissection type A: results of the German Registry for Acute Aortic Dissection Type A (GERAADA)," *Circulation*, vol. 124, no. 4, pp. 434–443, 2011.
- [9] P. G. Hagan, C. A. Nienaber, E. M. Isselbacher et al., "The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease," *JAMA*, vol. 283, no. 7, pp. 897–903, 2000.
- [10] M. A. Coady, J. A. Rizzo, G. L. Hammond et al., "What is the appropriate size criterion for resection of thoracic aortic aneurysms?," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 113, no. 3, pp. 476–491, 1997.
- [11] L. F. Hiratzka, G. L. Bakris, J. A. Beckman et al., "2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM Guidelines for the Diagnosis and Management of Patients

- With Thoracic Aortic Disease,” *Journal of the American College of Cardiology*, vol. 55, no. 14, pp. e27–129, 2010.
- [12] R. R. Davies, A. Gallo, M. A. Coady et al., “Novel measurement of relative aortic size predicts rupture of thoracic aortic aneurysms,” *The Annals of Thoracic Surgery*, vol. 81, no. 1, pp. 169–177, 2006.
  - [13] R. R. Davies, L. J. Goldstein, M. A. Coady et al., “Yearly rupture or dissection rates for thoracic aortic aneurysms: simple prediction based on size,” *The Annals of Thoracic Surgery*, vol. 73, no. 1, pp. 17–28, 2002.
  - [14] E. K. Kim, S. H. Choi, K. Sung et al., “Aortic diameter predicts acute type A aortic dissection in patients with Marfan syndrome but not in patients without Marfan syndrome,” *The Journal of Thoracic and Cardiovascular Surgery*, vol. 147, no. 5, pp. 1505–1510, 2014.
  - [15] L. A. Pape, T. T. Tsai, E. M. Isselbacher et al., “Aortic diameter  $\geq 5.5$  cm is not a good predictor of type A aortic Dissection,” *Circulation*, vol. 116, no. 10, pp. 1120–1127, 2007.
  - [16] M. H. Guo, J. J. Appoo, R. Saczkowski et al., “Association of mortality and acute aortic events with ascending aortic aneurysm. A systematic review and meta-analysis,” *JAMA Network Open*, vol. 1, no. 4, article e181281, 2018.
  - [17] E. Angeloni, A. Vitaterna, M. Pirelli, and S. Refice, “Effects of statin therapy on ascending aorta aneurysms growth: a propensity-matched analysis,” *International Journal of Cardiology*, vol. 191, pp. 52–55, 2015.
  - [18] M. Gagné-Loranger, É. Dumont, P. Voisine, S. Mohammadi, and F. Dagenais, “Natural history of 40-50 mm root/ascending aortic aneurysms in the current era of dedicated thoracic aortic clinics,” *European Journal of Cardio-Thoracic Surgery*, vol. 50, no. 3, pp. 562–566, 2016.
  - [19] B. W. Andrus, D. J. O’Rourke, L. J. Dacey, and R. T. Palac, “Stability of ascending aortic dilatation following aortic valve replacement,” *Circulation*, vol. 108, no. 90101, Supplement 1, pp. 295II–295299, 2003.
  - [20] C. D. Etz, S. Zoli, R. Brenner et al., “When to operate on the bicuspid valve patient with a modestly dilated ascending aorta,” *The Annals of Thoracic Surgery*, vol. 90, no. 6, pp. 1884–1892, 2010.
  - [21] D. Detaint, H. I. Michelena, V. T. Nkomo, A. Vahanian, G. Jondeau, and M. E. Sarano, “Aortic dilatation patterns and rates in adults with bicuspid aortic valves: a comparative study with Marfan syndrome and degenerative aortopathy,” *Heart*, vol. 100, no. 2, pp. 126–134, 2013.
  - [22] J. B. Kim, M. Spotnitz, M. E. Lindsay, T. E. MacGillivray, E. M. Isselbacher, and T. M. Sundt, “Risk of aortic dissection in the moderately dilated ascending aorta,” *Journal of the American College of Cardiology*, vol. 68, no. 11, pp. 1209–1219, 2016.
  - [23] S. Geisbusch, A. Stefanovic, D. Schray et al., “A prospective study of growth and rupture risk of small-to-moderate size ascending aortic aneurysms,” *The Journal of Thoracic and Cardiovascular Surgery*, vol. 147, no. 1, pp. 68–74, 2014.
  - [24] G. D. Aquaro, A. B. Vangosa, P. Toia et al., “Aortic elasticity indices by magnetic resonance predict progression of ascending aorta dilation,” *European Radiology*, vol. 27, no. 4, pp. 1395–1403, 2017.
  - [25] M. Alreshidan, N. Shahmansouri, J. Chung et al., “Obtaining the biomechanical behavior of ascending aortic aneurysm via the use of novel speckle tracking echocardiography,” *The Journal of Thoracic and Cardiovascular Surgery*, vol. 153, no. 4, pp. 781–788, 2017.
  - [26] S. Pasta, V. Agnese, M. Di Giuseppe et al., “In vivo strain analysis of dilated ascending thoracic aorta by ECG-gated CT angiographic imaging,” *Annals of Biomedical Engineering*, vol. 45, no. 12, pp. 2911–2920, 2017.
  - [27] S. Farzaneh, O. Trabelsi, B. Chavent, and S. Avril, “Identifying local arterial stiffness to assess the risk of rupture of ascending thoracic aortic aneurysms,” *Annals of Biomedical Engineering*, vol. 47, no. 4, pp. 1038–1050, 2019.
  - [28] S. Farzaneh, O. Trabelsi, and S. Avril, “Inverse identification of local stiffness across ascending thoracic aortic aneurysms,” *Biomechanics and Modeling in Mechanobiology*, vol. 18, no. 1, pp. 137–153, 2019.
  - [29] P. Youssefi, A. Gomez, C. Arthurs, R. Sharma, M. Jahangiri, and F. C. Alberto, “Impact of patient-specific inflow velocity profile on hemodynamics of the thoracic aorta,” *Journal of Biomechanical Engineering*, vol. 140, no. 1, 2018.
  - [30] F. Condemni, S. Campisi, M. Viallon et al., “Fluid- and biomechanical analysis of ascending thoracic aorta aneurysm with concomitant aortic insufficiency,” *Annals of Biomedical Engineering*, vol. 45, no. 12, pp. 2921–2932, 2017.
  - [31] F. Condemni, S. Campisi, M. Viallon, P. Croisille, J. F. Fuzelier, and S. Avril, “Ascending thoracic aorta aneurysm repair induces positive hemodynamic outcomes in a patient with unchanged bicuspid aortic valve,” *Journal of Biomechanics*, vol. 81, pp. 145–148, 2018.
  - [32] H. R. Underhill, T. S. Hatsukami, Z. A. Fayad, V. Fuster, and C. Yuan, “MRI of carotid atherosclerosis: clinical implications and future directions,” *Nature Reviews Cardiology*, vol. 7, no. 3, pp. 165–173, 2010.
  - [33] J. Chung, K. Lachapelle, E. Wener et al., “Energy loss, a novel biomechanical parameter, correlates with aortic aneurysm size and histopathologic findings,” *The Journal of Thoracic and Cardiovascular Surgery*, vol. 148, no. 3, pp. 1082–1089, 2014.
  - [34] T. J. Schlattmann and A. E. Becker, “Histologic changes in the normal aging aorta: implications for dissecting aortic aneurysm,” *The American Journal of Cardiology*, vol. 39, no. 1, pp. 13–20, 1977.
  - [35] C. Pisano, E. Maresi, C. R. Balistreri et al., “Histological and genetic studies in patients with bicuspid aortic valve and ascending aorta complications,” *Interactive CardioVascular and Thoracic Surgery*, vol. 14, no. 3, pp. 300–306, 2012.
  - [36] A. Curtis, T. Smith, B. Ziganshin, and J. Elefteriades, “The mystery of the Z-score,” *Aorta*, vol. 4, no. 4, pp. 124–130, 2016.
  - [37] D. Tang, C. Yang, P. J. del Nido et al., “Mechanical stress is associated with right ventricular response to pulmonary valve replacement in patients with repaired tetralogy of Fallot,” *The Journal of Thoracic and Cardiovascular Surgery*, vol. 151, no. 3, pp. 687–694.e3, 2016.
  - [38] C. Yang, R. G. Bach, J. Zheng et al., “In vivo IVUS-based 3-D fluid-structure interaction models with cyclic bending and anisotropic vessel properties for human atherosclerotic coronary plaque mechanical analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 10, pp. 2420–2428, 2009.
  - [39] L. Wang, D. Tang, A. Maehara et al., “Fluid-structure interaction models based on patient-specific IVUS at baseline and follow-up for prediction of coronary plaque progression by morphological and biomechanical factors: a preliminary study,” *Journal of Biomechanics*, vol. 68, pp. 43–50, 2018.
  - [40] A. Jarrahi, A. Karimi, M. Navidbakhsh, and H. Ahmadi, “Experimental/numerical study to assess mechanical properties of healthy and Marfan syndrome ascending thoracic aorta

under axial and circumferential loading,” *Materials Technology*, vol. 31, no. 5, pp. 1–8, 2015.

- [41] G. de Simone, R. B. Devereux, S. Chien, M. H. Alderman, S. A. Atlas, and J. H. Laragh, “Relation of blood viscosity to demographic and physiologic variables and to cardiovascular risk factors in apparently normal adults,” *Circulation*, vol. 81, no. 1, pp. 107–117, 1990.
- [42] D. Tang, H. Zuo, C. Yang et al., “Comparison of right ventricle morphological and mechanical characteristics for healthy and patients with tetralogy of Fallot: an in vivo MRI-based modeling study,” *Molecular & Cellular Biomechanics*, vol. 14, no. 3, pp. 137–151, 2017.

## Research Article

# DE-CNN: An Improved Identity Recognition Algorithm Based on the Emotional Electroencephalography

Yingdong Wang , Qingfeng Wu , Chen Wang , and Qunsheng Ruan

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

Correspondence should be addressed to Qingfeng Wu; [qfwu@xmu.edu.cn](mailto:qfwu@xmu.edu.cn)

Received 7 December 2019; Accepted 5 February 2020; Published 8 August 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Yingdong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the past few decades, identification recognition based on electroencephalography (EEG) has received extensive attention to resolve the security problems of conventional biometric systems. In the present study, a novel EEG-based identification system with different entropy and a continuous convolution neural network (CNN) classifier is proposed. The performance of the proposed method is experimentally evaluated through the emotional EEG data. The conducted experiment shows that the proposed method approaches the stunning accuracy (ACC) of 99.7% on average and can rapidly train and update the DE-CNN model. Then, the effects of different emotions and the impact of different time intervals on the identification performance are investigated. Obtained results show that different emotions affect the identification accuracy, where the negative and neutral mood EEG has a better robustness than positive emotions. For a video signal as the EEG stimulant, it is found that the proposed method with 0–75 Hz is more robust than a single band, while the 15–32 Hz band presents overfitting and reduces the accuracy of the cross-emotion test. It is concluded that time interval reduces the accuracy and the 15–32 Hz band has the best compatibility in terms of the attenuation.

## 1. Introduction

Identity systems are essential for security systems in many applications including payment systems, the Internet of Things (IoT), and health devices to protect personal data by verifying the identity of people. Moreover, these systems are often used in the process of human and machine interfaces. Conventional verification methods include setting a password or using a smart card for verification. However, these methods suffer from lots of problems, including forgetting or even stealing the password. Subsequently, the verification method through the password has been gradually replaced by the biometric verification system in the last few years. The biometric verification system authenticates through biometric information. A biological information system includes preprocessing of physiological signals, machine learning, and pattern recognition. Then, the system compares these features with the database. Physiological and behavioural biometrics include the fingerprint, face pattern, gait model, and electrocardiography (ECG).

Although these biological systems are popular now, there are still many unresolved problems. More specifically, identification systems based on the face [1], iris [2], sound [3], and the fingerprint [4] recognition can be deceived by high-quality images, sound, and feature extractions, respectively. Moreover, since fingerprints abundantly remain on ordinary surfaces, they can be easily abused by malefactors.

ID recognition method based on biometric, electroencephalography (EEG) provides more choices for identification [5]. Many scholars have focused on the EEG wave because it consists of invisible and untouchable electrical neural oscillations. Therefore, the EEG wave is highly attack-resilient and cannot easily be deceived. Moreover, the EEG is affected by the style of thinking, mood, and even the atmosphere so that it is a unique wave [6]. Benefiting from the deep learning (DL) techniques in various applications such as insomnia diagnosis, seizure detection, sleep studies, emotional recognition, and brain-computer interface (BCI) [7–10], the accuracy of the EEG-based ID recognition has



been remarkably improved. In order to ensure that EEG can be used for the identification, most studies focus on the design of experimental paradigms. For example, eye closing [11], visual stimulation [12], and multiple mental tasks [13] have been investigated and mentioned in this regard. Wu et al. [14] proposed the eye blinking and self- or non-self-rapid serial visual presentation, and then, they extracted features from the EEG and the eye blink. Finally, they applied a fusion technology with two features to obtain the final high estimation score. The average accuracy rate of the study cases reached 97.60%, the false acceptance rate (FAR) was 2.71%, and the false rejection rate (FRR) was 2.09%. Kang et al. [15] used an open-access motion-image EEG database for identification. They conducted the network analysis based on the phase synchronization and extracted 10 single-channel features and 10 multichannel features, and then, they calculated the Euclidean distance between each possible pair of row vectors in the training and validation data matrices. Finally, they found thresholds of different features and gained equal error rate (EER) and FRR when the FAR is set to 1%. They found that the EER for the Romberg test, eyes open (REO), and the Romberg test, eyes closed (REC), are 0.73% and 1.80%, respectively. Moreover, they showed that the FRR with 1% FAR for REO and REC is 1.10% and 2.20%, respectively. Sun et al. [16] used the biggest motion imagination EEG dataset. They applied the 1D-convolutional long short-term memory neural networks to identify 109 subjects, where the best result was about 0.0041 in terms of EER. Furthermore, Moctezuma et al. [17] utilized the imaginary speech EEG for 27 subjects and performed 33 repetitions of five imaginary words in Spanish and gained an accuracy of 97%.

Although reviewed models in Table 1 yield high precision results, it is a challenge for the brain to reproduce the same EEG. For example, Wu et al. [14] showed that as time passes by, people gradually became accustomed to the faces of strangers so that reproducing the original visual impact from the recorded data becomes difficult [19].

So, the question is how the specified content impact identity? Reviewing the literature indicates that few studies have been performed on the identity authentication of the EEG based on different stimuli. Zhang et al. [9] used the emotional EEG for the identification and found that the emotion has no impact on the identification of 12 s EEG. However, the method robustness for different emotions was not proved. The present study intended to select the SEED dataset, a public emotional EEG dataset, to eliminate the impact of different contents on the brain by watching long videos. It is believed that only during watching the video, underlying characteristics and rhythms of the individual can be discovered.

Many studies have investigated the acquisition of rhythm features from the EEG. Kang et al. [15] extracted 10 multichannel features and 10 single-channel features, including seven spectra, and three nonlinearities, based on the phase synchronization network analysis for the subject identification. The performed analysis showed that Maxlyp has outstanding results compared with other features. Shi et al. [20] proposed the differential entropy (DE) of the EEG-

based alert estimation and applied the proposed method to measure the complexity of EEG signals. Further studies [21, 22] showed that DE is a suitable scheme for emotional decoding. Moreover, Moctezuma et al. [17] applied the power spectral density (PSD) and autoregressive (AR) model coefficients for classification and obtained 99.76% accuracy in the studied cases. For the deep learning method, [18] applied the emotional EEG independent of any conventional features to predict the ID and achieved an accuracy of 94%. Therefore, the remaining question is “how to select the most suitable features and classification method for the model.”

In order to solve the problem, the computational expenses of features are initially compared. In this regard, two features are selected for the classification. Then, a novel preprocesses algorithm is proposed and the corresponding data is cut into small clips so that four-band data is obtained and each band's features are calculated. Finally, the model is designed with all features and the identification starts. In summary, the SEED dataset is utilized in the present study, which shields the correlation between the EEG and the specific content to find stable rhythm and characteristics in the EEG. Moreover, a novel method is proposed, which gains the algorithm accuracy higher than that of the algorithms. The present study contains three main highlights as the following:

- (i) Compared to other algorithms, our algorithm takes less computational expense, while it has higher accuracy compared with conventional algorithms
- (ii) The proposed method is better compatible with different emotions than other methods
- (iii) For the first time, the factor of the EEG authentication interval was considered, which proved to be a strong attenuation with our algorithm

The rest of the article is organized as follows. The literature review for designing extraction features and deep learning-based EEG biometrics are presented in Section 2. Moreover, the detailed information and methodology of the proposed EEG biometric identification system are discussed in Section 3. Then experimental results and evaluations are presented in Section 4. Discussions on the performance and the corresponding potentials are provided in Section 5, which is followed by conclusions and a brief description of future works in the final section.

## 2. Related Work

*2.1. Emotion and EEG.* Various psychophysiology studies have demonstrated the correlations between human emotion and EEG signals [23, 24]. Martini et al. [25] noticed an increase in P300 in the late positive potential and an increase in gamma activity during viewing unpleasant pictures when the comparison is made with neutral pictures. Zhang and Lu [26] showed that energy from the beta and gamma bands increases in positive emotions and decreases in neutral and negative emotions. Moreover, neuroscience studies [27, 28] showed that EEG alpha bands reflect attention processing, while beta bands reflect emotional and cognitive processing



TABLE 1: Summary of the literature review.

Papers	EEG content	Method	Time (s)	EER	AAC
[14]	Eye blinking and self- or non-self-rapid serial visual presentation	Machine learning	3	—	0.9076
[15]	Relax and eye-closed	Machine learning	60	0.0073	0.9893
[16]	MI-EEG	1DCNN-LSTM	1	0.0041	0.995
[17]	Imaginary speech	Deep learning	Four words	—	0.97
[18]	Relax and eye-closed	Attention-RNN	1/128	—	0.998
[9]	Emotion video (different stimulant)	2DCNN + LSTM	12	—	1

in the brain. Considering the latest improvements in the emotional classification, the ACC has reached about 91% for recognizing a single person and 86% for recognizing different persons [29, 30]. Therefore, different emotions affect the EEG so that the issue of emotional robustness should be addressed prior to applying the EEG as an ineffective and reliable authentication method. According to [31], the recorded EEG rhythms can be categorized into five different rhythms in accordance with their frequency ranges. These rhythms, which are presented in Table 2, are as follows: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta (15–32 Hz), gamma (32–40 Hz), and the other bands (40–75 Hz). Delta wave always appears when people are in deep sleep. Moreover, the theta wave is encountered in early sleep stages and drowsiness. Alpha and beta rhythm are the typical rhythms during the relaxed state with closed eyes and the prominent rhythm during stressful situations, respectively. Finally, gamma rhythm is always involved in higher-order functions of the brain such as the feature binding of a perceived image. Therefore, the most suitable band can be explored for the identification.

**2.2. Comparing Different Features.** Kang et al. [15] demonstrated that three types of nonlinear EEG features, including the maximum Lyapunov exponent (Maxlyp), sample entropy, and the permutation entropy, have a higher impact on EEG-based biometrics than conventional spectral features. Application of the Maxlyp scheme can reach the best result at EER of 0.043 so that many researchers applied the different entropy methods on the EEG classification [32, 33]. Moreover, PSD is the most common feature in the EEG data. In the next section, it is intended to compare these three features from the time-consuming point of view.

**2.2.1. Maximum Lyapunov Exponent.** Studies show that nonlinear methods, which mainly focus on the detection of characteristics of dynamic changes in a time series, are useful for clinical and scientific EEG applications [34, 35]. In the Maxlyp, the single-channel time series data  $A\{x_1, x_2, x_3, \dots, x_N\}$  are considered, where  $N$  denotes the data length. In order to calculate the maximum Lyapunov exponent, the time series must be embedded into a  $D$ -dimensional space  $x_j = [x_j, x_{j-t}, \dots, x_{j+(D-1)t}]$ .

The Lyapunov exponent characterizes the inherent instability of a time series by quantifying the average rate at which nearby trajectories in the phase space diverge or converge [22, 36]. This instability is based on the sensitive

dependence on the initial conditions. For two initial points in the phase close to each other space  $X_{j1}$  and  $X_{j2}$ ,  $\delta_0$  is defined as the distance between points in the phase space  $\|X_{j1} - X_{j2}\| \sigma_0 \ll 1$ , where the distance varies to  $\sigma_{\Delta n}$  after a certain time  $\Delta n$ .  $\|X_{j1+\Delta n} - X_{j2+\Delta n}\| = \sigma_{\Delta n}$ . The correlation between  $\sigma_0$  and  $\sigma_{\Delta n}$  can be expressed in terms of an exponential function as follows:

$$\delta_{\Delta n} \cong \delta_0 \cdot e^{\lambda \Delta n} (\Delta n \gg 1, \delta_{\Delta n} \ll 1), \quad (1)$$

$$\delta_{\Delta n} = \lim_{\Delta n \rightarrow \infty} \delta_0 \cdot e^{\lambda \Delta n}. \quad (2)$$

The constant term in the exponent  $\lambda$  describes the rate of change and can be expressed in the following form:

$$\lambda = \lim_{\Delta n \rightarrow \infty} \frac{1}{\Delta n} \ln \frac{\delta_{\Delta n}}{\delta_0}. \quad (3)$$

Every  $X_j$  has one  $\lambda$ , where the maximum  $\lambda$  is the maximum Lyapunov exponent.

**2.2.2. Differential Entropy (DE).** Differential entropy scheme is applied for EEG-based vigilance estimation to measure the complexity of EEG signals [20, 37]. The DE scheme is mathematically expressed in the following form:

$$h(x) = \int_x f(x) \log(f(x)) dx, \quad (4)$$

where  $X$  is a random variable and  $f(x)$  denotes the probability density function of  $X$ . For series with the Gauss distribution  $(N(\mu, \sigma^2))$ , the corresponding differential entropy can be expressed as

$$\begin{aligned} h(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \frac{1}{2} \log(2\pi e \sigma^2). \end{aligned} \quad (5)$$

**2.2.3. PSD Subheadings.** The periodic method PSD estimation is employed to simply find the discrete-time Fourier transform and scale the amplitude value of the result. In this scheme,  $L$  is defined as the signal  $x(n)$  length and  $F$  is the sampling frequency, respectively. In fact, the PSD value should be calculated at point  $N$ . The periodic estimation of the PSD method is expressed as follows:

TABLE 2: The range of applications for different waves.

Rhythm	Frequency domain	Brain states	Awareness degree
Delta	0.5–4 Hz	Deep dreamless sleep	Lower
Theta	4–8 Hz	Creative, intuitive, drowsy	Low
Alpha	8–15 Hz	Relax, not drowsy, tranquil, conscious, focus	Medium
Beta	15–32 Hz	Thinking, aware of self-surroundings	High
Gamma	32–40 Hz	Thinking, integrated thought	Very high

$$\text{psd} = \frac{1}{LF} \left( \sum_{n=0}^{L-1} \frac{x(n)e^{-2\pi j f(n)}}{F} \right)^2. \quad (6)$$

Table 3 shows a comparison of the three features and indicates that although the Maxlyp is the best feature, it consumes 12.591 s for calculation, where such high time consumption cannot be justified. Meanwhile, the complexity of the PSD and ED schemes is lower than that of the Maxlyp scheme.

**2.3. Normalization.** The EEG signal with  $N$  Channels is applied as the input for training the proposal neural network. When all the clips of DE values are obtained, the normalization overtime is required for each channel. The normalization is conducted as follows:

$$I_{i,j} = \left( \frac{\text{DE}_{i,j} - \left( \sum_{j=1}^N \text{DE}_{i,j} / N \right)}{\sigma_i} \right), \quad (7)$$

where  $i, j$ , and  $\sigma$  refer to the signal position, channel position, and the standard deviation of the DE at one position of the DE sequence, respectively.

**2.4. Changing Data from 1D to 2D.** The EEG-based BCI system uses a wearable headset with multiple electrodes to capture EEG signals. The International 10–20 System is an internationally recognized method of describing and applying the location of the scalp electrode and the underlying area of the cerebral cortex. It should be indicated that “10” and “20” numbers indicate that the actual distance between the adjacent electrodes is either 10% or 20% of the total front-back or right-left distance of the skull [38]. Although all positions in the data are meaningful, the sample EEG data is still a sequence after DE features are computed so that they are organized from the left to the right. In order to obtain the spatial features, data is converted into two-dimensional data in the manner of Figure 1. When there is no signal in the matrix, it will be replaced with zero.

### 3. Materials and Methods

**3.1. Data.** To develop the algorithm for EEG-based biometrics, the SEED<sup>1</sup> database [26] is utilized in this section, which is the largest publicly available database for emotional EEG. In this datasheet, 62-channel EEG signals are recorded from fifteen persons when they are watching fifteen 4-minute emotional video clips. EEG data of each person is recorded three times in different weeks, where each time

TABLE 3: The time of computing three features.

Feature name	Maxlyp	ED	PSD
Time (s)	12.591	0.00064	0.0019

contains fifteen sessions. Table 4 presents the distribution of labels per person for three kinds of emotions, including the neutral, positive, and negative emotions. The downsampling on data is performed with the frequency of 200 Hz, where the 0–75 Hz filter is applied.

**3.1.1. System Overview.** Figure 2 shows an overview of the proposed EEG-based identification system consisting of the training period and the identification phase. In this system, the EEG features of all users are learned and stored in the DE-CNN model in the training period. It should be indicated that the EEG data, either in the training phase or in the identification phase, are preprocessed. The conducted pre-processing consists of segmenting data into 1000-sample length, computing the DE feature, and normalizing prior to feeding into the CNN model. The identification phase is the result of three CNN layers and a fully connected (FC) network with a Softmax activation function. In the rest of the methodology section, data segmentation, computing DE features, normalization, and multichannel CNN model will be described in detail.

**3.1.2. Preparing the Dataset.** In the SEED dataset, the emotional EEG of each person is recorded three times. In the present study, two of three records are considered as the research data, while the last record will be tested for other purposes. Table 4 indicates that there are three kinds of emotional states in the data, and each affection has the same number five of EEG trails. Wilaiprasitporn [39] used 12 s long emotional EEG as the test data and proved that the emotion does not affect the result so that the EEG data can be used for the identification. It is intended to explore a method to achieve less identification time. In this regard, the identification time is set to 1000 samples (five-second  $\times$  200) to reflect the thinking rhythm. Thus, each EEG trial is simply segmented into 48 subsamples so that 720 subsamples (48 subsamples  $\times$  15 trials or clips) are obtained for each participant for one record. In summary, experiment labels are participant IDs. The data and labels of the present study can be described as follows:

- (1) Data:  $2 \times 15 \times 720 \times (1000 \times 62)$
- (2) Label:  $2 \times 15 \times 720 \times 1$

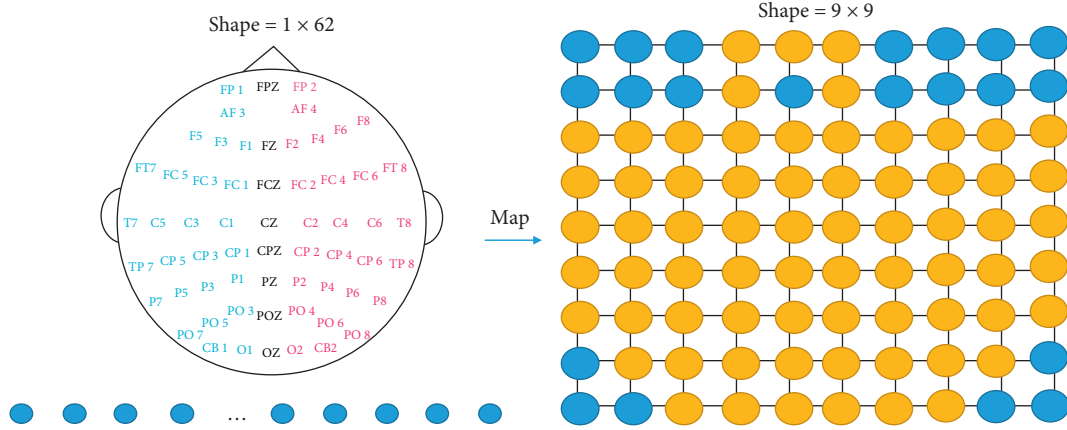


FIGURE 1: Schematic diagram of converting one-dimensional data into two-dimensional data.

TABLE 4: Number of emotions per person.

Label	Positive	Neutral	Negative	Totally
Number	5	5	5	15

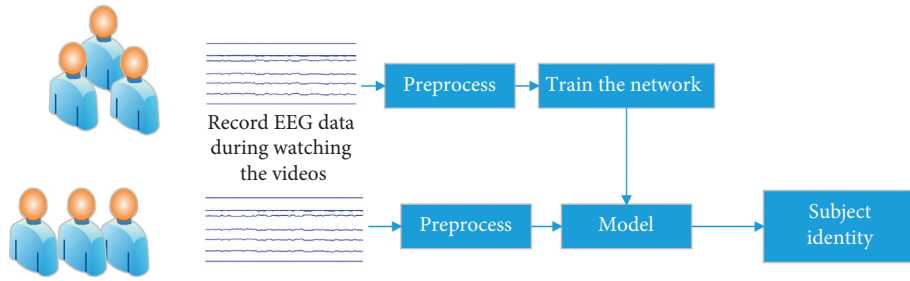


FIGURE 2: Overview of EEG-based identification.

**3.1.3. Preprocessing.** Figure 3 shows that the main flow of the preprocessing has five steps. In Step 1, in order to obtain the fine particle characteristics, the data is divided into five one-second sequences. Moreover, the features are extracted separately from sequences for finally being merged. In Step 2, decompose the EEG signal into four frequency bands ( $\theta, \alpha, \beta, \gamma$ ) by the Butterworth filter, which has been proved to be more useful than the whole data in many kinds of research [18]. After decomposition, one clip EEG data is converted from  $5 \times 200 \times 62$  to  $5 \times 200 \times 4 \times 62$ . In Step 3, DE is an excellent feature and it is the output of a chaotic degree of sequence. Therefore, after computing every bands' channels DE features, the data is converted to  $5 \times 1 \times 4 \times 62$ . In Step 4, for the purpose of preserving spatial information among multiple adjacent channels, the one-dimension DE feature vector of 62 lengths to the 2D plane ( $9 \times 9$ ) is transformed according to the electrode distribution map. So, the data size is  $5 \times 4 \times 9 \times 9$ . In Step 5, in order to speed up the convergence of the model, normalization is performed at the end. This preserves all the information to the utmost.

**3.1.4. Convolution Neural Network.** Figure 4 shows that a continuous convolution neural network with four convolution layers is used to extract features from the input cube. Moreover, a fully connected layer with dropout operation is

added for the feature fusion and the Softmax layer is used for final classification. It should be indicated that there is no pooling layer between two adjacent convolution layers. In each convolution layer, zero-padding is applied to prevent information missing at the edge of the cube. More specifically, in the first three convolution layers, the kernel size is set to  $4 \times 4$  and the stride is set to one. After the convolution operation, the RELU activation function is added to endow the model with nonlinear feature transformation capability. The first convolution layer with 64 feature maps is initiated and the feature maps in the following two convolution layers are doubled. Therefore, there are 128 and 256 feature maps in the second and third layers. In order to fuse different feature maps and reduce the computational cost, a one-by-one convolution layer with 64 feature maps is added. After these 4 continuous convolution layers, a fully connected layer is added to map the  $649 \times 9$  feature maps into a final feature vector  $f \in 1024$ . Then, the following Softmax layer receives  $f$  to predict the human emotional state.

## 4. Results and Discussion

**4.1. Comparison.** Table 5 shows the comparison results of the proposed method and the state-of-the-art EEG-based identification methods. Among the mentioned methods, two

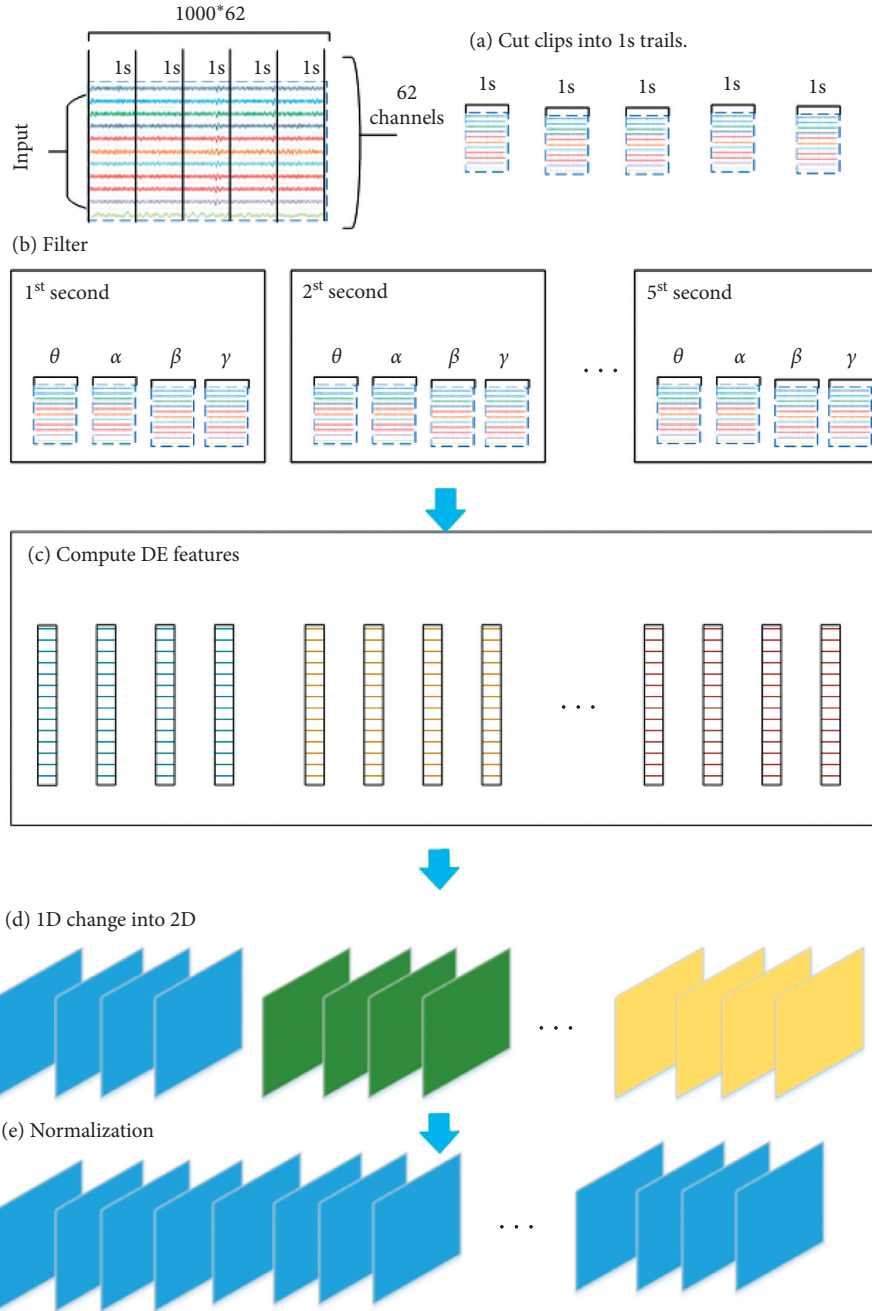


FIGURE 3: Main steps of the flow preprocessing: cutting, decomposing, computing, mapping, and normalization.

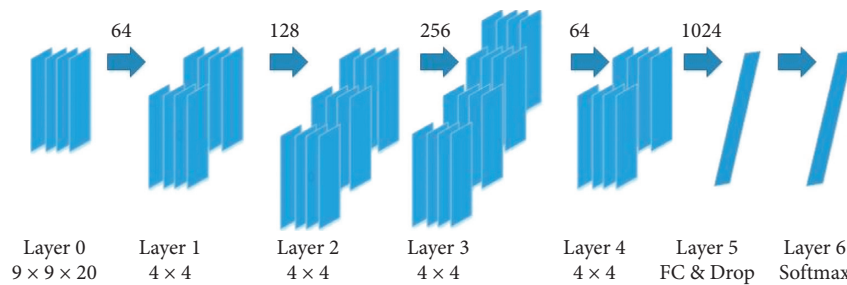


FIGURE 4: The classifier model for four layers. The number of  $4 \times 4$  kernel for each layer is 64, 128, 256, and 64.



TABLE 5: Comparison of the proposed method with some state-of-the-art EEG-based identification methods.

Name	Train time (s)	Rank-1	Parameters	EER
CNN + LSTM [9]	278	$0.98 \pm 0.002$	32361999	0.0121
D-LSTM [16]	60	$0.76 \pm 0.04$	1060943	0.011
DE-CNN	2	$0.997 \pm 0.0028$	7918095	0.002
PSD-CNN	2	$0.934 \pm 0.0034$	7918095	0.021

deep learning methods exist. The first one is introduced in [9], which applies CNN + STML to classify five kinds of motor imagination and the accuracy of this method can reach 99%. Moreover, in [40], the authors applied the same method to explore the emotional effects on the identification with the same method in the DEAP dataset. The obtained result can reach an accuracy of 95% for 12 s. In order to compare this method with the proposed method of the present study, parameters are adjusted to suit the dataset of this study. In CNN + LSTM, three layers of 2D CNNs with  $3 \times 3$  kernels are used. The number of filters starts with 128 in the first layer and continues with 64 and 32, respectively. It should be indicated that ReLu nonlinearity is used. Batch normalization and dropout are applied after every convolution layer. For the recurrent layers, 2 layers with 32 and 16 recurrent units are used, respectively. Moreover, recurrent dropout is applied. The dropout rates in each part of the model are fixed at 0.5. The RMSprop optimizer is used with a learning rate of 0.0005 and a batch size of 30.

Furthermore, 1D-convolution [16] applied the 1D-convolution long short-term memory neural network for the EEG-based user identification. The same parameters are applied in the present study for the proposed dataset. It is found that the model is different to converge. Therefore, three layers of 1D-convolution are used and the kernel sizes are 128, 256, and 512 with a dropout. Then, the results feed for the next two layers of LSTM with a kernel of 192. Finally, a dropout and a fully connected network with Softmax activation are applied to predict the probability of ID. The features of the last two methods are selected manually. Moreover, the same framework algorithm is compared based on the PSD. In the preprocessing, the average of the PSD is used in one second instead of DE. However, the other parts are the same as the DE-CNN.

In all experiments, the training, validation, and testing results are obtained by 10-fold cross-validation. It should be indicated that 90% of data are used as training and the left are used as the test dataset. The train time is the total time of computing one epoch at NVIDIA GEFDRCE RTX 2080ti. Table 5 shows that the 1D-LSTM cannot find enough information for the identification or overfit for the test dataset. Although CNN + LSTM has high accuracy, equal error rate (EER), it needs 278 seconds to train 32,361,999 parameters in one epoch. When the system adds a new user, it may take much time to update. The method proposed in the present study obtains the higher rank-1 of 0.997 and the EER of 0.00184 and it suits the high-demand security systems. Although the PSD is one of the most common used features, the result for PSD-CNN is lower 0.93 than that of the DE-

CNN. The code for all comparison algorithms can be found in <https://github.com/heibaipei/DE-CNN>.

*4.2. Comparison of Affective EEG-Based ID among Four Bands.* In order to explore the best band to reduce noise, only two methods within four bands are compared in a positive mood. Figure 5 shows the obtained results. All parameters are the same as the above-mentioned parameters. The number of training epoch is 50. Rank-1 with the  $\beta$  band and the 4–40 Hz band are a little higher than other bands, while the 4–40 Hz band is lower than the 14–31 Hz band. It is explicable that the beta band is highly correlated with attention and alertness. Moreover, it should be indicated that the wider band may have more noise.

It is found that the obtained results of the two methods in 14–31 Hz and 4–40 Hz have little difference in the final accuracy, while the processes of the training have a significant difference. The results of the DE-CNN in four bands are smoother and faster than the CNN-LSTM. Figure 6 shows the process of training. It should be indicated that beta waves work best in identity authentication consistent with the content of Table 6.

*4.3. Comparison of Affective EEG-Based ID among Three Affective States.* Many people have questioned the performance of the EEG for identity authentication. Moreover, the stability of the EEG-based method has attracted many scholars. Since different moods make different EEGs, it is of significant importance to train a robust model. Table 7 shows that three emotions have little impact on the identification with DE-CNN, where the currency and EER approach 0.99 and 0.001, respectively, while applying the CNN-LSTM to the neutral emotion yields the worst result, where the corresponding EER is only 0.06. Table 7 shows that two methods with different emotional EEG datasets almost have the same results at the end of the train. In fact, there is a little effect on rank-1 between three emotions. The left side of Figure 5 shows the test result of DE-CNN, while the right side shows the results of the CNN-LSTM. It is observed that the training process is more stable than the CNN-LSTM, while the DE-CNN convergences more quickly. In order to test the stability effect on the identification based on different moods, one of the affective models is utilized to predict the identity of the other two emotional states. It should be indicated that all methods are tested using default settings. Since performing all tests is not enough to find the most robust emotion, the one emotional model is also applied to the other two emotions. Table 8 shows the results obtained from the sympathetic EEG-based ID detection. Neu-pos and Neg-pos represent neutral and negative emotional data testing based on the positive model, respectively. Moreover, Pos-neu and Neg-neu represent positive and negative emotional data testing based on the neutral EEG dataset, respectively. Furthermore, Pos-neg and Neu-neg denote positive and neutral emotional data testing based on the negative EEG dataset, respectively. Comparing the results of identity authentication for two different bands, DE-RNN is



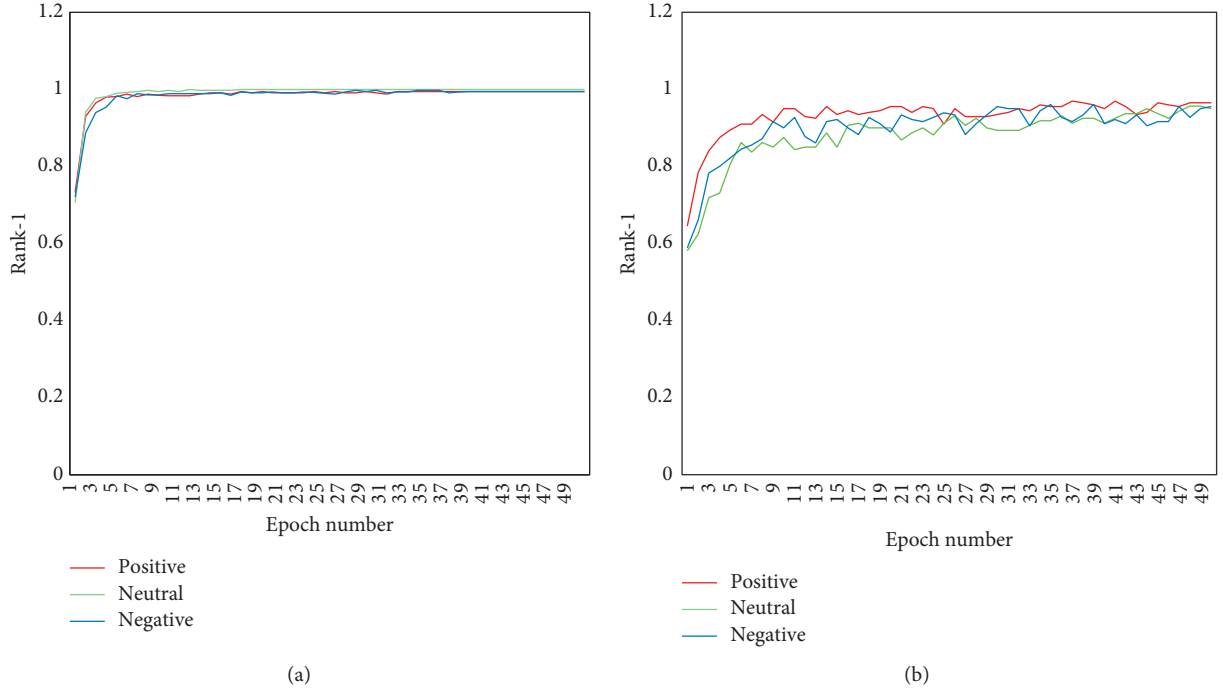


FIGURE 5: Training curve of the identity recognition based on three emotions for different algorithms: (a) DE-CNN algorithm and (b) CNN-LSTM algorithm.

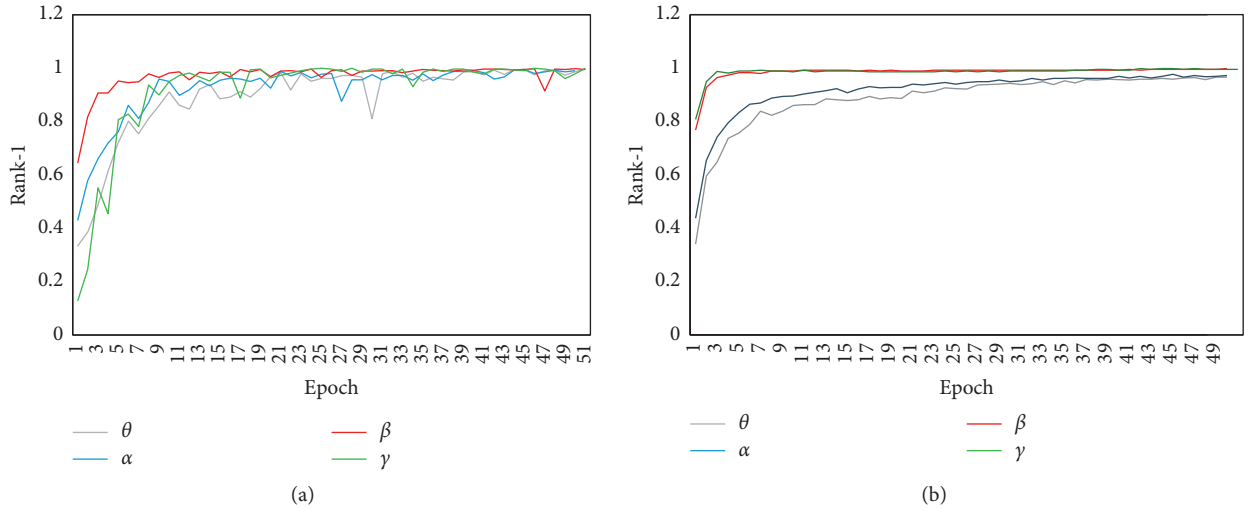


FIGURE 6: Testing curve of the identity recognition based on different bands: (a) results of the CNN-LSTM algorithm and (b) results of DE-CNN algorithm.

TABLE 6: Results of identification based on the four EEG bands.

Rank-1	4-7 Hz	8-14 Hz	14-31 Hz	32-40 Hz	4-40 Hz
CNN + LSTM	0.986	0.959	0.997	0.978	0.993
DE-CNN	0.864	0.972	0.998	0.994	0.996

better than CNN + LSTM. It is also observed that different emotions affect the identification and reduce rank-1 of the identification. Comparing the results of two methods indicates that positive emotions have a larger impact on identity authentication than other emotions. In other words,

TABLE 7: Rank-1 and EER of identification in three emotions.

	CNN + LSTM			DE + CNN		
Label	Positive	Neutral	Negative	Positive	Neutral	Negative
Rank-1	0.97	0.91	0.969	0.996	0.996	0.995
EER	0.012	0.06	0.01	0.001	0.001	0.002

the positive model is easier to overfit. It is found that neutral emotions are more robust to each other. Therefore, when the identification system based on the EEG is set up, the neutral

TABLE 8: Rank-1 of cross-emotion verification on identity recognition.

Rank	CNN + LSTM		DE + CNN	
	0-75hz	15-32hz	0-75hz	15-32hz
Neu-pos	0.83	0.95	0.915	0.83
Neg-pos	0.83	0.91	0.883	0.81
Pos-neu	0.89	0.84	0.91	0.914
Neg-neu	0.92	0.95	0.952	0.948
Pos-neg	0.86	0.78	0.897	0.845
Neu-neg	0.906	0.93	0.969	0.973

TABLE 9: Rank-1 of time interval emotion classification results.

DE-CNN	0.5-75 Hz		1-4 Hz		5-8 Hz		9-15 Hz		15-32 Hz	
Rank-1	Interval	NT	Interval	NT	Interval	NT	Interval	NT	Interval	NT
Positive	0.404	0.98	0.23	0.995	0.258	0.994	0.438	0.991	0.367	0.996
Neutral	0.4	0.996	0.3	0.978	0.326	0.992	0.324	0.995	0.402	0.993
Negative	0.336	0.952	0.264	0.957	0.276	0.984	0.276	0.991	0.495	0.995

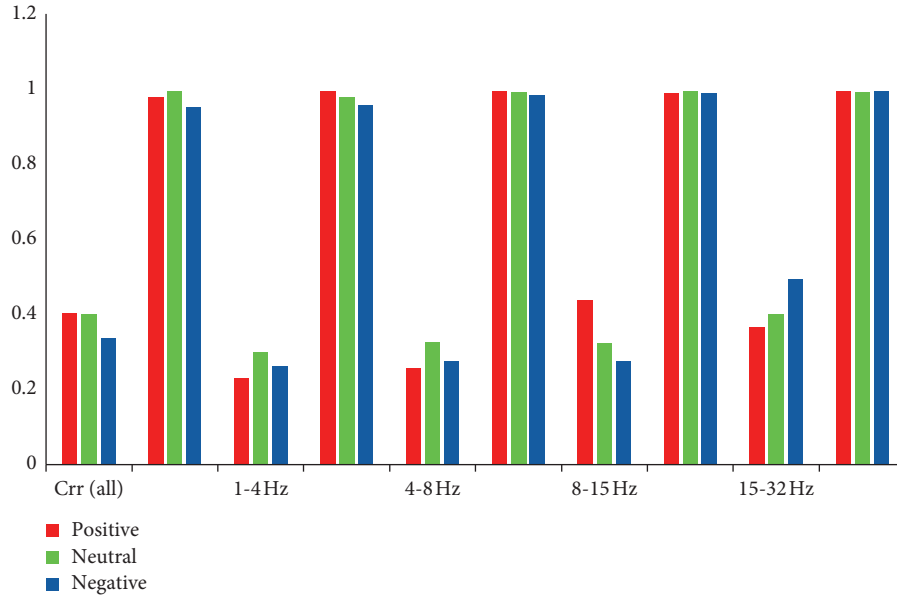


FIGURE 7: Rank-1 of the interval and noninterval verification. The graph in front of each band represents the interval result, while the latter represents the noninterval result. Red colour represents positive emotions, while green and blue colours represent neutral and negative emotion.

mood is proposed. Considering both the CNN-LSTM and the DE-CNN methods, beta waves (15-32 Hz) have worse results compared with 0-75 Hz. When robustness is required, the model needs wider EEG frequency band.

#### 4.4. Discussion

##### 4.4.1. Results of the Separated Dataset for the Identification.

In the SEED dataset, each subject performs three times of EEG signal collections with different time intervals. The longest time interval is four months, while the shortest time interval is three days. Each of the first two experiment EEG signals is used as the training dataset, while the last experimental EEG signal is used as the test dataset. Then the length of the longest interval recommended for the

identification is recorded. Moreover, the best compatible classification bands will be explored. It is found that EEG from different periods has different rank-1, where the details are presented in Table 9 and Figure 7. It indicates that the time interval rank-1 is lower than the noninterval. It should be indicated that NT represents the first two datasets of test rank-1, while the interval represents the last dataset accuracy based on the former train model. Moreover, different bands affect accuracy, where the 15-32 Hz band gives the best results. When time intervals of accuracy are obtained from the table, it is found that as the frequency increases, the results deteriorate. In this section, results are not compared with 32-75 Hz bands because the high frequency rarely appears in healthy people.

TABLE 10: Two ID recognition paradigms.

Paradigms	Content	Interpretability	Time validity
1	Same evoked content.	Strong	Weak attenuation
2	Different evoked content or the same evoked content; the order is different.	Weak	Strong attenuation

## 5. Conclusions

In order to evaluate the identity authentication based on the EEG, it is necessary to confirm that each person has a unique brain rhythm. In the present study, human EEG based on different contents is employed to recognize the identity. It is found that the proposed method can recognize the identity accurately.

Currently, there are two paradigms in processing the identification, where details are presented in Table 10. In the first paradigm, subjects see the same induced stimulus, where it reflects the different cognitive basis of a person. In the second paradigm, subjects see different contents induced stimuli, where longer EEG of each person should be divided into different parts to eliminate the variations caused by different contents. Personal characteristics are extracted during personal identification. There are many similarities between the brain wave and voice wave verification. Two main methods are proposed in this regard. In the first method, the extracted acoustic features are initially aligned with specific sounding units, features are projected into a lower linear space, and then speaker information is mined. Intuitively, “the difference between different people in the same tone [41–44]” can be understood as mining. The first method draws on some phonological knowledge that uses a vocal unit classification network for speech recognition. However, the second method “end-to-end deep learning-based authentication” is a purely data-driven approach. Through massive data samples and very deep convolution neural networks, the machine automatically explores speaker difference information in acoustic features to extract speaker information representations in acoustic features.

More specifically, the deep convolution neural network is trained by a large amount of voice wave data, and the output category is defined as the speaker ID. In actual training, tens of thousands of IDs are required for network training. Thus, a basic network capable of effectively characterizing the speaker is obtained. However, EEG is not easy to obtain a large number of samples due to equipment limitations. But EEG identification offers more choices for identity authentication.

The second most important feature in identity authentication is stability. In the present research, the second method is performed for identity authentication. In terms of stability, emotional factors and time factors are mainly considered. In the emotional factors, it is found that different emotions have little impact on the classification results with DE-CNN. While positive emotional robustness is the worst in terms of mutual migration, the neutral emotional robustness is best in three emotions. From the perspective of different frequency classification results, the frequency band of 15–32 Hz is superior to other frequency bands in both classification and migration. In terms of interval dimensions,

the 15–32 Hz band is more compatible than the other bands of the EEG and performs well at different points during the time.

Based on our current research, we will explore cross identity identification methods next and design models to eliminate the impact of time on individual EEG, which is very important, and this is the key point for EEG to become a daily identity authentication.

## Data Availability

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

## 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 Key Project of National Key R & D Project (no. 2017YFC1703303), Natural Science Foundation of Fujian Province of China (no. 2019J01846, no. 2018J01555, no. 2017J01773, and no. 2016J01674), External Cooperation Project of Fujian Province, China (no. 2019I0001), and Science and Technology Guiding Project of Fujian Province, China (2019Y0046).

## References

- [1] P. J. Phillips, G. H. Givens, J. R. Beveridge, B. A. Draper, D. S. Bolme, and Y. M. Lui, “Biometric face recognition: from classical statistics to future challenges I NIST,” *Computational Statistics*, vol. 5, no. 4, pp. 228–308, 2013.
- [2] R. P. Wildes, “Iris recognition: an emerging biometric technology,” *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.
- [3] K. I. Chang, K. W. Bowyer, and S. Sarkar, “Comparison and combination of ear and face images in appearance-based biometrics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1160–1165, 2003.
- [4] A. E. Vacarus, “Biometric security-fingerprint recognition system,” *Journal of Mobile, Embedded and Distributed Systems*, vol. 7, pp. 17–23, 2015.
- [5] C. Huang and G. Tian, “A new pulse coupled neural network (pcnn) for brain medical image fusion empowered by shuffled frog leaping algorithm,” *Frontiers in Neuroscience*, vol. 13, 2019.
- [6] Q. Gui, Z. Jin, and W. Xu, “Exploring EEG-based biometrics for user identification and authentication,” in *Proceedings of the 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1–6, Philadelphia, PA, USA, December 2014.
- [7] F. Movahedi, J. L. Coyle, and E. Sejdic, “Deep belief networks for electroencephalography: a review of recent contributions

- and future outlooks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 642–652, 2017.
- [8] M. Shahin, B. Ahmed, S. T.-B. Hamida, F. L. Mulaffer, M. Glos, and T. Penzel, “Deep learning and insomnia: assisting clinicians with their diagnosis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1546–1553, 2017.
  - [9] D. Zhang, L. Yao, X. Zhang, S. Wang, W. Chen, and R. Boots, *EEG-based Intention Recognition from Spatiotemporal Representations via Cascade and Parallel Convolutional Recurrent Neural Networks in Human-Computer Interaction*, 2017, <https://arxiv.org/abs/1708.06578>.
  - [10] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, “Emotion-meter: a multimodal framework for recognizing human emotions,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.
  - [11] E. Maiorana, J. Solé-Casals, and P. Campisi, “EEG signal preprocessing for biometric recognition,” *Machine Vision and Applications*, vol. 27, no. 8, pp. 1351–1360, 2016.
  - [12] Y. Chen, A. D. Atnafu, I. Schlattner et al., “A high-security EEG-based login system with rsvp stimuli and dry electrodes,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2635–2647, 2016.
  - [13] M. D. Pozobanos, J. B. Alonso, J. R. Ticayrivas, and C. M. Travieso, “Electroencephalogram subject identification: a review,” *Expert Systems With Applications*, vol. 41, pp. 6537–6554, 2014.
  - [14] Q. Wu, Y. Zeng, C. Zhang, L. Tong, and B. Yan, “An EEG-based person authentication system with open-set capability combining eye blinking signals,” *Sensors*, vol. 18, no. 2, p. 335, 2018.
  - [15] J.-H. Kang, Y. C. Jo, and S.-P. Kim, “Electroencephalographic feature evaluation for improving personal authentication performance,” *Neurocomputing*, vol. 287, pp. 93–101, 2018.
  - [16] Y. Sun, F. P.-W. Lo, and B. Lo, “EEG-based user identification system using 1d-convolutional long short-term memory neural networks,” *Expert Systems With Applications*, vol. 125, pp. 259–267, 2019.
  - [17] L. A. Moctezuma, A. A. Torres-García, L. Villaseñor-Pineda, and M. Carrillo, “Subjects identification using EEG-recorded imagined speech,” *Expert Systems with Applications*, vol. 118, pp. 201–208, 2019.
  - [18] X. Zhang, L. Yao, S. S. Kanhere, Y. Liu, T. Gu, and K. Chen, “Mind ID: person identification from brain waves through attention-based recurrent neural network,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–23, 2018.
  - [19] N. Thammasan, K. Moriyama, K.-I. Fukui, and M. Numao, “Familiarity effects in EEG-based emotion recognition,” *Brain Informatics*, vol. 4, no. 1, pp. 39–50, 2017.
  - [20] L. C. Shi, Y. Y. Jiao, and B. L. Lu, “Differential entropy feature for EEG-based vigilance estimation,” in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6627–6630, Osaka, Japan, July 2013.
  - [21] R. N. Duan, J. Y. Zhu, and B. L. Lu, “Differential entropy feature for EEG-based emotion classification,” in *Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Diego, CA, USA, November 2013.
  - [22] G. Boffetta, “Predictability: a way to characterize complexity,” *Physics Reports*, vol. 6, no. 356, pp. 367–474, 2002.
  - [23] D. Mathersul, L. M. Williams, P. J. Hopkinson, and A. H. Kemp, “Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety,” *Emotion*, vol. 8, no. 4, pp. 560–572, 2008.
  - [24] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, “Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music,” *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007.
  - [25] N. Martini, D. Menicucci, L. Sebastiani et al., “The dynamics of EEG gamma responses to unpleasant visual stimuli: from local activity to functional connectivity,” *NeuroImage*, vol. 60, no. 2, pp. 922–932, 2012.
  - [26] W. Zheng and B. Lu, “Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,” *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2017.
  - [27] W. Klimesch, M. Doppelmayr, H. Russegger, T. Pachinger, and J. Schwaiger, “Induced alpha band power changes in the human EEG and attention,” *Neuroscience Letters*, vol. 244, no. 2, pp. 73–76, 1998.
  - [28] W. Ray and H. Cole, “EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes,” *Science*, vol. 228, no. 4700, pp. 750–752, 1985.
  - [29] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, “Exploring EEG features in cross-subject emotion recognition,” *Frontiers in Neuroscience*, vol. 12, 2018.
  - [30] S. Taran and V. Bajaj, “Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method,” *Computer Methods and Programs in Biomedicine*, vol. 173, pp. 157–165, 2019.
  - [31] T. Wilaiprasitporn, A. Dittaporn, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, *Affective EEG-Based Person Identification Using the Deep Learning Approach*, IEEE, Piscataway, NJ, USA, 2018.
  - [32] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, “Entropies for detection of epilepsy in EEG,” *Computer Methods and Programs in Biomedicine*, vol. 80, no. 3, pp. 187–194, 2005.
  - [33] C. Wen Lin, H. Min Wei, J. BoLin, and C. Kuo Sheng, “Analysis of EEG entropy during visual evocation of emotion in schizophrenia,” *Annals of General Psychiatry*, vol. 16, pp. 1–10, 2017.
  - [34] R. G. Andrzejak, K. W. Bowyer, and S. Sarkar, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state,” *Physical Review*, vol. 64, p. 1, 2001.
  - [35] S. P. Kumar, N. Sriraam, P. G. Benakop, and B. C. Jinaga, “Entropies based detection of epileptic seizures with artificial neural net-work classifiers,” *Expert Systems with Applications*, vol. 37, pp. 3284–3291, 2010.
  - [36] A. Wolf, “Wolf Lyapunov exponent estimation from a time series (MATLAB central file exchange),” 2019, <https://www.mathworks.com/matlabcentral/fileexchange/48084-wolf-lyapunov-exponent-estimation-from-a-time-series>.
  - [37] X. Shungen, L. Shulin, S. Mengmeng, A. Nie, and Z. Hongli, “Coupling rub-impact dynamics of double translational joints with subsidence for time-varying load in a planar mechanical system,” *Multibody System Dynamics*, vol. 48, no. 4, pp. 451–486, 2019.
  - [38] H. H. Jasper, “The ten twenty electrode system: international federation of societies for electroencephalography and clinical neurophysiology,” *American Journal of EEG Technology*, vol. 1, no. 1, pp. 13–19, 1961.

- [39] T. Wilaiprasitporn, A. Ditthaporn, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-based person identification using the deep learning approach," *IEEE Transactions on Cognitive and Developmental System*, vol. 1, no. 1, 2019.
- [40] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, *Emotion Recognition from Multi-Channel EEG through the Parallel Convolutional Recurrent Neural Network*, pp. 1–7, IEEE, Piscataway, NJ, USA, 2018.
- [41] S. Garimella, A. Mandal, and N. Strom, "Robust vector-based adaptation of DNN acoustic model for speech recognition," in *Proceedings of the INTERSPEECH 2015 16th Annual Conference of the International Speech Communication Association*, pp. 2877–2881, Dresden, Germany, September 2015.
- [42] C. Huang, X. Shan, Y. Lan et al., "A hybrid active contour segmentation method for myocardial D-SPECT images," *IEEE Access*, vol. 6, no. 6, pp. 39334–39343, 2018.
- [43] C. Huang, C. Wang, J. Tong, L. Zhang, F. Chen, and Y. Hao, "Automatic quantitative analysis of bioresorbable vascular scaffold struts in optical coherence tomography images using region growing," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 1, pp. 98–104, 2018.
- [44] C. Huang, S. Wan, J. Yang et al., "Automatic side branch detection in optical coherence tomography images using adjacent frame correlation information," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1513–1518, 2018.



## Research Article

# An Automatic Epilepsy Detection Method Based on Improved Inductive Transfer Learning

Yufeng Yao <sup>1,2</sup> and Zhiming Cui <sup>3</sup>

<sup>1</sup>The Institute of Intelligent Information Processing and Application, Soochow University, Suzhou 215006, China

<sup>2</sup>Department of Computer Science and Engineering, Changshu Institute of Technology, Changshu 215500, China

<sup>3</sup>Suzhou University of Science and Technology, Suzhou 215009, China

Correspondence should be addressed to Yufeng Yao; yyf@cslg.edu.cn and Zhiming Cui; zmcui@usts.edu.cn

Received 20 June 2020; Accepted 15 July 2020; Published 3 August 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Yufeng Yao and Zhiming Cui. 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.

Epilepsy is a chronic disease caused by sudden abnormal discharge of brain neurons, causing transient brain dysfunction. The seizures of epilepsy have the characteristics of being sudden and repetitive, which has seriously endangered patients' health, cognition, etc. In the current condition, EEG plays a vital role in the diagnosis, judgment, and qualitative location of epilepsy among the clinical diagnosis of various epileptic seizures and is an indispensable means of detection. The study of the EEG signals of patients with epilepsy can provide a strong basis and useful information for in-depth understanding of its pathogenesis. Although, intelligent classification technologies based on machine learning have been widely used to the classification of epilepsy EEG signals and show the effectiveness. In fact, it is difficult to ensure that there is always enough EEG data available for training the model in real life, which will affect the performance of the algorithms. In view of this, to reduce the impact of insufficient data on the detection performance of the algorithms, a novel discriminate least squares regression-(DLSR-) based inductive transfer learning method was introduced which is on the basis of DLSR and the inductive transfer learning. And, it is applied to promote the adaptability and accuracy of the epilepsy EEG signal recognition. The proposed method inherits the advantages of DLSR; it can be more suitable for classification scenarios by expanding the interval between different classes. Meanwhile, it can simultaneously use the data of the target domain and the knowledge of the source domain, which is helpful for getting better performance. The results show that the improved method has more advantages in EEG signal recognition comparing to several other representative methods.

## 1. Introduction

At present, epilepsy has become a common disease in neurology. Its pathogenesis has not yet been fully elucidated, and it is usually defined as a chronic neurological disease caused by sudden abnormal discharge of brain neurons. The epileptic seizures are sudden and repetitive. Its onset is accompanied by clinical manifestations such as loss of consciousness, fainting, and twitching of extremities. It also has cognitive and mental disorders that seriously endanger patients' health, cognition, etc. [1, 2]. According to statistics, more than one percent of the world's population suffer from the disease [3], and there are approximately 9 million people with epi-

lepsy in China. Therefore, the depth research and prevention of epilepsy play an indispensable role to alleviate the suffering of patients, improve the quality of life, and promote healthy development. As an important method for studying epilepsy, EEG uses electrodes to record the electrical activity of nerve cells in the brain, which contains a large amount of physiological and pathological information, and is of great importance in the clinical examination, location, and therapy of epilepsy. Therefore, for people with a tendency to epilepsy, automatic detection of epilepsy can analyze and screen the EEG signals of people at high risk for epilepsy, so as to realize early detection, perform timely intervention, and reduce the impact of epilepsy on people and the incidence of epilepsy.

In view of this, it is of great value to study the epilepsy automatic detection algorithm based on EEG signals and develop an efficient and accurate epilepsy automatic detection system.

In fact, the study of automatic epileptic detection based on EEG signals has attracted extensive attention from scholars and experts at home and abroad since the 1970s. To predict the onset or preonset of epilepsy in the process of seizure detection, machine learning and pattern classification algorithms are generally applied to classify the EEG signals after extracting the characteristics of the time domain, frequency domain, time frequency domain, or nonlinear domain of the EEG. With the development of computer technology and digital signal processing technology, more and more methods are widely used in the study of seizure detection methods and have achieved certain research results, such as the Bayesian classifier [4], artificial neural network [5–9], support vector machine (SVM) [10–13], and fuzzy reasoning [14, 15]. For example, Obeyli extracted the Lyapunov exponential features of EEG signals and used probabilistic neural networks to classify EEG signals, so as to achieve high classification results [9]. Chan et al. extracted the time-frequency features of five subbands in the wavelet transform domain of epilepsy EEG signals and then used support vector machines and cluster regression models to recognize the onset of seizures. Aarabi et al. [14] extracted the features such as sample entropy, dominant frequency, average amplitude, and amplitude variation coefficient of intracranial EEG data of patients with epilepsy and used the established fuzzy inference rules to fuse the EEG feature information for seizure detection. Although many of the above intelligent classification methods have shown the effectiveness of epilepsy EEG signal classification, they still face a challenge, that is, it is very hard to get enough EEG data for epilepsy to train the model in real life. Therefore, it has important practical value to explore how to use the knowledge acquired from related fields to enhance the classification performance of EEG data in the current scenario [16].

To solve the above challenges, a novel inductive transfer learning method based on discriminant least squares regression (TDLSR) was proposed. Meanwhile, it was applied to specific medical application scenarios, namely, epilepsy EEG signal classification, so as to relieve the effect of severe data shortage on the performance of the algorithms. Transfer learning is an effective way to transfer knowledge from related fields and is helpful of obtaining more information in the absence of sufficient data or information. It focuses on how to use the useful information from similar but different source domains to improve the classification result of the classifier in the target domain. When studying epilepsy EEG signal classification, the inductive transfer learning method naturally becomes the first choice because of insufficient labeled epilepsy EEG samples in the target domain sometimes. What is more, since the discriminant LSR is still based on least squares regression (LSR) [17], which can explain the importance of each feature in the prediction model based on the original data space, we introduced the inductive transfer learning method based on DLSR to use for epilepsy EEG sig-

nal classification. In summary, the innovations of this work are summarized as follows:

*Point 1.* The improved method is on the basis of the inductive transfer learning, but it has some difference from the traditional inductive transfer learning. The latter directly transfers the samples or features used in the source domain to the target domain for transfer learning, while the former uses a knowledge lever mechanism that transfers some knowledge from the source domain to the target domain. Then, the security of the data in the source domain can be well protected, and the data in the target domain and the knowledge in the source domain can be used simultaneously, so that the classification effect is better.

*Point 2.* The improved method expands DLSR that can be well applied to classification scenes into a novel method with certain transfer learning ability, so that it can be used in more complex scenes.

*Point 3.* The improved method inherits the characteristics of DLSR, that is, it can be better applied to classification scenarios by expanding the interval between different categories. And, it can transfer knowledge from the source domain, thus ensuring the rationality of its training model.

Finally, to better illustrate the basic idea of this study, the structure of the paper is as follows:

*Part 1.* Introduction to the research background, status, and significance of the thesis.

*Part 2.* The related work, including the related technology of epilepsy EEG signal detection and the epilepsy EEG signal classification based on transfer learning, is summarized in advance so that the following sections become more readable.

*Part 3.* The notations of the inductive transfer learning algorithm based on DLSR was introduced in detail.

*Part 4.* The reliability and validity of TDLSR algorithm in detection of epilepsy EEG signals based on a series of experimental were verified.

## 2. Related Work

Automatic epilepsy detection is based on signal processing technology and pattern recognition. It analyzes EEG data to identify the location and duration of seizures. Usually, the EEG signals collected during seizures was called seizure EEG, and the EEG signals collected in the nonseizure are called nonseizure EEG. The problem of automatic detection of epilepsy EEG is to effectively judge the above two types of EEG signals and identify seizures. The related detection technologies are introduced as follows.

*2.1. The Related Technology of Epilepsy EEG Signal Detection.* Because the EEG signal of epilepsy is easily interfered by many factors, it is very random, and it is a nonstationary signal, and its rule is generally difficult to grasp. Therefore, researchers often use quantitative analysis to extract characteristic information of epilepsy EEG signals. The existing methods of automatic epilepsy detection include the following:

- (1) Time-domain analysis: time-domain analysis is one of the earliest methods used in signal analysis. It

analyzes the time-domain waveforms of EEG signals to study the difference between EEG waveform during seizures and EEG waveform during nonseizures, and directly extracts the waveform characteristics of the signals to distinguish the two types of EEG signals. The time-domain analysis method has the characteristics of intuitiveness and clear physical meaning, and it can reflect the important information in transient waveform such as spike wave and harp wave. The representative methods of time-domain analysis have a template matching method, time-domain waveform, or energy characteristics, etc. In 2001, Litt et al. [18] performed feature extraction on EEG signals based on time-domain analysis methods. In addition, researchers performed time-domain analysis of EEG signals to extract waveforms or energy features different from background activities and use amplitude, rhythm, period, and other parameters as classification criteria to identify epilepsy EEG signals. Gotman et al. performed a “half-wave” decomposition of EEG signals and then extracted EEG features, including average amplitude, duration, and coefficient of variation relative to the background, and set thresholds based on expert experience. The characteristic parameters are compared with the threshold to judge whether it is an epilepsy signal [19–21]. Time-domain analysis only performs statistical analysis from the time domain, and it is easy to miss other important changes in abnormal signals, such as slow waves

- (2) Frequency analysis: unlike time-domain analysis, which mainly analyzes the waveform characteristics of epilepsy EEG, frequency domain analysis analyzes the frequency characteristics of EEG. It recognizes different rhythms according to the frequency of brain waves. Each brain wave of different rhythms corresponds to epilepsy EEG signals in different time periods or different parts of the brain [22]. Frequency domain analysis is based on the Fourier transform and is mainly used for power spectrum analysis of EEG signals. It performs the Fourier transform on the EEG signal to obtain its frequency components and spectrum distribution and extracts the corresponding EEG features in the frequency domain for epilepsy detection and recognition. Representative methods include power spectrum estimation, autoregressive (AR) model spectrum estimation, and higher order spectrum [23]. Among them, the power spectrum estimation transforms the EEG signal whose amplitude changes with time into the EEG spectrum chart with power varying with frequency and analyzes the distribution and change of each frequency band of the EEG signal intuitively and quantitatively [24, 25]. Although frequency domain analysis can provide a lot of effective information, allowing researchers to detect epilepsy based on the frequency domain characteristics of EEG, the overall spectrum of the signal obtained by the Fourier transform nei-

ther can reflect the local characteristics of the signal nor can reflect the signal frequency component changes with time. Therefore, the detection results obtained by frequency domain analysis are not very satisfactory and are greatly restricted in practical applications

- (3) Time-frequency analysis: the epilepsy EEG is a typical nonstationary signal, which contains not only the waveform parameter characteristics in the time domain but also the energy distribution characteristics in the frequency domain. However, neither the above two methods can fully extract the transient characteristics and information of the EEG signals and can get the ideal results. With the development of digital signal theory and methods, the time-frequency analysis method combining time domain and frequency domain is widely used in the analysis of nonstationary EEG signals. It can obtain time and frequency domain information at the same time and capture transient information of EEG. In recent years, more and more studies have adopted time-frequency analysis methods to analyze EEG signals, among which various wavelet change methods are represented. The wavelet transform uses the translation and expansion of the window function to implement a wide time window for the low-frequency components of the signal and a narrow time window for the high-frequency components to complete the multiscale analysis of the signal. This analysis method conforms to the laws of nature and has a good ability to characterize the local characteristics of the signal [26]. It can capture the transient characteristics of the EEG signal and accurately locate it in the time and frequency domains. In addition to wavelet transform, commonly used time-frequency analysis methods also include empirical mode decomposition [27–30], the Wigner-Ville distribution [31, 32], and the Stockwell transform [33, 34]. However, most of these time-frequency analysis methods can only be used for multiresolution analysis of the original signal and then need to be combined with other algorithms to achieve the feature extraction and selection of EEG. Figure 1 shows the comparison of time-domain analysis, frequency domain analysis, and wavelet transform analysis.

Observe the above figures, it is easy to draw a conclusion that the time-frequency analysis can provide more useful information compared to time-domain analysis and frequency domain analysis. In Figure 1(c), the wavelet transform improves the time resolution at the high frequency of the signal by changing the time window and improves the frequency resolution at the low frequency, which has a better classification effect

- (4) Nonlinear dynamic analysis: nowadays, with the progress of nonlinear dynamics theory, researchers are devoted to studying the nonlinear of EEG signals

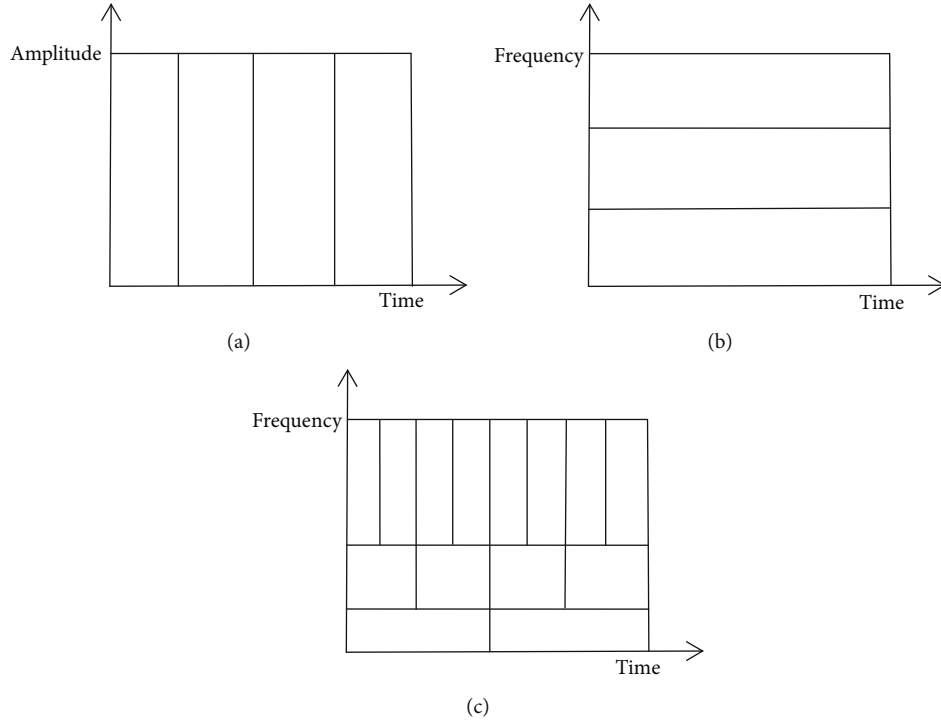


FIGURE 1: Comparison of time-domain analysis, frequency domain analysis, and time-frequency analysis. (a) Time-domain analysis. (b) Frequency analysis (Fourier transform). (c) Wavelet transform.

to solve the problem of automatic detection of epilepsy. Using nonlinear dynamics theory methods for EEG signal analysis, various nonlinear features of EEG signals can be extracted to distinguish epilepsy EEG signals from normal EEG signals. This provides some new research ideas for automatic epilepsy detection technology. Kannathal et al. used different entropies to measure the chaotic characteristics of EEG signals and used them as EEG features to distinguish EEG signals in different periods [35], including the Shannon entropy, Renyi entropy, Kolmogorov-Sinai entropy, and approximate entropy. The results show that the complexity of the EEG signal in patients with epilepsy during the intermittent period is higher than that during the seizure period, that is, the complexity of the EEG signal during the seizure is reduced, and the bet value is less than the normal EEG signal. Although nonlinear analysis can reflect the dynamic mechanism of seizures well, most of the nonlinear features are computationally intensive and generally time-consuming, which is not suitable for real-time epilepsy automatic detection systems.

**2.2. The Epilepsy EEG Signal Classification Based on Transfer Learning.** Traditional classification methods use a large amount of data with label information to train a decision function and then use this function to classify and identify test samples with unknown label information. However, these classification methods all have a presupposition: training data and test data need to obey the same distribution characteristics, as shown in Figure 2. For the differences in

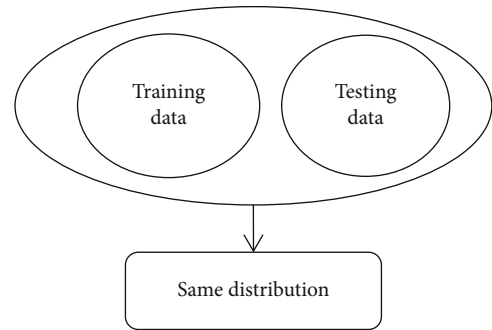


FIGURE 2: Scene suitable for the traditional method.

the distribution of training samples and test samples as described above, the performance of the traditional methods significantly decrease, as shown in Figure 3. In response to this challenge, transfer learning is a promising research direction. Transfer learning focuses on knowledge transfer problems that are similar to different domains or have different data distributions. It enhances the performance of the classifier used for target area recognition by learning useful knowledge from the source domain. According to whether the target domain used contains samples with labeled information, transfer learning techniques are divided into three categories: inductive transfer learning method, direct transductive transfer learning method, and unsupervised transfer learning method [36]. In the paper, we will focus on an inductive transfer learning method with good performance, that is, inductive transfer learning method

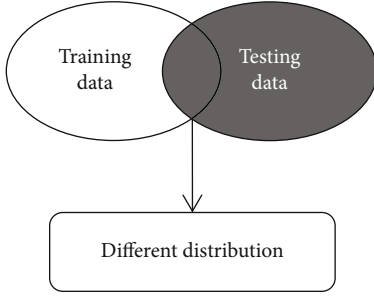


FIGURE 3: New challenge for the traditional methods.

based on discriminant least squares regression (TDLSR). And, its application and actual effect in EEG signal detection of epilepsy will be studied. The framework structure of epilepsy EEG signal detection based on transfer learning theory is given, as shown in Figure 4.

In short, transfer learning is to transfer the knowledge (useful knowledge) from the source domain with a large amount of labeled data for learning in the target domain with no or little labeled data, thereby improving the training quality of the target domain. This can reduce the workload of collecting labeled data in the target domain.

### 3. The Inductive Transfer Learning Algorithm Based on DLSR

To better describe the algorithm proposed in this paper, Table 1 gives a detailed description of the symbols in the algorithm.

Since DLSR is a nontransforming algorithm, the symbols in Table 1 refer to the parameter variables of the original training sample.

**3.1. The Least Squares Regression.** As a widely used method based on statistical theory, LSR has become a typical method. LSR uses the Frobenius norm to constrain the matrix of representation coefficients. In the paper, to expand the classification ability of the LSR algorithm, we preconstructed the binary label matrix  $Y$  corresponding to the training samples  $X$ , so that it can better cope with more complex classification scenarios. The  $j$ th column of  $Y$  indicates that only the data belonging to the  $j$ th class corresponds to an element equal to 1, and all other elements are 0. Then, the objective function of LSR can be redefined as follows:

$$\arg \min_Z \|XZ - Y\|_F^2 + \lambda \|Z\|_F^2. \quad (1)$$

Since the LSR has an analytical solution, which can be easily obtained, formula (1) can be rewritten as:

$$J(Z) = \arg \min_Z \|XZ - Y\|_F^2 + \lambda \|Z\|_F^2. \quad (2)$$

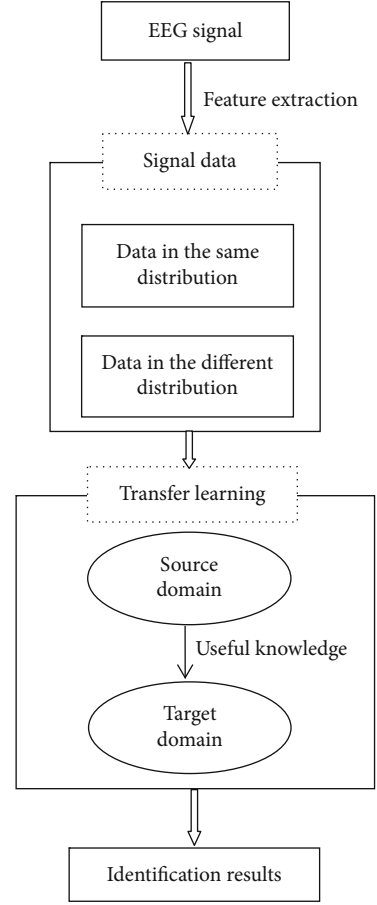


FIGURE 4: Framework of adaptive recognition for epileptic EEG signals based on transfer learning.

TABLE 1: Symbol description.

Symbol	Description
$X$	Training matrix of the target domain, $X = [x_1, x_2, \dots, x_N]$ and $x_i \in R^d$
$N$	Total number of training samples of the target domain
$d$	Dimension of sample
$C$	The number of classes
$Y$	The binary label matrix corresponding to the training samples in the target domain. The $j$ th column of $Y$ indicates that only the data belonging to the $j$ th class corresponds to an element equal to 1, and all other elements are 0
$B$	Indicator matrix constructed based on $Y$
$W$	Label offset matrix of the target domain
$Z$	Mapping matrix of target domain
$Z_S$	Mapping matrix of source domain
$e$	Vector representing all 1, $e \in R^N$
$\ominus$	Hadamard product operator for matrix
$p$	The mapping offset vector of target domain
$\lambda, \eta$	The regularization parameter



Then, the solution process of the analytical solution of LSR is as follows: let

$$\begin{aligned} \frac{\partial J(Z)}{\partial Z} &= 0, \\ \implies 2X^T XZ - 2X^T Y + 2\lambda Z &= 0, \\ \implies (X^T X + \lambda I_d)Z &= X^T Y. \end{aligned} \quad (3)$$

Finally, the analytical solution of LSR can be obtained as

$$Z = (X^T X + \lambda I_d)^{-1} X^T Y. \quad (4)$$

**3.2. The Discriminate Least Squares Regression.** As we know, LSR can be directly used for classification tasks. However, since the interval between any two different classes in the constructed binary class label matrix is  $\sqrt{2}$ , the DLSR proposed in literature [37] introduces the relaxation technique into the LSR so as to expand the interval between the two data from different classes. To improve the compactness of the classification task, DLSR will comprehensively consider the class factors and build an indicator matrix  $B$  on the basis of the binary label matrix  $Y$  of the sample. Each element of matrix  $B$  is defined as follows:

$$B_{ij} = \begin{cases} +1, & \text{if } y_i = 1, \\ -1, & \text{otherwise.} \end{cases} \quad (5)$$

In essence, each element of matrix  $B$  represents the offset direction of the corresponding class label. Then,  $\varepsilon$  relaxation on each element of  $Y$  is performed, and the amount of  $\varepsilon$  relaxation through the matrix  $W$  is recorded. Then, the objective function of DLSR can be expressed as

$$\min_{s.t. W \geq 0} \|XZ + e p^T - (Y + B\Theta W)\|_F^2 + \lambda \|Z\|_F^2. \quad (6)$$

The objective function of DLSR is a convex optimization problem. However, it cannot directly optimize the solution. Literature [37] adopts an alternative optimization strategy and ensures that a closed solution is obtained at each step. The specific derivation process is as follows:

- (1) Fix  $W$ , and update  $Z$  and  $p$ . Let  $L = Y + B\Theta W$ ; formula (6) can be rewritten as

$$J(Z, p) = \arg \min_{Z, p} \|XZ + e p^T - L\|_F^2 + \lambda \|Z\|_F^2. \quad (7)$$

According to the optimization theory, we take a partial derivative of  $p$ , namely

$$\begin{aligned} \frac{\partial J(Z, p)}{\partial p} &= 0, \\ \implies Z^T X^T e + p e^T e - L^T e &= 0, \\ \implies p &= \frac{(L^T e - Z^T X^T e)}{N}. \end{aligned} \quad (8)$$

Furthermore, we find the partial derivative of  $Z$ , and we can get

$$\begin{aligned} \frac{\partial J(Z)}{\partial Z} &= 0, \\ \implies 2X^T XZ - 2X^T (e p^T - L) + 2\lambda Z &= 0, \\ \implies X^T XZ - X^T \left( e \frac{(e^T L - e^T XZ)}{N} - L \right) + \lambda Z &= 0, \\ \implies X^T \left( I_N - \frac{e e^T}{N} \right) XZ - X^T \left( I_N - \frac{e e^T}{N} \right) L + \lambda Z &= 0, \\ \implies Z &= (X^T HX + \lambda I_d)^{-1} X^T HL \end{aligned} \quad (9)$$

- (2) Fix  $Z$  and  $p$ , and update  $W$ . Let  $G = XZ + e p^T - Y$ ; then,  $W$  can be solved as follows:

$$\begin{aligned} \arg \min_W \|G - B\Theta W\|_F^2, \\ s.t. W \geq 0 \end{aligned} \quad (10)$$

According to literature [37], the Frobenius norm square of a matrix can be solved element by element, so Eq. (10) can be equivalent to solve  $N \times C$  subproblems. For element  $W_{ij}$  in row  $i$  and column  $j$ , there can be

$$\min_{W_{ij}} (G_{ij} - B_{ij} W_{ij})^2, \quad s.t. W_{ij} \geq 0 \quad (11)$$

where  $G_{ij}$  and  $B_{ij}$  represent the  $j$ th element of the  $i$ th row of the matrix  $G$  and matrix  $B$ , respectively and satisfy  $B_{ij}^2 = 1$ . Then, we can get

$$(G_{ij} - B_{ij} W_{ij})^2 = (B_{ij} G_{ij} - W_{ij})^2. \quad (12)$$

And, because each element of  $W$  satisfies  $W_{ij} \geq 0$ , formula (12) can be written as

$$W_{ij} = \max(B_{ij} G_{ij}, 0). \quad (13)$$

The description of DLSR

**Input:**

The training samples  $X = [x_1, x_2, \dots, x_N] \in R^{N \times d}$ , and their corresponding class labels  $y_i \in \{1, 2, \dots, C\} (i = 1, 2, \dots, N)$ , where  $x_i \in R^d (i = 1, 2, \dots, N)$ . The maximum number of iterations is  $T$ .

**Output:**

The mapping matrix  $Z$  and the translation vector  $p$ .

**Training:**

Step 1: Construct the label matrix  $Y$  and the indicator matrix  $B$ , respectively.

Step 2: Get the value of  $\lambda$  by grid searching in the set of  $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10, 10^2\}$ .

Step 3: Initialize  $Z' = 0, p = 0$ , and  $W = 0$ . Set  $t = 1$ .

Step 4: Update  $Q$  by  $Q = (X^T H X + \lambda I_d)^{-1} X^T H$ .

Repeat

Step 5: Update the label matrix  $L$  by  $L = Y + B \Theta W$ .

Step 6: Update  $p$  and  $Z$  by Eqs. (8) and (10).

Step 7: Update the regression error matrix  $G$  by  $G = XZ + ep^T - Y$ .

Step 8: Update the label shift matrix  $W$  by Eq. (14).

Step 9: Let  $Z' = Z$  and  $t = t + 1$

Until  $\|Z' - Z\|_F^2 \leq 10^{-4}$  or  $t > T$ .

Step 10: Output  $Z$  and  $p$ .

ALGORITHM 1: The description of the DLSR algorithm.

The description of TDLSR

**Input:**

The training samples in target domain  $X = [x_1, x_2, \dots, x_N] \in R^{N \times d}$ , and their corresponding class labels  $y_i \in \{1, 2, \dots, C\} (i = 1, 2, \dots, N)$ , where  $x_i \in R^d (i = 1, 2, \dots, N)$ . The maximum number of iterations is  $T$ .  $A_S$  is learned in advanced by using DLSR from source domain.

**Output:**

The mapping matrix  $Z$  and the translation vector  $p$ .

**Training:**

Step 1: Construct the label matrix  $Y$  and the indicator matrix  $B$ , respectively.

Step 2: Get the value of  $\lambda$  and  $\eta$  by grid searching in the set of  $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10, 10^2\}$ , respectively.

Step 3: Initialize  $Z' = 0, p = 0$ , and  $W = 0$ . Set  $t = 1$ .

Step 4: Compute  $Q$  by  $Q = (X^T H X + \lambda I_d)^{-1} X^T H$  and  $V$  by  $V = \eta(X^T H X + (\lambda + \eta)I_d)^{-1} Z_S$ , respectively.

Repeat

Step 5: Update the label matrix  $L$  by  $L = Y + B \Theta W$ .

Step 6: Update  $p$  and  $Z$  by Eqs. (17) (18).

Step 7: Update the regression error matrix  $G$  by  $G = XZ + ep^T - Y$ .

Step 8: Update the label shift matrix  $W$  by Eq. (20).

Step 9: Let  $Z' = Z$  and  $t = t + 1$

Until  $\|Z' - Z\|_F^2 \leq 10^{-4}$  or  $t > T$ .

Step 10: Output  $Z$  and  $p$ .

ALGORITHM 2: The description of the TDLSR algorithm.

Therefore, the final solution formula of  $W$  is:

$$W = \max(B \Theta G, 0). \quad (14)$$

According to the above derivation, Algorithm 1 gives a detailed description of the DLSR algorithm.

**3.3. The Inductive Transfer Learning Algorithm Based on DLSR.** Most inductive transfer learning algorithms are implemented by directly learning from the data in the source domain through some classes. However, in the paper, we

used a knowledge-based inductive transfer learning framework instead of raw data to study inductive transfer learning methods based on source domain knowledge. Inspired by this, an inductive transfer learning algorithm based on DLSR was introduced. Its objective function is

$$\min_{Z, p} \|XZ + ep^T - (Y + B \Theta W)\|_F^2 + \lambda \|Z\|_F^2 + \eta \|Z - Z_s\|^2. \quad (15)$$

*s.t.*  $W \geq 0$

It can be found from the formula (15) that the first two items directly inherit the DLSR for learning in the target domain. The third item is used to transfer the knowledge  $Z_s$  of the source domain to the target domain. When  $\eta = 0$ , DLSR is DSLR.

In short, TDLSR summarizes DLSR from the perspective of transfer learning, but it has more transfer learning capabilities than DLSR and has better applicability. Similar to DLSR, the objective function of TDLSR can also be solved using an alternate optimization strategy. The specific derivation process is as follows:

- (1) Fix  $W$ , and update  $Z$  and  $p$ . Let  $L = Y + B\Theta W$ ; formula (15) can be rewritten as

$$J(Z, p) = \arg \min_{Z, p} \|XZ + ep^T - L\|_F^2 + \lambda \|Z\|_F^2 + \eta \|Z - Z_s\|_F^2. \quad (16)$$

According to the optimization theory, we take a partial derivative of  $p$ , namely

$$\begin{aligned} \frac{\partial J(Z, p)}{\partial p} &= 0, \\ \implies Z^T X^T e + p e^T e - L^T e &= 0, \\ \implies p &= \frac{(L^T e - Z^T X^T e)}{N}. \end{aligned} \quad (17)$$

Furthermore, we find the partial derivative of  $Z$ , we can get

$$\begin{aligned} \frac{\partial J(Z)}{\partial Z} &= 0, \\ \implies 2X^T XZ - 2X^T (ep^T - L) + 2\lambda Z + 2\eta(Z - Z_s) &= 0, \\ \implies X^T XZ - X^T \left( e \frac{(e^T L - e^T XZ)}{N} - L \right) + \lambda Z \\ &\quad + \eta(Z - Z_s) = 0, \\ \implies X^T \left( I_N - \frac{ee^T}{N} \right) XZ - X^T \left( I_N - \frac{ee^T}{N} \right) L \\ &\quad + (\lambda + \eta)Z - \eta Z_s = 0, \\ \implies Z &= (X^T HX + (\lambda + \eta)I_d)^{-1} (X^T HL + \eta Z_s) \end{aligned} \quad (18)$$

- (2) Fix  $Z$  and  $p$ , and update  $W$ . Let  $G = XZ + ep^T - Y$ ; then,  $W$  can be solved as:

$$\begin{aligned} \arg \min_W &\|G - B\Theta W\|_F^2. \\ s.t. &W \geq 0 \end{aligned} \quad (19)$$

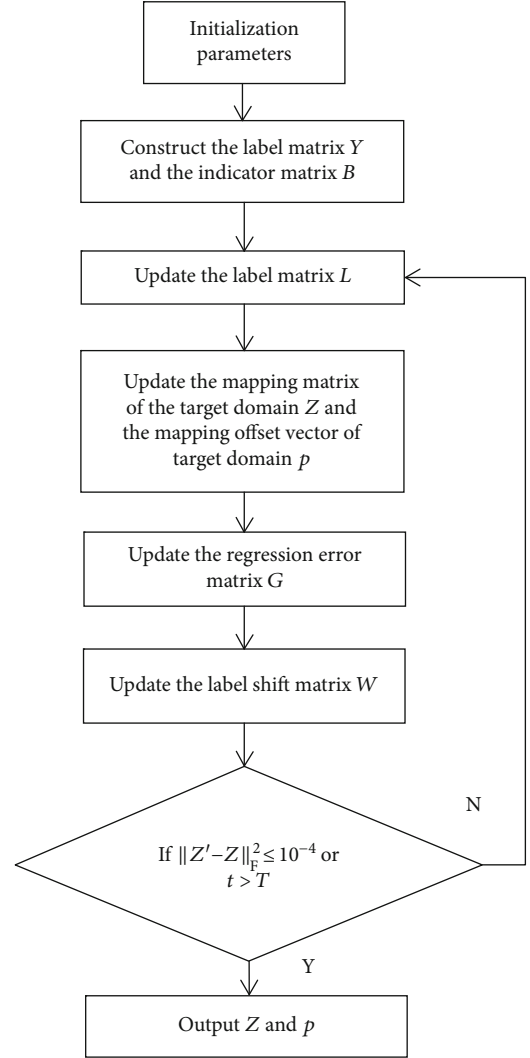


FIGURE 5: The process of the TDLSR algorithm.

Similar to the way of solving  $W$  in DLSR, the final solution formula of  $W$  is:

$$W = \max(B\Theta G, 0). \quad (20)$$

According to the above derivation process, Algorithm 2 gives a detailed description of the TDLSR algorithm.

To understand the TDLSR algorithm more clearly, Figure 5 shows the specific process of the TDLSR.

## 4. Experimental Research

**4.1. The Environment and Parameter Settings.** To evaluate the performance of the TDLSR, a lot of experiments were carried out based on EEG signal recognition. Experiments will be conducted from two aspects: (1) the comparison experiments with classic LSR, SRC, RLR, DLSR, and SVM and (2) the comparison experiments with related methods with transform ability, such as AuSVM, Tr-Adaboost,

TABLE 2: The parameter settings in the experiments.

Methods	Parameter settings
LSR	The regularization parameter $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10, 10^2\}$
SRC	The regularization parameter $\alpha \in \{10^{-3}, 5 \times 10^{-3}, 10^{-2}, 2 \times 10^{-2}, 5 \times 10^{-2}, 10^{-1}\}$
RLR	The regularization parameter $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0\}$ ; the heart parameter $\delta \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$
DLSR	The regularization parameter $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10, 10^2\}$
SVM	The penalty factor $c \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$ , the kernel parameter $\delta \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$
TSVM	The Lagrange multiplier upper bound $c \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$ , the kernel parameter $\delta \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$
Au-SVM	The penalty factor $c \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$ , the kernel parameter $\delta \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$
Tr-Adaboost	The penalty factor $c \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$ , the kernel parameter $\delta \in \{2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^3, 2^4, 2^5\}$
LMPROJ	The regularization parameters $\tau \in \{0.25, 0.1, \dots, 200\}$ , $\sigma \in \{0.25, 0.1, \dots, 200\}$ , the kernel parameter $\delta \in \{0.25, 0.1, \dots, 200\}$
TDLSR	The regularization parameter $\lambda \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10, 10^2\}$ , $\eta \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10, 10^2\}$

TABLE 3: The construction of experimental datasets.

The subdataset	Source domain	Target domain
D1	AE-each 75	AE-each 25
D2	BDE-each 75	BDE-each 25
D3	ABCD-each 75	ABCD-each 25
D4	BCDE-each 75	BCDE-each 25
D5	BE-each 75	BC-each 25
D6	ACE-each 75	BCE-each 25
D7	ADE-each 75	BDE-each 25
D8	ACDE-each 75	BCDE-each 25

TSVM, and LMPROJ. The hardware environment used in the experiment is Intel (R) Core(TM) I7-9700 3.0GHz×2, 8G RAM, and the software environment is Windows 10 64bit, MATLAB R2012b. Table 2 shows the parameter settings of the above algorithms.

**4.2. Experimental Dataset.** In the experiments, the EEG dataset used for epilepsy is the Bonn dataset [38, 39], which was collected by Andrzejak et al. at an epilepsy center at the University of Bonn. The EEG dataset contains five datasets, denoted by A to E. This dataset compares the EEG of the patient during the onset and nononset period with the EEG of the normal person. Dataset A and dataset B are EEG signals collected by healthy testers with their eyes open and closed. Dataset C and dataset D are the EEG signals collected by epilepsy patients outside and inside the lesion during the seizure period, and dataset E is the EEG signals collected by the patients in dataset C and dataset D during the seizure. Each of the 5 datasets includes 100 single-channel EEGs (that is 100 samples), the sampling frequency is 173.61 Hz, each segment of the signal collects 4097 frequency points, and each EEG segment lasts 23.6 s.

In our experiments, the use of the Bonn dataset is significantly different from many previous works. We constructed 8 subdatasets from the original 5 datasets to simulate different scenarios in the experiments. The source and target

TABLE 4: Experimental results based on nontransfer learning algorithms.

Datasets	LSR	SRC	RLR	DLSR	SVM
D1	0.87	0.84	0.89	0.90	0.82
D2	0.79	0.72	0.81	0.80	0.74
D3	0.71	0.62	0.72	0.72	0.68
D4	0.71	0.64	0.70	0.72	0.66
D5	0.82	0.78	0.83	0.84	0.78
D6	0.73	0.60	0.72	0.74	0.61
D7	0.70	0.60	0.70	0.72	0.59
D8	0.63	0.53	0.65	0.64	0.50

TABLE 5: Experimental results based on transfer learning algorithms.

Datasets	Au-SVM	TSVM	Tr-Adaboost	LMPROJ	TDLSR
D1	0.90	0.92	0.89	0.92	0.96
D2	0.82	0.85	0.82	0.84	0.90
D3	0.72	0.75	0.73	0.76	0.83
D4	0.72	0.75	0.71	0.78	0.82
D5	0.85	0.86	0.84	0.87	0.91
D6	0.75	0.78	0.75	0.80	0.83
D7	0.71	0.78	0.74	0.78	0.83
D8	0.68	0.70	0.66	0.70	0.73

domains of the 8 subdatasets of experiments are composed of the partial data extracted from the 5 sets. We randomly select 75% of the data from a certain dataset as the source domain, and the remaining 25% as the target domain. The sample data of source and target domains of subdatasets 1-4 are derived from the same distribution, but the samples taken are different. The sample data of the source domain and the target domain of subdatasets 5-8 have different distributions. Finally, for the data in the target domain, 20% is randomly selected for testing, and the remaining 80% is used for

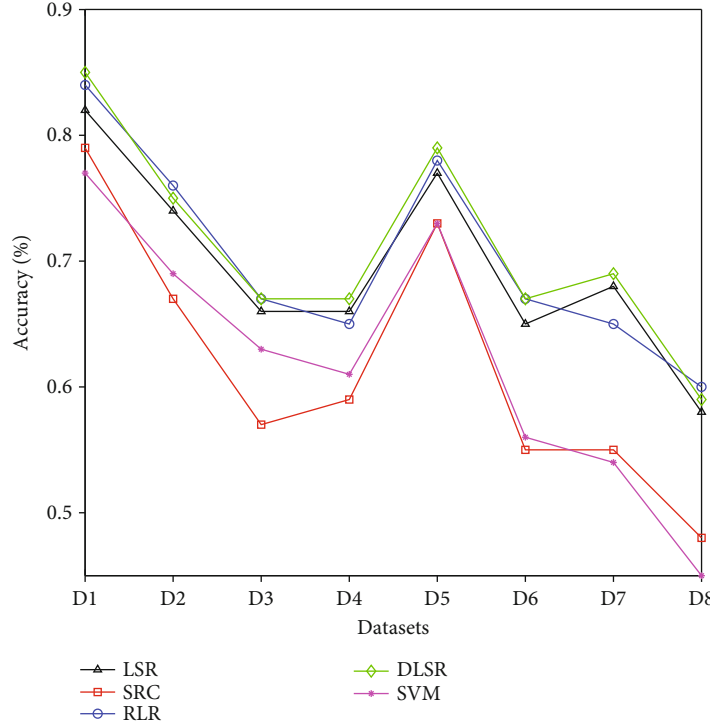


FIGURE 6: The average classification accuracy of all nontransfer learning algorithms.

training. In addition, the training and test datasets do not contain the same samples and are independent of each other. Table 3 lists the detailed information of the 8 subdatasets.

**4.3. Experimental Results and Analysis.** To verify the effectiveness of TDLSR in epilepsy EEG data recognition, the comparative experiments were conducted among the classic algorithms. The experimental results are shown in Tables 4 and 5, respectively.

As shown in the above experimental results, the conclusions are summarized as follows:

- (1) In the case of the same distribution of the source domain dataset and the target domain dataset, both the nontransfer learning algorithms and the transfer learning algorithms achieve good classification results. For the datasets with certain differences in distribution, the classification effects of algorithms without transforming abilities are quite different, and the experimental results of the algorithms in Table 5 are generally better than results of the algorithms in Table 4
- (2) As a whole, the performance of TDLSR introduced in this paper is obviously superior to all other algorithms. It means that by using the knowledge transferred from the source domain to the target domain, the TDLSR algorithm obtains better performance and becomes effective for epilepsy EEG recognition
- (3) Comparing the performance of the algorithms in Table 4, it can be found that DLSR has the best performance, while SRC and SVM have poor performance. This is because DLSR expands the ability to distinguish between classes by using the class information in the label space for the classification task. As the results shown in Table 5, the TDLSR algorithm performs best. This is because it not only inherits the advantages of DLSR by increasing the interval between different classes but also transfers more useful information from the source domain to the target domain. It has stronger transfer learning ability than several other algorithms. And because the TDLSR method requires fewer parameters to adjust, it is easier to use, and the stability and fault tolerance are stronger than the transfer learning algorithms such as LMPROJ.

In addition, to further observe and compare the overall classification performance of all algorithms, Figures 6 and 7 also give the average classification accuracy of each algorithm on all datasets.

Observing Figures 6 and 7, it can be seen that the classification accuracy of the algorithms on the same distributed dataset is higher than the result on the different distributed dataset. Secondly, compared with the traditional nontransfer learning algorithms, the algorithm with transfer learning ability has more advantages in classification performance, and the TDLSR has a significant improvement in classification performance. Finally, for all datasets with different classes, the performance of all algorithms decreases continuously



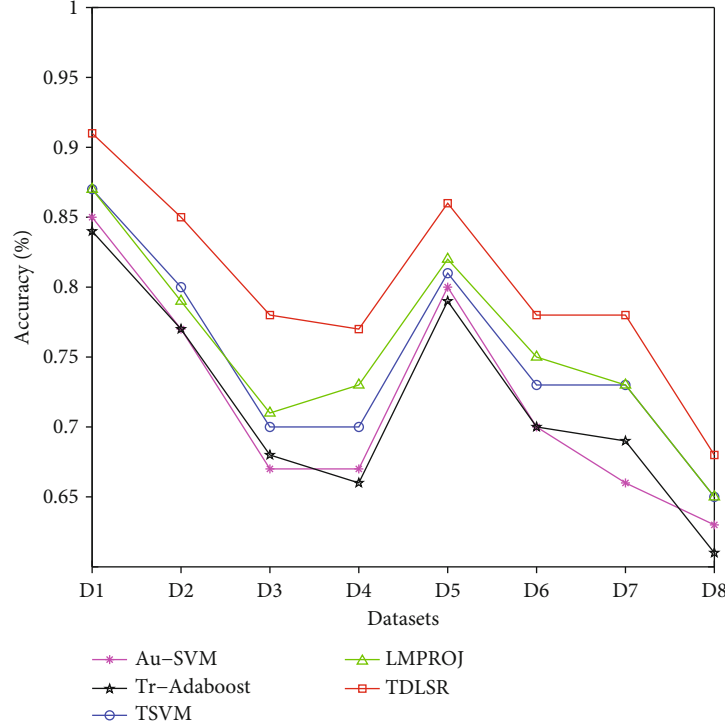


FIGURE 7: The average classification accuracy of all transfer learning algorithms.

as the number of categories increases. This is because as the number of classes increases, the information in the label space becomes more complicated, and learning the information in the label space becomes more difficult.

Meanwhile, to further verify the reliability and stability of the algorithms, we randomly added 15% white Gaussian noise to the data in the source domain to prove that the algorithm in this paper can adapt to more complex scenarios. The experimental results are shown in Tables 6 and 7, respectively.

Through the above experimental results, it can be found that the classification performance of all nontransfer learning algorithms under noisy conditions decreases more. The reason is that they cannot obtain useful knowledge from noisy data (source domain) for classification. However, all the algorithms in Table 7 can transfer some useful knowledge from the source domain for classification in the target domain, so their performance is better than the results in Table 6.

In summary, the TDLSR algorithm introduced in the paper is superior to the other algorithms in the detection of epilepsy EEG signals. And, it is easy to learn and train, has high stability, and shows certain advantages compared with other intelligent algorithms.

## 5. Conclusion

To solve the problem of serious shortage of training data in the current scene and improve the accuracy of classification, a novel DLSR-based inductive transfer learning algorithm (TDLSR) was introduced for the detection of epilepsy EEG signals. It can take advantage of both inductive transfer learn-

TABLE 6: Experimental results based on nontransfer learning algorithms with 15% noise in the source domain.

Datasets	LSR	SRC	RLR	DLSR	SVM
D1	0.83	0.80	0.84	0.85	0.77
D2	0.74	0.67	0.76	0.76	0.70
D3	0.66	0.59	0.68	0.69	0.65
D4	0.66	0.61	0.67	0.68	0.63
D5	0.78	0.74	0.80	0.80	0.74
D6	0.65	0.56	0.69	0.69	0.58
D7	0.64	0.57	0.67	0.71	0.56
D8	0.60	0.50	0.62	0.61	0.47

TABLE 7: Experimental results based on transfer learning algorithms with 15% noise in the source domain.

Datasets	Au-SVM	TSVM	Tr-Adaboost	LMPROJ	TDLSR
D1	0.87	0.86	0.89	0.89	0.93
D2	0.79	0.79	0.82	0.81	0.87
D3	0.69	0.70	0.72	0.73	0.80
D4	0.69	0.68	0.72	0.75	0.79
D5	0.82	0.81	0.83	0.84	0.88
D6	0.72	0.72	0.75	0.77	0.80
D7	0.68	0.71	0.75	0.75	0.80
D8	0.65	0.63	0.67	0.67	0.70

ing and DLSR. On the one hand, it can not only protect the security of the source domain data but also use the data of the target domain and the knowledge of the source domain

to get better performance. On the other hand, it inherits the DLSR's characteristics of being more suitable for real classification scenarios by expanding the interval between different categories. Therefore, compared with DLSR, the new algorithm not only enhances the ability of transfer learning but also ensures that the model is more reasonable. The results reflect that the improved algorithm has more advantages in epilepsy EEG signal classification compared with the traditional algorithms. However, it is found that the results are easily affected by the parameters. In a word, the quality of the parameter selection will directly affect the final detection accuracy. Therefore, to obtain higher detection accuracy, it is worth further study the characteristics of the EEG signal to guide the setting range of the parameters in the transfer learning algorithm.

### Data Availability

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

### Conflicts of Interest

The authors declare no conflict of interest regarding the publication.



### References

- [1] K. S. Wilcox, P. J. West, and C. S. Metcalf, "The current approach of the Epilepsy Therapy Screening Program contract site for identifying improved therapies for the treatment of pharmacoresistant seizures in epilepsy," *Neuropharmacology*, vol. 166, p. 107811, 2020.
- [2] L. Harris and H. Angus-Leppan, "Epilepsy: diagnosis, classification and management," *Medicine*, 2020.
- [3] T. Gandhi, B. K. Panigrahi, M. Bhatia, and S. Anand, "Expert model for detection of epileptic activity in EEG signature," *Expert System with Applications*, vol. 37, no. 4, pp. 3513–3520, 2010.
- [4] R. Sahu, S. R. Dash, L. A. Cacha, R. R. Poznanski, and S. Parida, "Epileptic seizure detection: a comparative study between deep and traditional machine learning techniques," *Journal of Integrative Neuroscience*, vol. 19, no. 1, pp. 1–9, 2020.
- [5] S. Panahi, T. Shirzadian, M. Jalili, and S. Jafari, "A new chaotic network model for epilepsy," *Applied Mathematics and Computation*, vol. 346, no. 1, pp. 395–407, 2019.
- [6] S. Panahi, Z. Aram, S. Jafari, J. Ma, and J. C. Sprott, "Modeling of epilepsy based on chaotic artificial neural network," *Chaos, Solitons & Fractals*, vol. 105, pp. 150–156, 2017.
- [7] S. T. George, M. S. P. Subathra, N. J. Sairamya, L. Susmitha, and M. J. Premkumar, "Classification of epileptic EEG signals using PSO based artificial neural network and tunable-Q wavelet transform," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 2, pp. 709–728, 2020.
- [8] S. P. Kumar, N. Sriraam, P. G. Benakop, and B. C. Jinaga, "Entropies based detection of epileptic seizures with artificial neural network classifiers," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3284–3291, 2010.
- [9] E. D. Uneyli, "Lyapunov exponents/probabilistic neural networks for analysis of EEG signals," *Expert Systems with Applications*, vol. 37, no. 2, pp. 985–992, 2010.
- [10] M. Sharma, S. Patel, and U. R. Acharya, "Automated detection of abnormal EEG signals using localized wavelet filter banks," *Pattern Recognition Letters*, vol. 133, pp. 188–194, 2020.
- [11] S. Raghu, N. Sriraam, Y. Temel, S. V. Rao, A. S. Hegde, and P. L. Kubben, "Performance evaluation of DWT based sigmoid entropy in time and frequency domains for automated detection of epileptic seizures using SVM classifier," *Computers in Biology and Medicine*, vol. 110, pp. 127–143, 2019.
- [12] A. Temko, E. Thomas, W. Marnane, G. Lightbody, and G. Boylan, "EEG-based neonatal seizure detection with support vector machines," *Clinical Neurophysiology*, vol. 112, no. 3, pp. 464–473, 2011.
- [13] N. Mahmoodian, A. Boese, M. Friebe, and J. Haddadnia, "Epileptic seizure detection using cross-bispectrum of electroencephalogram signal," *Seizure: European Journal of Epilepsy*, vol. 66, pp. 4–11, 2019.
- [14] S. Ramakrishnan and A. S. Muthanantha Murugavel, "Epileptic seizure detection using fuzzy-rules-based sub-band specific features and layered multi-class SVM," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 1161–1176, 2019.
- [15] H. S. Alaei, M. A. Khalilzadeh, and A. Gorji, "Optimal selection of SOP and SPH using fuzzy inference system for on-line epileptic seizure prediction based on EEG phase synchronization," *Australasian Physical & Engineering Sciences in Medicine*, vol. 42, no. 4, pp. 1049–1068, 2019.
- [16] J. M. Lichtenberg and O. Simsek, "Simple regression models," in *Imperfect Decision Makes: Admitting Real-World Rationality*, pp. 13–25, Barcelona, 2017.
- [17] N. Imran, R. Tongeri, and M. Bennamoun, "Linear regression for face classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, 2011.
- [18] B. Litt, R. Esteller, J. Echaz et al., "Epileptic seizures may begin hours in advance of clinical onset," *Neuron*, vol. 30, no. 1, pp. 51–64, 2001.
- [19] A. Sharmila and P. Geethanjali, "Effect of filtering with time domain features for the detection of epileptic seizure from EEG signals," *Journal of medical engineering & technology*, vol. 42, no. 3, pp. 217–227, 2018.
- [20] J. Gotman, "Automatic seizure detection: improvements and evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 76, no. 4, pp. 317–324, 1990.
- [21] A. Sharmila and P. Mahalakshmi, "Wavelet-based feature extraction for classification of epileptic seizure EEG signal," *Journal of Medical Engineering & Technology*, vol. 41, no. 8, pp. 670–680, 2017.
- [22] I. Goker, O. Osman, S. Ozekes, M. B. Baslo, M. Ertas, and Y. Ulgen, "Classification of juvenile myoclonic epilepsy data acquired through scanning electromyography with machine learning algorithms," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2705–2711, 2012.
- [23] O. Faust, U. R. Acharya, L. C. Min, and B. H. Spath, "Automatic identification of epileptic and background EEG signals using frequency domain parameters," *International journal of neural systems*, vol. 20, no. 2, pp. 159–176, 2012.
- [24] R. Hopfengartner, F. Kerling, V. Bauer, and H. Stefan, "An efficient, robust and fast method for the offline detection of epileptic seizures in long-term scalp EEG recordings," *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2332–2343, 2007.
- [25] A. Maheshwari, "Rodent EEG: expanding the spectrum of analysis," *Epilepsy Currents*, vol. 20, no. 3, pp. 149–153, 2020.

- [26] S. Ashok and G. Purusothaman, "A review on the pattern detection methods for epilepsy seizure detection from EEG signals," *Biomedical Engineering/Biomedizinische Technik*, vol. 64, no. 5, pp. 507–517, 2019.
- [27] R. B. Pachori, "Analysis of normal and epileptic seizure EEG signals using empirical mode decomposition," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 373–381, 2011.
- [28] R. J. Martis, U. R. Acharya, H. J. Tan et al., "Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals," *International Journal of Neural Systems*, vol. 22, no. 6, 2012.
- [29] V. Bajaj and R. B. Pachri, "Epileptic seizure detection based on the instantaneous area of analytic intrinsic mode functions of EEG signals," *Biomedical Engineering Letters*, vol. 3, no. 1, pp. 17–21, 2013.
- [30] R. B. Pachori, R. Sharma, and S. Patidar, "Classification of normal and epileptic seizure EEG signals based on empirical mode decomposition," in *Complex system modelling and control through intelligent soft computations*, pp. 367–388, Springer, Cham, 2015.
- [31] C. Guerrero-Mosquera, A. M. Trigueros, J. I. Franco, and Á. Navia-Vázquez, "New feature extraction approach for epileptic EEG signal detection using time-frequency distributions," *Medical & Biological Engineering & Computing*, vol. 48, no. 4, pp. 321–330, 2010.
- [32] E. Ebrahimzadeh, S. M. Alavi, A. Bijar, and A. Pakkhesal, "A novel approach for detection of deception using Smoothed Pseudo Wigner-Ville Distribution (SPWVD)," *Journal of Biomedical Science and Engineering*, vol. 6, no. 1, pp. 8–18, 2013.
- [33] H. Polat, M. U. Aluçlu, and M. S. Özerdem, "Evaluation of potential auras in generalized epilepsy from EEG signals using deep convolutional neural networks and time-frequency representation," *Biomedical Engineering / Biomedizinische Technik*, 2019.
- [34] M. A. Al-Manie and W. J. Wang, "A modified S-transform for EEG signals analysis," *International Journal of Computational Methods*, vol. 12, no. 3, 2015.
- [35] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 3, pp. 187–194, 2005.
- [36] A. Arnold, R. Nallapati, and W. W. Cohen, "A comparative study of methods for transductive transfer learning," in *Proceedings of the 7th IEEE International Conference on Data Mining Workshops(ICDMW'07)*, pp. 77–82, IEEE Computer Society, Washington, DC, USA, 2007.
- [37] S. Xiang, F. Nie, G. Meng, C. Pan, and C. Zhang, "Discriminative least squares regression for multiclass classification and feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1738–1754, 2012.
- [38] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [39] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.

## Research Article

# An EEG Database and Its Initial Benchmark Emotion Classification Performance

Ayan Seal <sup>1,2</sup> Puthi Prem Nivesh Reddy,<sup>1</sup> Pingali Chaithanya,<sup>1</sup> Arramada Meghana,<sup>1</sup> Kamireddy Jahnvi,<sup>1</sup> Ondrej Krejcar <sup>2,3</sup> and Radovan Hudak<sup>4</sup>

<sup>1</sup>PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur 482005, India

<sup>2</sup>Faculty of Informatics and Management, Center for Basic and Applied Research, University of Hradec Kralove, Rokitanskeho 62, Hradec Kralove, 50003, Czech Republic

<sup>3</sup>Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

<sup>4</sup>Department of Biomedical Engineering and Measurement, Faculty of Mechanical Engineering, Technical University of Kosice, Letna 904200 Kosice, Slovakia

Correspondence should be addressed to Ayan Seal; [ayanseal30@ieee.org](mailto:ayanseal30@ieee.org)

Received 6 June 2020; Accepted 6 July 2020; Published 3 August 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Ayan Seal 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.

Human emotion recognition has been a major field of research in the last decades owing to its noteworthy academic and industrial applications. However, most of the state-of-the-art methods identified emotions after analyzing facial images. Emotion recognition using electroencephalogram (EEG) signals has got less attention. However, the advantage of using EEG signals is that it can capture real emotion. However, very few EEG signals databases are publicly available for affective computing. In this work, we present a database consisting of EEG signals of 44 volunteers. Twenty-three out of forty-four are females. A 32 channels CLARITY EEG traveler sensor is used to record four emotional states namely, happy, fear, sad, and neutral of subjects by showing 12 videos. So, 3 video files are devoted to each emotion. Participants are mapped with the emotion that they had felt after watching each video. The recorded EEG signals are considered further to classify four types of emotions based on discrete wavelet transform and extreme learning machine (ELM) for reporting the initial benchmark classification performance. The ELM algorithm is used for channel selection followed by subband selection. The proposed method performs the best when features are captured from the gamma subband of the FP1-F7 channel with 94.72% accuracy. The presented database would be available to the researchers for affective recognition applications.

## 1. Introduction

Emotion is a mental condition [1] of a person. It is directly or indirectly connected to high intensity and high hedonic content namely, thoughts, feelings, behavioral responses, and a level of happiness or unhappiness [2]. The objective of emotion recognition is to identify a human's temporal emotional state based on facial expressions and verbal expressions automatically. Multimedia elements such as texts, audio, and video are also used frequently to induce human expressions for emotion recognition [3]. Body movements with gestures and speech are also combined with facial expressions to classify emotion more accurately [4]. Unavoidably, feelings play

a significant role in making relations with others. It also helps to increase interaction with computers. Affective computing is a branch of study, where human interaction with computers using emotion and its applications are taught [5]. Affective computing finds its applications spanning in various areas like medicine, e-learning, monitoring, marketing, entertainment, and law. One can find some real-life examples of affective computing such as counseling and determining client's medical state, determining patients sentiment and relief level about the treatment during healthcare, analyzing the emotion and thereby adjusting the learning technique and presentation according to the style of a learner in the case of e-learning, remote health checkup of senior citizens,

expanding means of communication and assisting the people with autism, and determining fatigue in the case of driving and alerting in advance. Since the emotional state of a human may affect attentiveness, problem-solving, and policy-making skills, the vision of affective computing is to build systems that are capable to identify and influence a person's emotion to increase yield and effectiveness of working with computers. Electroencephalography (EEG) is an electrophysiological checking strategy to record the electrical activity of the human brain. It is ordinarily noninvasive, with the electrodes placed along the scalp. EEG estimates voltage variations because of ionic current within the neurons of the cerebrum. The state-of-the-art emotion recognition methods [6, 7] can be broadly divided into three types namely, knowledge-based methods, statistical approaches, and hybrid techniques [8]. In knowledge-based, emotion detection is carried out by utilizing domain understanding and the semantic and syntactic features of language [9] using WordNet, SenticNet [10], ConceptNet, and EmotiNet [11]. However, it is unable to handle concept nuances and complex linguistic rules [8]. In statistical techniques, generally different supervised machine learning algorithms such as Support Vector Machines [12], Naive Bayes, and Maximum Entropy [13] are considered in emotion classification using a large set of annotated data. The classification accuracy is higher in the case of statistical methods, and it depends on the sufficiently large training set. Few works consider both knowledge-based techniques and statistical methods in emotion identification by exploiting complementary features [14, 15]. The classification accuracy provided by hybrid approaches is higher as compared to the previous two methods. However, it is computationally more challenging [9]. Nowadays, researchers are using physiological signals such as EEG and electrocardiogram (ECG) signals for emotion recognition [16, 17]. The different sources of physiological signals in the human body are shown in Figure 1.

However, when compared to emotion portrayed through facial expressions, emotion captured through EEG signals is more reliable because someone can fake his/her facial expressions; in other words, he/she may feel a different emotion from within and act and express something else, or in some cases, he/she might not be able to express through gestures or facial expressions the feeling they are experiencing inside. Interested readers can go through [19] for knowing other advantages of EEG signals over facial expression in emotion recognition. Data is an integral component of any emotion recognition task. It is rare to get publicly available annotated data which is essential for training and testing of machine learning algorithms. Some of the publicly available multimodal sources of data in the form of multimedia contents such as texts, audio, videos, or physiological signals to perform emotion recognition research are briefly described as follows: Belfast facial expression database [20] was created mainly for the investigation of sex differences, cultural differences, and individual differences in the expression and perception of a wide range of emotions using short TV programs and interview recordings [20]. HUMAINE facial expressions database [21] consists of almost all audio-visual recordings in diverse emotion-rich scenarios. SEMAINE is

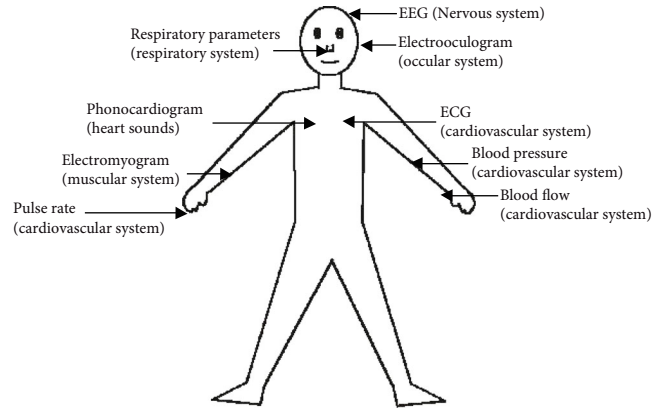


FIGURE 1: Sources of physiological signals [18].

also an audio-visual recording [22] between a person and a virtual assistant and comprises of annotated emotion like angry, happy, fear, disgust, sadness, contempt, and amusement. IEMOCAP is also audio-visual recordings of dyadic sessions between actors and consists of emotion annotations such as happiness, anger, sadness, frustration, and neutral state [23]. eINTERFACE considers pleasure, irritation, sorrow, amazement, disgust, and panic in the form of audio-visual recordings from seven nationalities [24]. DEAP [16] and DREAMER [17] databases consist of EEG, ECG, and face video recordings [16]. This database is built, while people were watching film clips. All the data are annotated properly in terms of valence, arousal, and dominance of people. In [25], Alakus et al. offered an EEG database called GAMEEMO for various games.

It is clear from the literature that a few publicly available EEG signals databases exist. So, the salient contributions of the proposed study are as follows:

- (i) A new database is presented, which consists of four emotions namely, happy, sad, neutral, and fear in the form of EEG signals of 44 subjects
- (ii) The captured EEG signals are classified based on discrete wavelet transform (DWT) and extreme learning machine (ELM).
- (iii) The ELM algorithm is used for channel selection followed by subband selection

The rest of the work is as follows: the background study related to the human brain and emotion is discussed in Section 2. Section 3 presents a new database for affective computing. Materials and methods used for noise removal, feature extraction, and classifier are discussed in Section 4. Section 5 presents the experimental results. Finally, the conclusion is drawn in Section 6.

## 2. Background Study

In the following section, we give a brief introduction of the human brain and the lobes present in it which are responsible for emotion and also about the characteristics of EEG signals.



**2.1. Human Brain.** The brain is the command center for the human nervous system. It is responsible for all of the body's activities. It receives sensory input from the many sensory receptors. Then, it processes, interprets, and decides what actions should be taken. It also sends the output to the muscles. The brain has three main parts viz., cerebellum, brain stem, and cerebrum. The cerebellum accepts data from the many sensory receptors, the spinal cord, and other parts of the brain and then manages motor movements. It also involved involuntary movements such as balance, posture, coordination, and speech, governing smooth muscular function. The brain stem manages the movement of messages between the brain and the other parts of the body. The basic body activities like breathing, swallowing, heart rate, blood pressure, and consciousness are managed by stem. On the other hand, the cerebrum is the largest part of the human brain, and it is further divided into hemispheres viz., left and right. Each hemisphere has four lobes namely, frontal, temporal, occipital, and parietal as in Figure 2. The cerebrum is mainly responsible for the action and thought. The various regions of cerebrum control various actions as described below:

- (i) The frontal region is responsible for the movement of the body, speech, concentration, processing memories from the limbic region, emotional reaction, problem-solving, the meaning of words, personality, and planning. In the frontal lobe region, neurons of dopamine-sensitive type are located which help with the feelings of attention, memory, reward, pleasure, and planning. It contains a part of the limbic region which is responsible for some basic emotions as stated above
- (ii) The temporal region is responsible for hearing, emotion, and regulate emotions like hunger, thirst, recognizing faces, and long-term memory which are responsible for storing, forming, and processing visual memory. It contains some part of the limbic region which is responsible for some basic emotion as stated above
- (iii) The occipital region is responsible to evaluate the distance, depth, size, helps in color recognition, interprets the visual stimuli, and transmits the visual data to the other regions of the brain and other regions respond to this information
- (iv) The parietal region is responsible to feel the sensations of touch, process, taste, and temperature

**2.2. Emotions.** Emotion is a state of feeling and change in both physical and psychological influence of our behavior which we people experience in our daily routine, and it comprises three components namely, abstract understanding, a physiological reaction, and a social or expressive reaction [26, 27]. Emotion has been portrayed as discrete and steady reactions to occasions outside or inner with hugeness for the life form [28]. Emotion plays an important role in human-human communication. Emotion is often intercon-

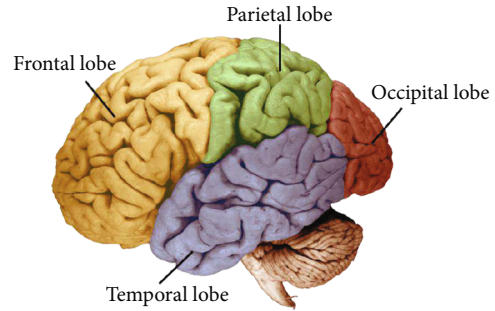


FIGURE 2: Sectional part of human brain [13].

nected with several factors like mood, personality, temperature, one's circumstances, or relationships with others. The emotional state can be expressed through two dimensions valence and arousal. Valence is the dimension that measures positive or negative effects due to emotion. Arousal is the dimension that measures how soothing or exciting the emotion is, and it is represented in terms of high and low. There are two different aspects of representing emotion. The first aspect shows that essential feelings have advanced through regular choices. Plutchik suggested eight primitive emotions anger, fear, sadness, disgust, surprise, curiosity, acceptance, and joy [29]. In the second aspect, given perception, the feelings are mapped into the valence, arousal, and dominance measurements. Valence goes from extremely positive sentiments to exceptionally negative or unpleasant to joy; excitement goes from states like drowsy to energized; lastly, predominance compares to the quality of the feeling [30]. A considerable lot of our regular exercises rely upon our emotional state. Assessing emotion is the key to understanding people. Consequently, human emotion acknowledgment is vital for the precise classification of human emotion. We will in this manner center around the strategies utilized for recognizing emotion.

As human-human association happens, here come brain-computer interface (BCI) systems. For a technologically invaded life, the brain-computer interaction system plays an important role. As machines impacted our lives from numerous points of view in the previous decade, machines ought to be able to find, translate, get to, and coordinate the emotional states of humans during an interaction. This human-machine association is made simple with BCI frameworks. Nowadays, these BCI frameworks have the capacity for emotion recognition. To have an effective emotion acknowledgment, a huge assortment of strategies like emotion recognition from outward appearances, the pitch of voice, signals from self-ruling sensory system, signals from EEG.

### 3. EEG Signals Database for Affective Computing

A very limited number of EEG signals databases are observed in the literature. This motivates us to develop an EEG signals database for research purposes. This section presents the different protocols that are considered while capturing signals

of a human. The detailed description of each protocol is discussed in subsequent sections.

**3.1. EEG Sensor.** EEG is a therapeutic imaging procedure that interprets scalp electrical movement produced by cerebrum structures, i.e., it measures voltage variances coming about because of ionic current streams inside the neurons of the cerebrum. EEG machine used here is EEG traveler braintech 32+ CMEEG-01, which is of 32 channels. This model has many characteristics such as the most friendly used toolbar, a facility of deleting events and comments, online/offline reformatting of gain, filter, and sweep speed. It also had EEG flash with high-intensity LED with a built-in rechargeable Li-ion battery. It is a highly advanced database management with different searching options. It operates with computer USB power.

Some of the specifications of this EEG traveler are it has 32 input channels and 32 display channels. The frequency band range is from 0.1 Hertz (Hz) to 100 Hz. The maximum impedance limit is  $10M\Omega$ . The sampling frequency is 1024 Hz for at ADC and 256 Hz for internal storage. It has a high-cut digital filter of the range 0.5 Hz to 100 Hz and low-cut digital filter of the range 0.1 Hz to 7 Hz. There is EEG paste which is used to adhere to the 32 electrodes or channels to the scalp for recording signals. A piece of EEG setup utilized for recording appears in Figure 3.

**3.2. EEG Signals.** EEG signals are principally used to track the action of the human brain utilizing electrical signals. The identification of emotional changes is normally made simple through EEG signals. Along these lines, we have picked this EEG flag-based approach for classification of human emotion to comprehend the essential feelings like happy, fear, sad, and neutral. The EEG recording is carried out with the assistance of the electrodes that are associated with the CLARITY EEG setup. These electrodes are little and level metallic discs that are joined to the scalp utilizing EEG gel/paste. For each action of the human body, cerebrum cells communicate with each other and these exercises of the brain will be recorded in a computer with the assistance of electrodes, and it is estimated in small scale voltage contrast. Voltage contrasts estimated at electrodes are little so they are digitized send to enhancer, and we can see the amplified data in terms of voltage. The EEG signals are helpful for doctors to diagnose disorders like a brain tumor, head injury, and encephalitis. This EEG signal is additionally used to measure the bloodstream level in the cerebrum all through the careful procedure and to screen the brain activity of the individual in unconsciousness/coma. EEG is considered a protected technique. It will not cause any distress or any sensation. It is not ok for a person who is experiencing seizure issue or epilepsy, since it might cause seizure during the procedure as blazing of light on the face, deep breathing is associated with the procedure. It might cause different issues relying upon the medical state of the person experiencing EEG recording.

**3.3. Brainwaves.** Generally, five types of electrical patterns or brain waves have been observed simultaneously by the

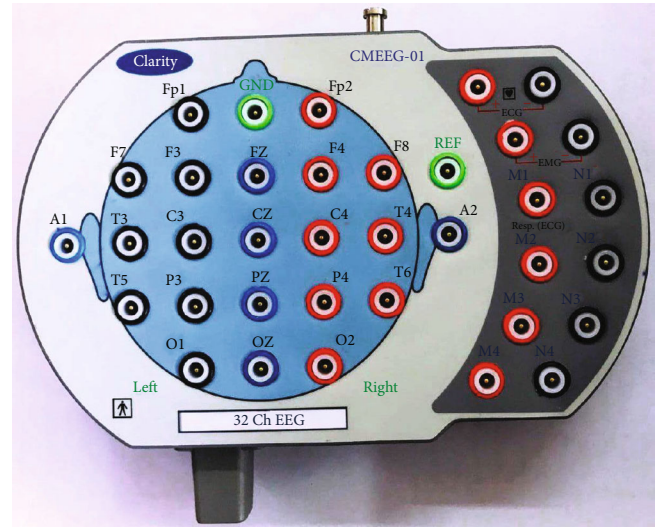


FIGURE 3: EEG machine for placing electrodes.

researchers across the cortex throughout a day in our awake state or even when we are asleep. These waves in order of lowest frequency to highest are as follows: delta (0-4 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-63 Hz). It is observed that the range of these waves fluctuates a few Hz among various studies. These waves jointly perform mental functioning effectively and efficiently. However, each brain wave has a unique role. One particular brain wave will be dominant at a time while performing a particular task, and it depends on the state of cognizance that a person is in. If one of the five types of brain waves is either oversynthesized and/or undersynthesized in our brain, it can cause problems. The detail description of each of the useful brain wave is as follows:

- (i) Delta waves are related to the deepest levels of pleasure and therapeutic rest, healing sleep. These are the slowest captured brain waves in people. Generally, higher levels are noticed in infants and kids. These are produced in a lesser amount as one grows old even during deep sleep. These are also observed too many of our unconscious bodily functions such as regulating heartbeat and digestion. Adequate levels of delta waves help to feel jubilant after we wake up refreshed. However, irregular delta activity may hamper normal learning ability or maintaining awareness
- (ii) Theta waves are linked in daydreaming and sleep. These are linked to us experiencing and feeling deep and raw emotions. However, higher levels of theta activity may result in an episode of depression and may make people "highly suggestible" according to the fact that they are in a deeply relaxed, semihypnotic state. These waves help to improve our intuition, creativity, and make us feel more natural. It is also involved in therapeutic rest. The range of these waves is very helpful until not generated in excess during waking

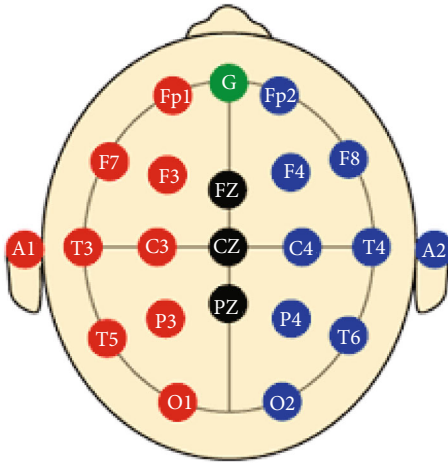


FIGURE 4: The international 10/20 system for electrode positioning.

- (iii) Alpha waves establish a “frequency bridge” between our cognizant reasoning (beta) and intuitive (theta) mind. These are known to enable quiet us to down and advance sentiments of deeper relaxation and substance. Beta waves assume a functioning part in arranging coordination and correspondence and do not happen until three years of age in person. Higher levels of beta activity and lower levels of alpha activity may result in an emotional state called “alpha blocking.”
- (iv) Beta waves are the low-amplitude high-frequency brain waves that are generally found in an awake person. These are produced when a person is incognizant in states like cognitive reasoning, calculation, reading, speaking, or thinking. Higher levels of beta waves may result in stress and anxiety. The beta activity of a person will increase when he or she drinks a stimulant called caffeine that is found especially in tea or coffee
- (v) In the area of neuroscience, a more recent discovery of brainwaves is gamma waves. Researchers are constantly exploring to know more about gamma waves. They have examined a strong relation between meditation and gamma waves. In addition to a healthy cognitive task, these perform more complex tasks. Gamma waves are important for learning, memory, and information processing. It is thought that the 40 Hz gamma wave is important for the binding of our senses in regard to perception and are involved in learning new information. Lower levels of gamma activity are observed in people with mentally challenged

**3.4. EEG Electrode Positioning.** The recordings of the EEG signals are performed according to the international standard 10/20 electrode system [31] as shown in Figure 4. Figure 4 depicts the 10/20 international standard electrode system in which the distance between any two adjoining electrodes can be either 10% or 20% of the total front to back or right

to the left of the skull. This 10/20 system is based on the connection between the position of the electrode and base zone of the cerebral cortex. Every electrode is demonstrated with a letter and a number which is utilized to recognize the lobe and hemisphere where it must be placed (F, T, P, C, and O letters alludes to frontal, temporal, parietal, central, and occipital lobes separately). There are numbers from 1 to 8 given to the electrodes among which 1, 3, 5, and 7 stand for the electrodes to be placed on the left hemisphere and 2, 4, 6, and 8 stand for the electrodes to be placed on the right hemisphere. “Z” refers to the electrode that is to be placed on the midline. Besides these electrodes, we also have ground and reference electrodes. Four anatomical reference points on head are utilized for the outright position of electrodes specifically nasion, inion, and both preauricular focuses. Nasion is the middle of forehead and nose, inion is the point from the posterior of the head which is the lowest of the skull, and preauricular focuses are those which are front to ear [32].

**3.5. Impedance Check.** After placing all the electrodes in their desired positions, we should check the impedance, which is whether all the electrodes are connected properly or not. If not, we should again check those electrodes which are not connected properly and connect them. The impedance is verified in the software which is provided with the CLARITY EEG machine. During the recording procedure, the impedance of the electrodes should not exceed the maximum impedance limit. The impedance corresponding to each electrode is shown in Figure 5. If the limit exceeds the maximum impedance level, color shown in the impedance check option of the software used to record the EEG signals changes from green to red. During this impedance check, the recording is stopped.

**3.6. Stimuli.** The stimuli selected in this experiment through various steps. First, we should consider the different types of emotion, and among them, we should select some which we want for classification. Here, four emotions namely, sad, happy, fear, and neutral are considered. Total stimuli have 12 videos, which means each emotion has 3 videos. Each video is of approximately 150 seconds which is 2 minutes and 30 seconds [33]. After watching every video, the person should relax for 15-20 seconds for neutralizing the emotion that he had felt during that period. During relaxing for a certain period, the person should close his/her eyes and relax to avoid the surroundings and prevent disturbances around them. The signals have been recorded in a close room with no lights and room temperature is 25-30 degrees Celsius. We have selected the stimuli of our required emotion from featured films based on the criteria like comedy scenes for happy emotion and emotional scenes for sad motion and scary scenes for fear emotion and neutral scenes for neutral emotion. There should be no noise or disturbance around the one who is watching the stimuli because, if any disturbance has taken place, the person might get deviated from the emotion, and the signals during that period of disturbance will be useless. The snapshot of happy stimuli, the reaction of a subject while watching that stimuli, and the



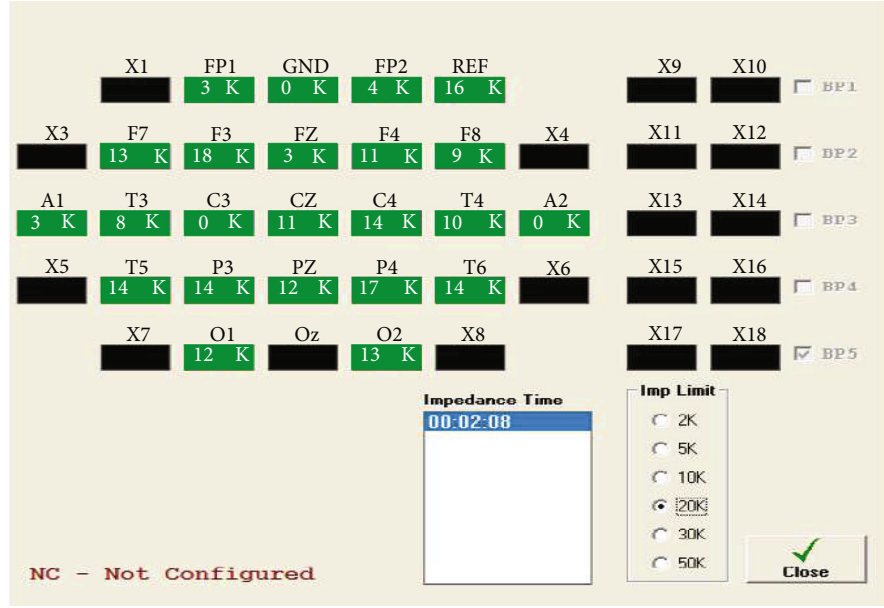
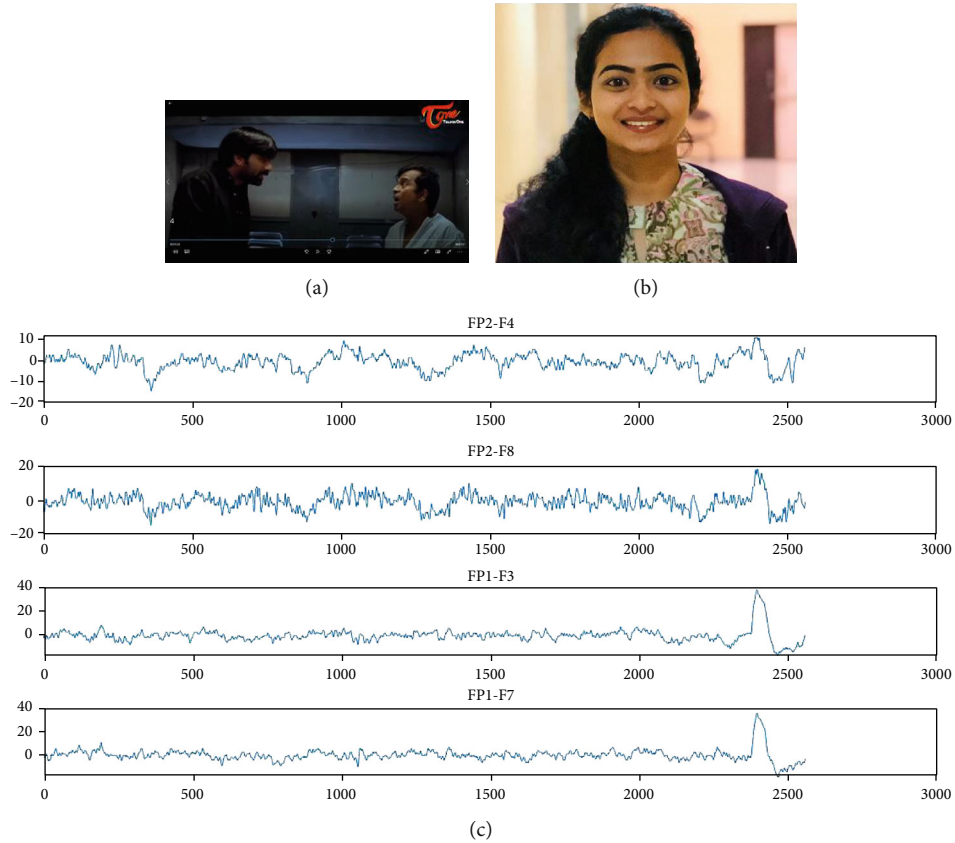


FIGURE 5: Impedance check.

FIGURE 6: (a) The snapshot of happy stimuli (b) reaction of a subject (c) captured EEG signals (x-axis and y-axis show time in second and amplitude in  $\mu V$ , respectively).

captured EEG signals of four channels namely, FP2-F4, FP2-F8, FP1-F3, and FP1-F7 are shown in Figure 6. Figure 7 shows the snapshot of sad stimuli, the subject's reaction, and the acquired EEG signals, respectively. The snapshot of fear stimuli, the reaction of a subject while watching that stimuli, and

the captured EEG signals are shown in Figure 8. Figure 9 illustrates the snapshot of neutral stimuli, the subject's reaction, and the acquired EEG signals, respectively. The horizontal and vertical axes of Figures 6(c), 7(c), 8(c), and 9(c) are mean sample sizes and amplitudes of EEG signal, respectively.



FIGURE 7: (a) The snapshot of sad stimuli (b) reaction of a subject (c) captured EEG signals ( $x$ -axis and  $y$ -axis show time in second and amplitude in  $\mu V$ , respectively).

**3.7. Pattern Matching.** These 12 videos are divided into 3 patterns in which each pattern consists of 4 videos of a different emotion. As the person after watching the first round of the pattern, he needs to match the videos according to the emotion provided on the left side. We should take care of creating these patterns while arranging the videos, because no pattern should be the same as of the previous one.

*Pattern 1:* neutral–sad–fear–happy (Figure 10).

*Pattern 2:* sad–happy–neutral–fear (Figure 11).

*Pattern 3:* happy–neutral–sad–fear (Figure 12).

This is because of the person who is matching the pattern should remember the four videos and match the pattern as to how he felt. For better results, the patterns should be varied. It is because if we change pattern we can make sure that the person who is matching the pattern is matching as how he felt rather than remembering the previous pattern and matching as it is in the new one.

**3.8. Database Summary.** No. of videos: 12. Video duration: 150 seconds (approximately). Selection method: manually. Total no. of participants: 44 (23 male + 21 female).

## 4. Materials and Methods

A schematic block diagram of the proposed system for classifying emotion is displayed in Figure 13. The proposed method includes EEG signal acquisition, removal of noise

and division of subbands, extraction of features, and classification. EEG signals acquisition procedure is presented in Section 3. The rest of the steps would be described in the subsequent sections.

**4.1. Discrete Wavelet Transform.** It is clear from Section 3.3 that EEG signals consist of mainly five types of brain waves. However, EEG signals also include noise. It is an unwanted signal that corrupts the signal of interest. So, the noise has to be removed from the raw signals to extract useful information. Thus, the first step in the processing of EEG signals is the filtering of raw signals. EEG signal is nonstationary, and multiresolution DWT is a widely used decomposition method for eliminating noise from a nonstationary signal [26]. The frequency of the EEG signal that we have captured is 256 Hz. Daubechies 8 DWT (db8) is adopted here to decompose raw EEG signal into detail coefficients and approximation coefficients up to 6<sup>th</sup> level. So, a total of 6 detail subbands (D1 to D6) and 1 approximation subband (A6) would be generated. These subbands are also called waves. Here, D1, D2, D3, D4, D5, D6, and A6 subbands are known as noise (125–256 Hz), noise (63–125 Hz), gamma (30–63 Hz), beta (13–30 Hz), alpha (8–13 Hz), theta (4–7 Hz), and delta (0–4 Hz), respectively. The various subbands obtained using “db8” DWT of a randomly selected channel from neutral emotion are shown in Figure 14.



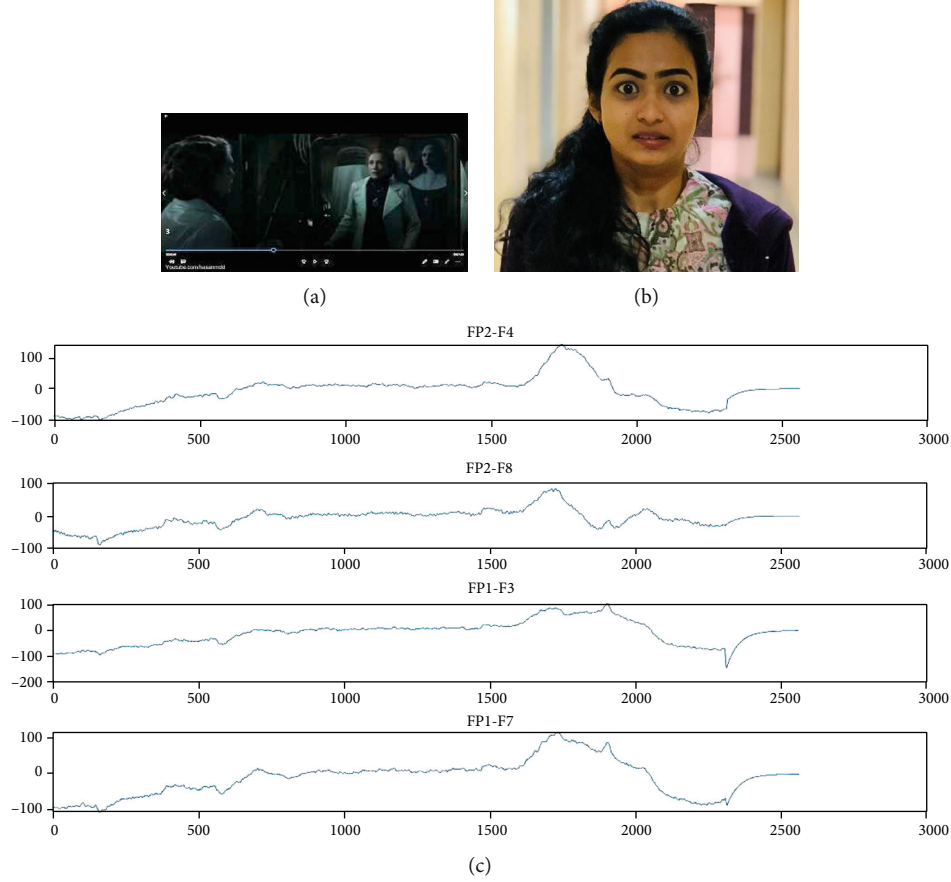


FIGURE 8: (a) The snapshot of fear stimuli (b) reaction of a subject (c) captured EEG signals (x-axis and y-axis show time in second and amplitude in  $\mu V$ , respectively).

**4.2. Extraction of Features.** The extraction of features is a process of unveiling hidden characteristics from the input signal. In other words, a set of features can represent an input signal. Moreover, this feature set illustrates a specific behavior or pattern depicted by the original input signal. Every signal is normalized to zero mean and unit variance. Ten popularly used features are extracted from each subband for emotion classification considered in this work, which is as follows:

(i) Average amplitude change

Average amplitude change is the mean value of the changes between two consecutive sample values. The value of AAC is estimated using Eq. (1).

$$f_1 = \frac{1}{C} \sum_{i=1}^{C-1} |s_{i+1} - s_i|, \quad (1)$$

where  $s_i$  is the  $i^{th}$  sampling point of an EEG signal and  $C$  is the total number of sample points. In our case, the value of  $C$  is 2560 because the sampling frequency of the device used is 256 Hz, and the duration of the signal is 10 s. It is also referred to as difference absolute mean value.

(ii) Activity

Activity is the measure of the mean power which is calculated by Eq. (2). It also represents the activeness of the signal. It also represents the variance of the EEG signal.

$$f_2 = \sigma_a^2, \quad (2)$$

where  $\sigma_a^2$  represents the variance of an EEG signal.

(iii) Absolute square root sum

The absolute square root sum is calculated using Eq. (3) which is the sum of the square root of each sample point in a signal.

$$f_3 = \sum_{i=1}^C \sqrt{s_i}, \quad (3)$$

where  $C$  is the total number of sample points.

(iv) Clearance factor

The clearance factor is calculated through Eq. (4), and it is a ratio of the peak value of a signal to the

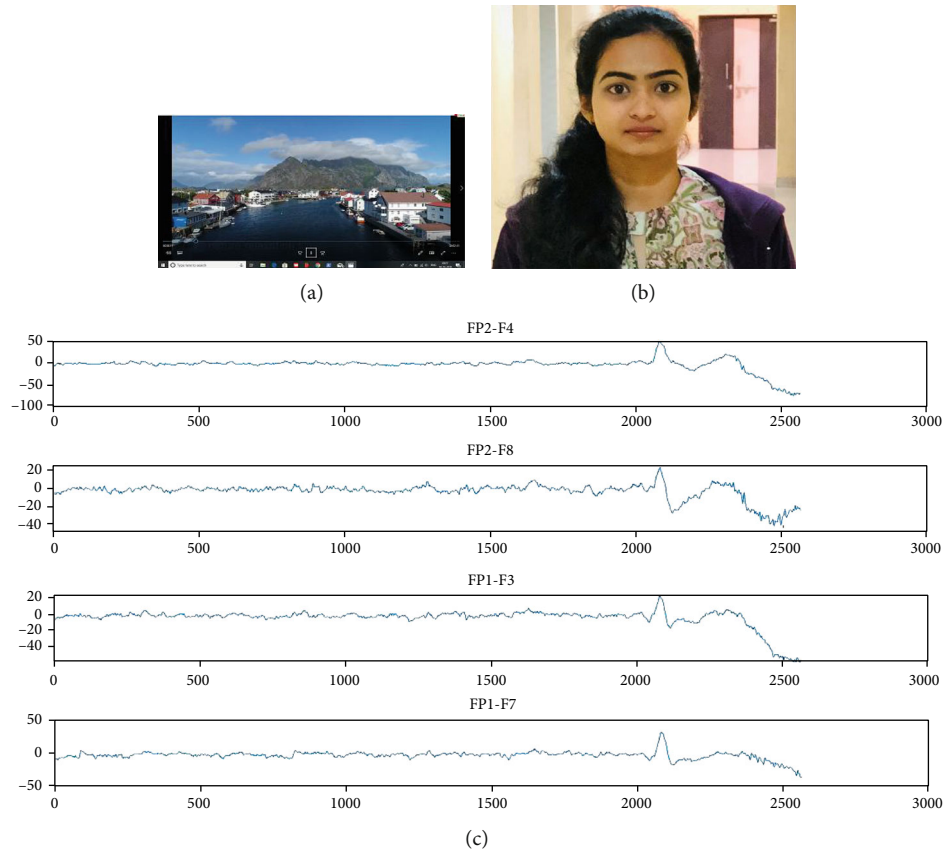


FIGURE 9: (a) The snapshot of neutral stimuli (b) reaction of a subject (c) captured EEG signals (x-axis and y-axis show time in second and amplitude in  $\mu V$ , respectively).

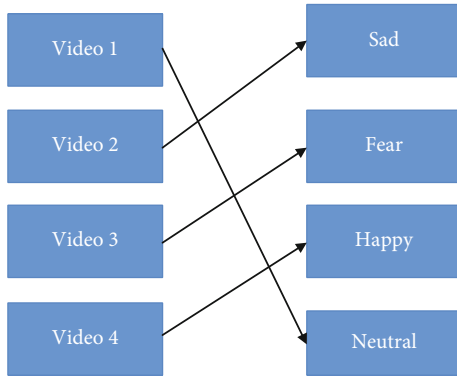


FIGURE 10: Pattern 1 for watching stimuli.

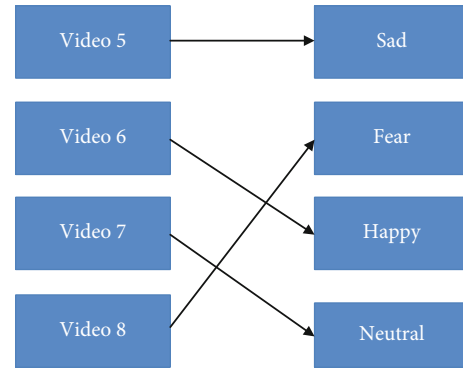


FIGURE 11: Pattern 2 for watching stimuli.

square of the average of the square root of the absolute value of the EEG signal.

$$f_4 = \left( \frac{\text{peak} - \text{value}}{\left( \frac{1}{C} \sum_{i=1}^C \sqrt{|s_i|} \right)^2} \right), \quad (4)$$

where peak value is obtained using  $(1/2)(\max(s_i) - \min(s_i))$ .

(v) Root mean square

Root mean square (RMS) is defined as the square root of the average of the square of each sample point of a signal which is estimated using Eq. (5).

$$f_5 = \sqrt{\frac{1}{C} \sum_{i=1}^C (s_i)^2}. \quad (5)$$

(vi) Crest factor

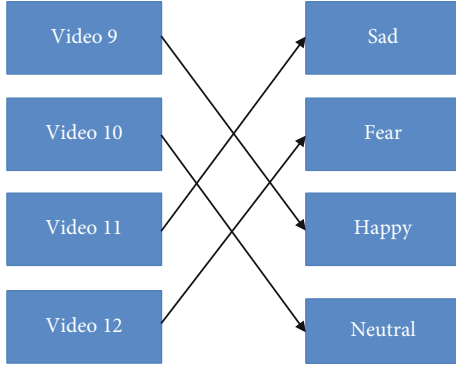


FIGURE 12: Pattern 3 for watching stimuli.

Crest factor is the ratio of the peak value of the signal to the RMS value, and it is found using Eq. (6).

$$f_6 = \left( \frac{\text{peak} - \text{value}}{\text{RMS}} \right). \quad (6)$$

(vii) Shape factor

The shape factor is the ratio of RMS value to the average of the absolute value of the signal which is computed by Eq. (7).

$$f_7 = \left( \frac{\text{RMS}}{1/C \sum_{i=1}^C |s_i|} \right). \quad (7)$$

(viii) Log detector

The nonlinear detector is defined as Eq. (8).

$$f_8 = e^{1/C \sum_{i=1}^C \log(s_i)}. \quad (8)$$

(ix) Mobility

Mobility is approximated from Eq. (9) which is defined as the square root of the ratio of the variance of the first derivative of signal to the variance of the signal.

$$f_9 = \frac{\sigma_d}{\sigma_a}. \quad (9)$$

(x) Absolute sum

Absolute sum is defined by Eq. (10).

$$\text{AS} = \sum_{i=1}^C |s_i|. \quad (10)$$

Every machine learning algorithm is associated with some objective functions. Sometimes, these objective functions may not work properly without

normalization, because the range of values of raw data varies greater than usual. For example, most of the classifiers estimate the distance between two points by the Euclidean norm. If one of the features or attributes has a wide range of values, the distance will be governed by this specific feature. Thus, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance [34]. So, normalization is done on extracted features also.

**4.3. Extreme Learning Machine Classifier.** The ELM is an efficient single-layer feedforward neural networks [35]. It is mainly considered as a classifier, regressor, and cluster analysis. However, it is adopted here for emotion classification because of the better generalization, controllability and robustness, and fast learning rate. It can solve the overfitting problem as compared to other conventional neural networks. It uses the empirical risk minimization theory. Moreover, it averts local minimization and multiple iterations. Furthermore, the learning process of the ELM takes a single iteration only. Finally, tune is not required for updating the parameters of hidden nodes. These nodes can be acquired from their forefathers without any modification or can be randomly assigned. Normally, the outcome weights of hidden nodes are gained in a single step. Let us consider an ELM with a single hidden layer and the output function of the  $j^{\text{th}}$  hidden node is  $m_j(x) = f(\alpha_j, \beta_j, x) = 1/1 + \exp(-(\alpha_j x + \beta_j))$ , where  $f(\cdot)$  is a sigmoid function and  $\alpha_j$  and  $\beta_j$  are the parameters of the  $j^{\text{th}}$  hidden node. The output function of the ELM for single-layer feedforward network with  $k$  hidden nodes is expressed by Eq. (11).

$$g_k(x) = \gamma_1 m_1(x) + \gamma_2 m_2(x) + \dots + \gamma_j m_j(x) + \dots + \gamma_k m_k(x), \quad (11)$$

where  $\gamma_j$  is the output weight of the  $j^{\text{th}}$  hidden node. The outcome of the hidden layer is estimated using Eq. (12).

$$m(x) = [f(m_j(x), \dots, m_k(x))]. \quad (12)$$

The hidden layer yield matrix  $M$  of ELM is expressed by Eq. (13), when the number of training samples is  $N$ .

$$M = \begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix} = \begin{bmatrix} f(\alpha_1, \beta_1, x_1) & \cdots & f(\alpha_l, \beta_l, x_1) \\ \vdots & \vdots & \vdots \\ f(\alpha_1, \beta_1, x_N) & \cdots & f(\alpha_l, \beta_l, x_N) \end{bmatrix}. \quad (13)$$

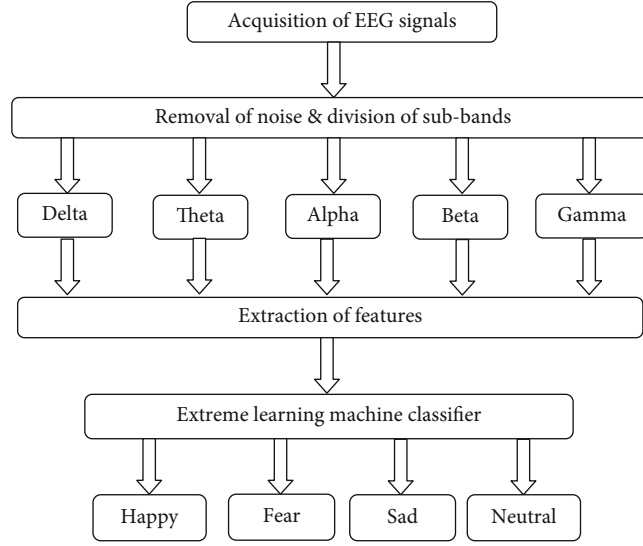
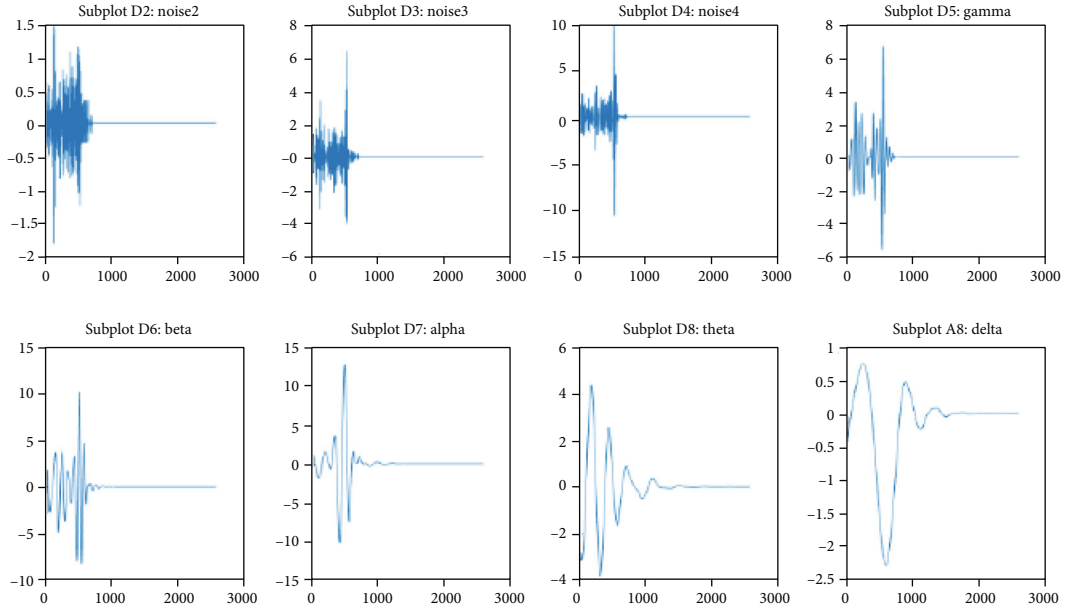


FIGURE 13: Block diagram of the proposed methodology.

FIGURE 14: The different subbands of a randomly selected channel from neutral emotion produced by "db8" DWT ( $x$ -axis and  $y$ -axis show time in second and amplitude in  $\mu V$ , respectively).

The training data target vector is represented by Eq. (13).

$$L = \begin{bmatrix} l_1 \\ \vdots \\ l_N \end{bmatrix}. \quad (14)$$

The objective function of ELM is shown in Eq. (14).

$$\|\gamma\|_p^{\sigma_1} + C\|M\gamma - L\|_q^{\sigma_2}, \quad (15)$$

where  $\sigma_1, \sigma_2 > 0$  and  $p, q = 0, 0.5, 1, 2, \dots$ . The basic form of ELM algorithm considered in this work is as follows:

$$\hat{Y} = W^2 \sigma(W^1 x), \quad (16)$$

where  $W^1$  and  $W^2$  are the two matrices used to denote weights of input to hidden layer and hidden to output layer, respectively, and  $\sigma$  is the sigmoid activation function. The steps of the ELM algorithm are as follows:

$$W^2 = \sigma(W^1 X)^+ Y, \quad (17)$$

where  $X$ ,  $Y$ , and  $+$  denote feature matrix, response variables, and pseudoinverse, respectively. Scholars use the ELM due to the property of low computational complexity. Most training can be accomplished in minutes, and even in some cases, the ELM takes seconds to complete the training process. The complexity of the ELM can be represented by  $O(nd)$ , where  $n$  and  $d$  denote input weights and biases, respectively [36].

*Step 1.*  $W^1$  is filled by some random values.

*Step 2.*  $W^2$  is estimated as per Eq. (17).

## 5. Experimental Results and Discussions

All the programs are implemented in Matlab 2017a, and these programs are executed on a laptop of the following specification: Intel(R) Core(TM) i5 5200U CPU @ 2.20 GHz, 4 GB RAM, 64-bit Windows Operating System.

At the very first step, DWT takes noisy EEG signals as inputs and produces noise-free EEG signals in the form of five subbands as outputs. Every subband is normalized to zero mean and unit variance. Ten features are extracted from these subbands in the second step. Then, all the features are normalized in the range of 0 and 1. In this study, 44 volunteers and their four emotions are considered for classification. Three videos are captured for each emotion. Ten pages of information for each video are stored while capturing videos. Four channels namely, FP2-F4, FP2-F8, FP1-F3, and FP1-F7 are selected out of 32 channels as these channels are placed in the frontal region of the head, and the frontal region is mainly responsible for identifying emotion correctly. Each channel has 5 subbands after noise removal. So, we have  $(44 \times 3 \times 10 \times 5 \times 4) = 6600$  samples altogether for each emotion. It means we get a feature matrix of size  $26400 \times 10$ , where 26400 is the sample size and 10 features are there for each sample. Three experiments are performed in this study. In the first experiment, channel-wise classification is carried out to know which channel is more responsible for emotion recognition. Several samples of each channel would be equal to  $(44 \times 3 \times 10 \times 5 \times 4) = 6600$ , and each channel will have 10 features. The feature matrix is divided into two sets namely, training and testing sets. The former one has 70% of total samples, and the rest of the samples are part of the testing set. Training samples and their class labels are fed into ELM for building a trained model. Finally, the test sample would be provided one after another as input and predicts its class label as output with the help of the trained model. Five measures namely, accuracy, specificity, sensitivity/recall, precision, and f1-score [37–39] are considered in this study for measuring the emotion recognition performance of each channel. The obtained value of each measure is reported in Table 1. It is clear from Table 1 that the performance of the first channel, i.e., FP1-F7 is better than the other three channels as the values of accuracy, sensitivity, specificity, precision, and f1-score are 94.72, 98.24, 94.72, 94.71, and 94.71, respectively, which are the highest among the four channels.

Subband wise emotion classification is performed in the second experiment. The number of samples of each subband

TABLE 1: Channel-wise performance comparison in percentage.

	FP2-F4	FP2-F8	FP1-F3	FP1-F7
Accuracy	90.85	91.23	94.01	94.72
Specificity	96.95	97.07	98.03	98.24
Sensitivity	90.85	91.23	94.01	94.72
Precision	90.82	91.02	94.09	94.71
F1-score	90.82	91.20	94.09	94.71

TABLE 2: Subband wise performance comparison for different channels.

Channels	Alpha	Beta	Gamma	Theta	Delta
FP2-F4	86.53	88.01	90.85	73.86	84.05
FP2-F8	85.79	88.21	91.23	80.67	85.95
FP1-F3	92.34	91.09	94.01	75.93	83.73
FP1-F7	94.70	94.23	94.72	76.44	86.16

TABLE 3: Emotions classification performance using gamma subband of FP2-F4 channel.

	Happy	Sad	Neutral	Fear
Accuracy	95.71	93.63	94.15	97.68
Specificity	96.83	96.01	96.29	98.13
Sensitivity	92.59	86.76	88.00	96.48
Precision	91.26	88.30	89.17	94.98
F1-score	91.92	87.52	88.58	95.72

would be equal to  $44 \times 3 \times 10 \times 4 = 5280$ . So, the size of feature matrix would be  $5280 \times 10$ . Then, 70% of the total samples are allocated for training. The remaining 30% samples are referred to as the testing set. ELM is trained by the training samples and their class labels. Test set is used then to predict the class labels of the testing samples. Accuracy is computed based on actual and predicted classes, and subband wise performance accuracy is reported in Table 2.

It is observed from Table 2 that the gamma subband provides the highest overall accuracy among all the subbands. Here, no emotion is considered separately. So, in the final experiment, only features are extracted from the gamma subband of the FP2-F4 channel. The number of samples would be equal to  $44 \times 3 \times 10 = 1320$  for each emotion. Again, the same procedure is adopted for classifying emotion by ELM. Emotion-wise classification performances are shown in Table 3. Table 4 reports the classification performance that is achieved with the help of features extracted from the gamma subband of the FP2-F8 channel. Similarly, the recognition performances are reported in Tables 5 and 6 only when features are obtained on gamma subband of the FP1-F7 channel.

## 6. Conclusion

Classification of human emotion using EEG signals is the best possible method, because it provides the actual neurophysiology of brain states. In this work, DWT decomposes



TABLE 4: Emotions classification performance using gamma subband of FP2-F8 channel.

	Happy	Sad	Neutral	Fear
Accuracy	96.07	94.85	93.92	96.95
Specificity	97.42	96.85	95.54	98.38
Sensitivity	92.31	90.02	89.08	93.14
Precision	92.75	90.18	87.01	95.57
F1-score	92.53	90.10	88.03	94.34

TABLE 5: Emotions classification performance using FP1-F3 channel and gamma subband.

	Happy	Sad	Neutral	Fear
Accuracy	98.33	95.25	94.72	99.74
Specificity	98.72	96.77	96.59	99.75
Sensitivity	97.22	90.74	89.17	99.72
Precision	96.42	90.50	89.78	99.30
F1-score	96.82	90.62	89.47	99.51

TABLE 6: Emotions classification performance using FP1-F7 channel and gamma subband.

	Happy	Sad	Neutral	Fear
Accuracy	97.95	95.86	95.75	99.69
Specificity	96.85	97.56	97.16	99.87
Sensitivity	97.42	90.85	91.59	99.18
Precision	94.95	92.66	91.59	99.61
F1-score	96.17	91.75	91.59	99.39

each EEG signal into subbands which are further used for extraction of features from that subbands. The channel-wise extracted features are fed as inputs to the ELM classifier for the classification of happy, fear, sad, and neutral emotion. This method provides better results using a single EEG channel. The single-channel-based emotion classification ability of the proposed method shows its adroitness for the practical applications of emotion classification in BCI systems. Recent state-of-the-art methods would also be considered to compare our proposed method. Deep learning convolution neural network may be employed to denoise the EEG signals followed by classification emotions. Here, we mainly concentrated to capture EEG signals database. We are also planning to consider more number of subjects to extend the database. We will focus to make a multimodality based BCI system for emotion recognition using face biometrics and physiological signals such as EEG signals, blood pressure, respiratory rate, body temperature, and galvanic skin response in near future.

## Data Availability

The “EEG signals related to emotions” data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data, 6 months after publication of this article, will be considered by the corresponding author under license agreement.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgments

This work is partially supported by the project “Prediction of diseases through computer assisted diagnosis system using images captured by minimally-invasive and non-invasive modalities”, Computer Science and Engineering, PDPM Indian Institute of Information Technology, Design and Manufacturing, Jabalpur India (under ID: SPARC-MHRD-231). This work is also partially supported by the SPEV project “Smart Solutions in Ubiquitous Computing Environments”, 2020, University of Hradec Kralove, Faculty of Informatics and Management, Czech Republic (under ID: UHK-FIM-SPEV-2020-2102) and by the Research University Grant Vot-20H04 at Universiti Teknologi Malaysia (UTM), Malaysia Research University Net-work (MRUN) Vot 4L876. We would also like to thank you for consulting to Sebastien Mam-bou and Ayca Kiritmat, Ph.D. students at FIM UHK.

## References

- [1] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press, 1998.
- [2] D. L. Schacter, D. T. Gilbert, and D. M. Wegner, *Introducing Psychology*, Macmillan, 2009.
- [3] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, “A review of affective computing: from unimodal analysis to multimodal fusion,” *Information Fusion*, vol. 37, pp. 98–125, 2017.
- [4] G. Caridakis, G. Castellano, L. Kessous et al., “Multimodal emotion recognition from expressive faces, body gestures and speech,” in *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 375–388, Springer, 2007.
- [5] A. Kolakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wrobel, “Emotion recognition and its applications,” in *Human-Computer Systems Interaction: Backgrounds and Applications 3*, pp. 51–62, Springer, 2014.
- [6] G. Chen, X. Zhang, Y. Sun, and J. Zhang, “Emotion feature analysis and recognition based on reconstructed eeg sources,” *IEEE Access*, vol. 8, pp. 11907–11916, 2020.
- [7] Y. Luo, Q. Fu, J. Xie et al., “Eeg-based emotion classification using spiking neural networks,” *IEEE Access*, vol. 8, pp. 46007–46016, 2020.
- [8] E. Cambria, “Affective computing and sentiment analysis,” *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.
- [9] M. S. Rani and S. Sumathy, “Perspectives of the performance metrics in lexicon and hybrid based approaches: a review,” *International Journal of Engineering and Technology*, vol. 6, no. 4, pp. 108–115, 2017.
- [10] E. Cambria, S. Poria, R. Bajpai, and B. Schuller, “Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives,” in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, pp. 2666–2677, Osaka, Japan, 2016.
- [11] A. Balahur, J. M. Hermida, and A. Montoyo, “Detecting implicit expressions of emotion in text: a comparative analysis,” *Decision Support Systems*, vol. 53, no. 4, pp. 742–753, 2012.

- [12] J. A. Dominguez-Jimenez, K. C. Campo-Landines, J. Martinez-Santos, E. J. Delahoz, and S. Contreras-Ortiz, "A machine learning model for emotion recognition from physiological signals," *Biomedical Signal Processing and Control*, vol. 55, article 101646, 2020.
- [13] S. Sun, C. Luo, and J. Chen, "A review of natural language processing techniques for opinion mining systems," *Information Fusion*, vol. 36, pp. 10–25, 2017.
- [14] E. Cambria and A. Hussain, *Sentic computing: Sentic computing: a common sense-based framework for concept-level sentiment analysis*, Springer, 2015.
- [15] M. Araujo, P. Goncalves, M. Cha, and F. Benevenuto, "ifeel: a system that compares and combines sentiment analysis methods," in *Proceedings of the 23rd International Conference on World Wide Web, ACM*, pp. 75–78, Seoul Korea, 2014.
- [16] S. Koelstra, C. Muhl, M. Soleymani et al., "Deap: a database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.
- [17] S. Katsigiannis and N. Ramzan, "Dreamer: a database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 98–107, 2018.
- [18] K. Tripathi, *Important Physiological Signals in the body*, 2019, January 2019, <http://biomedikal.in/2011/05/important-physiological-signals-in-the-body/>.
- [19] T. Matlovic, P. Gaspar, R. Moro, J. Simko, and M. Bielikova, "Emotions detection using facial expressions recognition and eeg," in *2016 11th international workshop on semantic and social media adaptation and personalization (SMAP)*, pp. 18–23, Thessaloniki, Greece, 2016.
- [20] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach, "Emotional speech: towards a new generation of databases," *Speech Communication*, vol. 40, no. 1-2, pp. 33–60, 2003.
- [21] P. Petta, C. Pelachaud, and R. Cowie, *Emotion-Oriented Systems: The HUMAINE Handbook*, Springer, 2011.
- [22] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, "The semaine database: annotated multimodal records of emotionally colored conversations between a person and a limited agent," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 5–17, 2012.
- [23] C. Busso, M. Bulut, C.-C. Lee et al., "Iemocap: interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 2008.
- [24] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The enterface'05 audio-visual emotion database," in *22nd International Conference on Data Engineering Workshops (ICDEW'06)*, p. 8, Atlanta, GA, USA, 2006.
- [25] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on eeg signals and various computer games-GAMEEMO," *Biomedical Signal Processing and Control*, vol. 60, p. 101951, 2020.
- [26] D. H. Hockenbury and S. E. Hockenbury, *Discovering Psychology*, Macmillan, 2010.
- [27] I. B. Mauss and M. D. Robinson, "Measures of emotion: a review," *Cognition and Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [28] E. Fox, *Emotion Science Cognitive and Neuroscientific Approaches to Understanding Human Emotions*, Palgrave Macmillan, 2008.
- [29] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.
- [30] P. J. Lang, "The emotion probe: studies of motivation and attention," *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.
- [31] "Manuals," 2016, February 2014 <https://www.trans-cranial.com/manuals/>.
- [32] "Clinically Validated tDCS Technology," 2016, <https://www.trans-cranial.com/>.
- [33] J. A. M. Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: a dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, 2018.
- [34] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 563–582, 2001.
- [35] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2012.
- [36] Y. Wang, F. Cao, and Y. Yuan, "A study on effectiveness of extreme learning machine," *Neurocomputing*, vol. 74, no. 16, pp. 2483–2490, 2011.
- [37] A. Seal, D. Bhattacharjee, M. Nasipuri, C. Gonzalo-Martin, and E. Menasalvas, "À-trous wavelet transform-based hybrid image fusion for face recognition using region classifiers," *Expert Systems*, vol. 35, no. 6, article e12307, 2018.
- [38] K. K. Sharma and A. Seal, "Modeling uncertain data using Monte Carlo integration method for clustering," *Expert Systems with Applications*, vol. 137, pp. 100–116, 2019.
- [39] A. Karlekar and A. Seal, "Soynet: soybean leaf diseases classification," *Computers and Electronics in Agriculture*, vol. 172, article 105342, 2020.

## Research Article

# Research and Verification of Convolutional Neural Network Lightweight in BCI

Shipu Xu <sup>1,2</sup>, Runlong Li,<sup>3</sup> Yunsheng Wang,<sup>2</sup> Yong Liu,<sup>2</sup> Wenwen Hu,<sup>2</sup> Yingjing Wu,<sup>2</sup> Chenxi Zhang <sup>1</sup>, Chang Liu <sup>4</sup>, and Chao Ma <sup>2</sup>

<sup>1</sup>Department of Software Engineering, Tongji University, Shanghai 201804, China

<sup>2</sup>Agricultural Information Institutes of Science and Technology, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China

<sup>3</sup>Department of Railway Transportation, Shanghai Institute of Technology, Shanghai 201418, China

<sup>4</sup>School of Information Engineering, Nanchang Hangkong University, Nanchang Jiangxi 330038, China

Correspondence should be addressed to Chenxi Zhang; [xzhang2000@163.com](mailto:xzhang2000@163.com), Chang Liu; [lcsszz83@gmail.com](mailto:lcsszz83@gmail.com), and Chao Ma; [machao@saas.sh.cn](mailto:machao@saas.sh.cn)

Received 26 March 2020; Revised 3 June 2020; Accepted 30 June 2020; Published 1 August 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Shipu Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increasing of depth and complexity of the convolutional neural network, parameter dimensionality and volume of computing have greatly restricted its applications. Based on the SqueezeNet network structure, this study introduces a block convolution and uses channel shuffle between blocks to alleviate the information jam. The method is aimed at reducing the dimensionality of parameters of in an original network structure and improving the efficiency of network operation. The verification performance of the ORL dataset shows that the classification accuracy and convergence efficiency are not reduced or even slightly improved when the network parameters are reduced, which supports the validity of block convolution in structure lightweight. Moreover, using a classic CIFAR-10 dataset, this network decreases parameter dimensionality while accelerating computational processing, with excellent convergence stability and efficiency when the network accuracy is only reduced by 1.3%.

## 1. Introduction

In the 5G era, with the development of emerging technologies such as the Internet of Things and big data, related applications in smart terminals are becoming more and more widespread. As a support for these intelligent applications, brain-computer-interface (BCI) technology plays an essential role in intelligent identification, classification, and computing. Our work mainly focuses on the intelligent recognition of images and videos, which is an indispensable intelligent application in life.

Since the publication of the 2006 Hinton research [1], deep learning algorithms have evolved rapidly. Based on the traditional artificial neural network (ANN) and the processing power of modern computers, it has achieved remarkable results in image processing, speech recognition, and scene analysis. The processing power and algorithm performance of complex problems have been greatly improved,

which has attracted widespread attention from academia and industry. The idea of deep learning can be summarized as unsupervised learning from bottom to top and parameter adjustment from top to bottom. Its adjustment process is based on the traditional BP algorithm. Typical deep learning algorithm models mainly include encoder, deep belief networks (DBN), and convolutional neural networks.

The development of ANN can be tracked back to the 1940s, and its development process is roughly divided into three stages. The first stage was the submission of the neuron model and learning rules from 1947-1969, such as perceptron, HEBB learning rules, binary neuron model (MP model), etc. The second stage is the HNN neural network model introduced by Professor Hopfield in 1982 by introducing the concept of Energy Function. The third stage is the classic back-propagation algorithm proposed by Professor Rumelhart in 1986. This algorithm is now known as the BP algorithm [2]. A typical three-layer ANN model is shown in Figure 1.

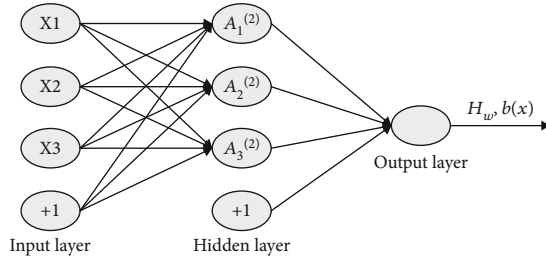


FIGURE 1: Traditional ANN model.

One of the benefits of deep learning frameworks in image recognition is that they do not need the traditional classification algorithms. It requires a lot of artificial processing of image features.

It is an adaptive algorithm. Through multilayer convolution and a nonlinear activation function, the algorithm classifies and regresses all image features through MLP [3]. It has the characteristics of migration invariance, scale invariance, and radiance image brightness. The overall CNN model starts with a convolutional layer. A convolution layer is a series of feature filter pairs (filter set), which contains multiple convolution kernels, and the output is feature maps. Then, convolution results are linearly modified; using RELU [4] function. After the convolution, a pooling layer is usually added. Generally, there is a mean pooling layer and a maximum pooling layer to compress the image.

Many classic deep learning network architectures have been developed based on the ILSVRC platform, such as AlexNet [5], ZFNet [6], VGGNet [7], GoogleNet [8], and ResNet. VGGNet is an improved framework based on the 1000-class image classification and localization model using the imageNet model. Due to the characteristics of neural networks, in order to obtain high accuracy, existing pipelines tend to increase the depth and complexity of their networks continuously. The number of internal parameters and the nonlinear mapping tends to be huge, which makes the deep network structure perform well in competition and data reflection. However, real-world applications are often constrained by storage space, computing power, and computing speed of the terminals. For example, in practical applications such as automatic driving, face recognition on mobile phones, video classification, etc., learning results are often demanded in milliseconds. Additionally, these devices often have limited processor performance with no prior trainings in the lab. Therefore, the practicality of CNN could be limited.

Two lines of work have been proposed to make deep learning networks applicable to daily lives. One is to improve hardware. The other is to improve the computing power of mobile terminals and to improve network structures, with a goal to minimize the training time and the amount of data required without affecting the accuracy. The development speed of the hardware is relatively slow, and its update iteration is far behind the speed required for the evolution of the network structure. Therefore, reducing the calculation parameters and calculation complexity of traditional network frameworks has gained the most research interests in deep learning.

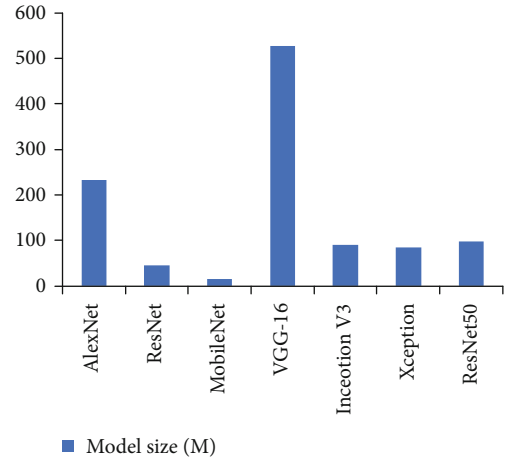


FIGURE 2: Comparison of famous model quality.

## 2. Related Works

Since the discovery of electrodes that can be used to collect EEG signals from the subcortex in the 1930s, research on EEG signals has provided experimental tools to decode neural substrates that are associated with thoughts and feelings of study subjects. With the rapid development of pattern recognition algorithms, ANNs, and deep learning frameworks, research on brain-computer interface (BCI) systems is in full swing.

BCI system-evoked potential collection methods include nonimplantable electroencephalogram (EEG) [7], implantable electroencephalogram (EcoG) [8], and functional magnetic resonance imaging (functional magnetic resonance imaging) [9]. The acquisition of nonimplantable EEG signals will not cause damage to the cerebral cortex. As a convenient, simple, and low-cost method, it has been widely used in the field of brain-computer interface system research.

The five-layer CNN has AlexNet and its optimized networks such as ZF, VGG, GoogleNet, ResNet, and DenseNet. Their performance is gradually improving, but the amount of parameters is also increasing. See Figure 2 for a comparison of the performance and quality of some of the more popular network models recently. It can be seen that the quality of these volume computer networks is mostly in the tens to hundreds of megabytes.

UC Berkeley proposed the SqueezeNet convolutional network model in 2016. This model can reduce these tens of megabytes and hundreds of megabytes of network structure to about 4.6 megabytes without affecting accuracy. This paper proposes three improvement strategies for SqueezeNet's core module, Fire Module. The first strategy is to improve on the dense  $1 \times 1$  convolution [9] kernel by using  $1 \times 1$  grouping convolution to reduce the number of calculations. This strategy can also solve the problem of noncirculation of channel information in grouped convolution. The second strategy is to add channel shuffle [10] (cross grouping) operation. This strategy reorganizes the different feature maps after grouping convolutions, so that the next grouping convolutions come from different groups, making



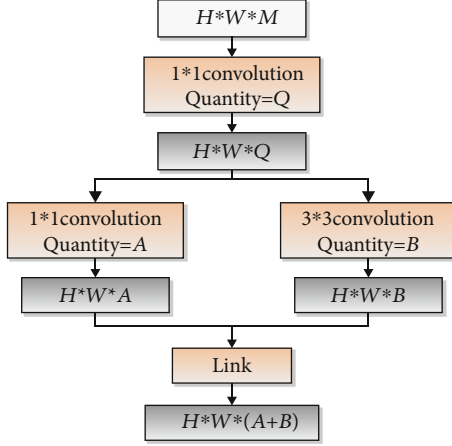


FIGURE 3: Composition of Fire Module

information flow in various groups. The third strategy is to adjust at the SoftMax layer and use SoftMax Loss [11] and Center Loss [12] to monitor training jointly. This training can be used to compensate for the high similarity image recognition effect.

In order to research the lightweight of CNNs in BCI, in 2015, Professor He Kaiming introduced a new structure of deep residual neural networks [13, 14]. The CNN trained on this structure has a depth from the AlexNet8 layer to the VGG19 layer to the ResNet152 layer and can converge and train regularly. ResNet won the championship with an accuracy of 16% over the second place in the ImageNet detection task and surpassed the second place by 27% in the ImageNet positioning task.

### 3. Improvements and Design of Fire Module

**3.1. Improvements of Fire Module.** As shown in Figure 3, the core module of the SqueezeNet network model is the Fire Module. It consists of the Squeeze layer and the Expand layer. The Squeeze layer is composed of a  $1 \times 1$  convolution kernel. The  $1 \times 1$  convolution kernel can change or reduce the number of channels when the model is input by changing its own convolution kernel number. Finally, the purpose of reducing the number of parameters and computational complexity is achieved. The Expand layer consists of a  $1 \times 1$  convolution kernel and a  $3 \times 3$  convolution kernel. Because the  $1 \times 1$  convolution calculation in the convolution operation accounts for most of the entire module, the calculation complexity is still high. This paper replaces the original  $1 \times 1$  conventional convolution kernel with a  $1 \times 1$  packet convolution on the basis of Fire-Module. In addition, batch normalization (BN) is used for the input of the model to speed up the training and convergence process and improve the classification accuracy. Finally, the strategy cascades improved Fire Modules.

**3.2. Grouping Convolutional Appointments.** The comparison between grouped convolution and regular convolution is shown in Figure 4. The traditional convolution is shown in Figure 1. The convolution kernel is completely converted

for training. In the grouping convolution, the convolution kernel is divided into  $N$  ( $N = 3$  in the figure) parts, and the input dimension is  $D_{in}/N$ . In the grouping convolution, the convolution kernel corresponds to the input  $[:, : 0 : D_{in}/N]$  dimension part for convolution operation. The second set of convolution kernels and the input  $[:, : D_{in}/N : 2D_{in}/N]$  dimensions are used for convolution. According to this, it is concluded that the output after the convolution operation of each group has become a convolution kernel in the  $D_0/N$  dimension. And each set of input and output operations is independent convolution operations. The input is convolved only with the current grouping convolution kernel and not with other grouping convolution kernels. After all the grouping convolutions are completed, the outputs of all  $D_0/N$  dimensions are superimposed to obtain the complete output of the final grouping convolution.

It is clear from Table 1 that the comparison of the conventional convolution parameters of the grouped convolution kernel is directly proportional to their ratio and number of groups. When the input size is  $W_1$  (width),  $H_1$  (height), and  $C_1$  (size), assuming a  $C_2$  convolution kernel and the size of the convolution kernel is  $h \times w$ , the calculations in the above table can be regarded as grouping convolution and regular convolution calculations and parameters. It can be concluded that the use of grouped convolution can significantly reduce the number of parameters and the calculation of the entire model.

**3.3. Channel Shuffle.** Because the input is a whole, the output after convolution is also mapped to the whole of the input. In the grouping convolution, the training of each grouping convolution is performed independently for each channel. This is equivalent to dividing the overall input into many independent parts for convolution. Therefore, the independent operation between each group will cause the information of each group channel to flow. To strengthen the information exchange between each packet, this article adds channel shuffle operation on this basis.

As shown in Figure 5, cross scramble the grouped features to form a new feature and input it into the next round of convolution operations. This allows the input of the grouping convolution to come from different groups and allows the information between different independent groups to circulate.

**3.4. Improvement of Loss Function SoftMax.** SqueezeNet uses a conventional SoftMax classifier, and the SoftMax function is a finite discrete probability distribution function. The SoftMax function, as shown in

$$L_S = - \sum_{i=1}^m \log \frac{e^{w_{ji}^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{w_{ji}^T x_i + b_{y_i}}} \quad (1)$$

For probabilistic multiclassification problems, it is simple and effective. However, the high similarity and features of human facial pictures are not apparent, and the class spacing of their features is often substantial. The intraclass distance is likely to be larger than the interclass distance. This will result



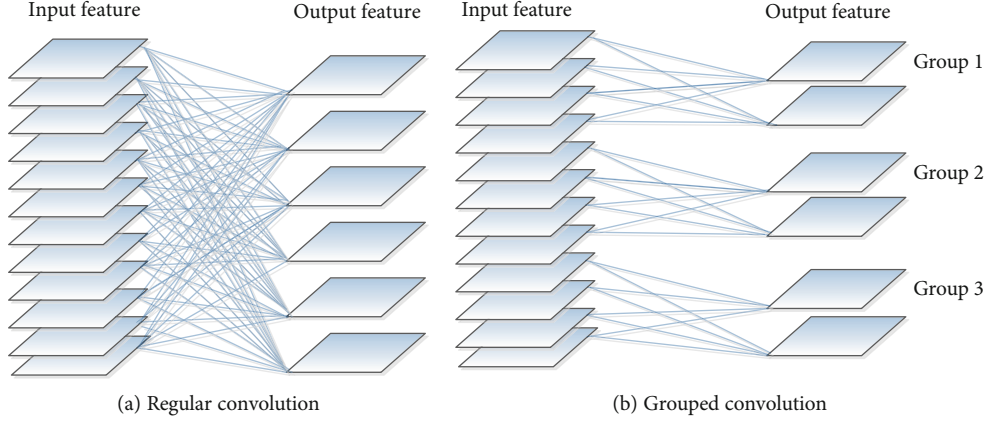


FIGURE 4: Regular convolution and grouped convolution.

TABLE 1: Comparison of parameters and calculations of traditional convolution and grouping convolution.

	Parameters	Calculations
Regular convolution	$C_1 \times C_2 \times 9$	$2 \times C_1 \times h \times w \times H_1 \times W_1$
Grouping convolution	$(C_1/g) \times (C_2/g) \times 9 \times g$	$2 \times (C_1/g) \times (C_2/g) \times h \times w \times H_1 \times W_1 \times g$
$R$	$1 : (1/g)$	$1 : (1/g)$

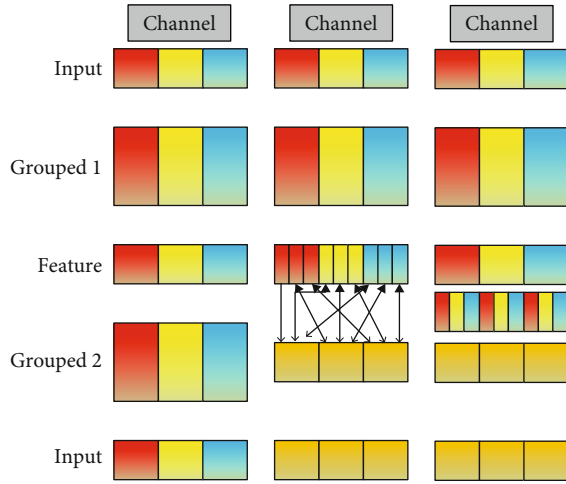


FIGURE 5: The principle of group convolution. (a) The channels are grouped into 3 groups, and there is no communication between different groups of feature maps. (b) Reconstructed feature maps. (c) Channel shuffle after convolution

in a lower recognition rate under complex face pictures. The Center loss function is shown in

$$L_C = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2. \quad (2)$$

The mixed loss function is shown in

$$L = L_s + \gamma L. \quad (3)$$

In Equation (3),  $x_i$  represents the features before the fully connected layer, and  $c_{y_i}$  represents the center of the  $y_i$

branch. The difference of Center Loss is that it adds a center to each class branch and, on this basis, increases the distance between other class branches and the center. So it makes the gap between classes smaller and the distance between classes larger. These features are more useful for classifying some complex images. Based on this, the SoftMax classification layer of SqueezeNet is improved to become a joint classification of SoftMax-Center Loss to reduce class spacing, which makes it more useful to recognize complex and similar face models.

**3.5. NVMNet.** The Fire Module improved based on the above method is shown in Figure 6. This article is named the NVM (New Visual Module). Based on this, the NVMNet (New Visual Module Net) structure is established. Compared with the previous Fire Module, the NVM is also composed of a compression layer and an expansion layer. This paper replaces the last  $1 \times 1$  conventional convolution with a grouped convolution to reach the model reduction. It reduced the number of input channels by  $1 \times 1$  grouping convolution in the compression layer before and added batch normalization after the  $1 \times 1$  convolution to speed up the training process. Then, channel shuffle allows the data to circulate the packet training information in different channels.

According to the hyperparameters of the previous module, the grouping number  $g$  of the grouping convolution of the Squeeze layer and the Expand layer, the number of convolution kernels  $h$  of the layer Squeeze, the number of convolution kernels  $w$  of the convolution layer, and the number of convolution kernels  $n$  of the expansion layer are set. Among them,  $h = m$ ,  $h < n$ , and  $w < n$ .

The dimensions and dimensions of the input and output of the entire model are the same. This article refers to the structure of the SqueezeNet model by linking the improved

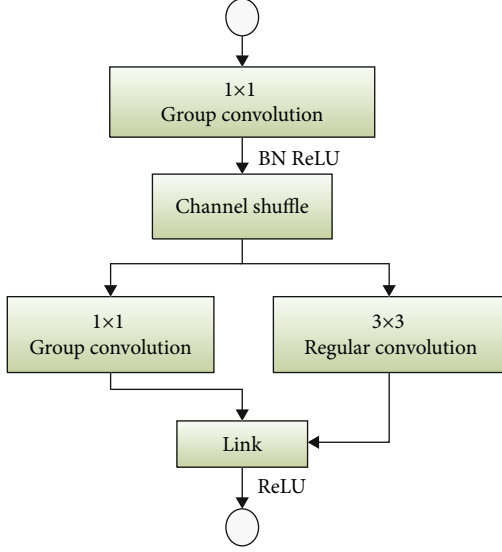


FIGURE 6: Improved NVM.

NVM. And add a pooling layer and two NVM to the fabric. Besides, add the SoftMax-Center Loss function at the end of the structure. The new SoftMax-Center Loss function maps the model's output value to the  $[0, 1]$  interval during training. But the overall output sum is still 1. It is the probability value of the classification result needed in this paper. Table 2 shows the whole structure of NVM.

#### 4. Improvement and Design Based on SE-ResNet Module

**4.1. Shortcut Connects and Bottleneck in ResNet.** The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

It is not difficult to know from Figure 7 that shortcut connect transfers the top-level information in the CNN to the bottom layer of the network similar to a fast link, if we consider a 50-layer CNN as a process in which 50 people send notifications in sequence. In the transmission process of the warning to the 50 people, inevitably, the expression of the notice and the initial notice issued at the time of the 50th person are inconsistent due to some reasons. Shortcut connect is similar to a pager, and when it is passed from the first person to the second person, the third person is notified through the pager at the same time. This can effectively avoid the loss or miscommunication of information. The notification process from the first person to the third person is a complete residual connection module.

It is not difficult to understand why ResNet is superior to other CNNs in terms of convergence speed and classification accuracy. In each remaining module, shortcut connect makes convolutional layer learning difficult. Secondly, it guarantees efficient transfer of gradients.

Figure 8 is the bottleneck structure [15, 16]. As shown in this figure, we can see that the entire structure is firstly transformed from 256-dimensional input features to 64-dimensional by  $1 \times 1$  convolution. Then, the feature extrac-

TABLE 2: NVMNet structure table.

Layer name	Output size (number of parameters)	$h$	$w$	$n$
Input	$224 \times 224 \times 3$			
Convolution layer	$111 \times 111 \times 64$			
Maximum pooling	$55 \times 55 \times 64$			
NVM2	$55 \times 55 \times 64$	16	16	64
NVM3	$55 \times 55 \times 64$	16	16	64
Maximum pooling 3	$27 \times 27 \times 128$			
NVM	$27 \times 27 \times 256$	32	32	128
NVM	$27 \times 27 \times 256$	32	32	128
Maximum pooling 5	$13 \times 13 \times 256$			
NVM6	$13 \times 13 \times 384$	48	48	192
NVM7	$13 \times 13 \times 384$	48	48	192
NVM8	$13 \times 13 \times 512$	64	64	256
Maximum pooling 9	$6 \times 6 \times 512$			
NVM11	$6 \times 6 \times 512$	64	64	256
NVM12	$6 \times 6 \times 512$	64	64	256
NVM13	$6 \times 6 \times 512$	64	64	256
Average pooling	$1 \times 1 \times 512$			
Full connection	Classification number			
SoftMax-Center Loss	Classification number			

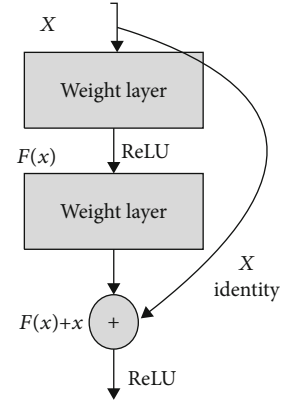


FIGURE 7: Structure diagram of shortcut connect.

tion of a  $3 \times 3$  convolution kernel is performed. Finally, the output dimension is restored to 256 dimensions through a  $1 \times 1$  convolution kernel. The calculation of the entire operation parameter amount is analogous to the grouped convolution and depth separable convolution mentioned in the previous research, and the parameter amount can be reduced by 17 times.

**4.2. Squeeze-and-Excitation Module.** In the ILSVRC2017 computer vision competition, SeNet won the classification championship [17]. Its core module is Squeeze-and-Excitation module [18]. As shown in Figure 9, it first uses global average pooling as the channel for the Squeeze operation for features. Then, a Bottleneck structure composed of

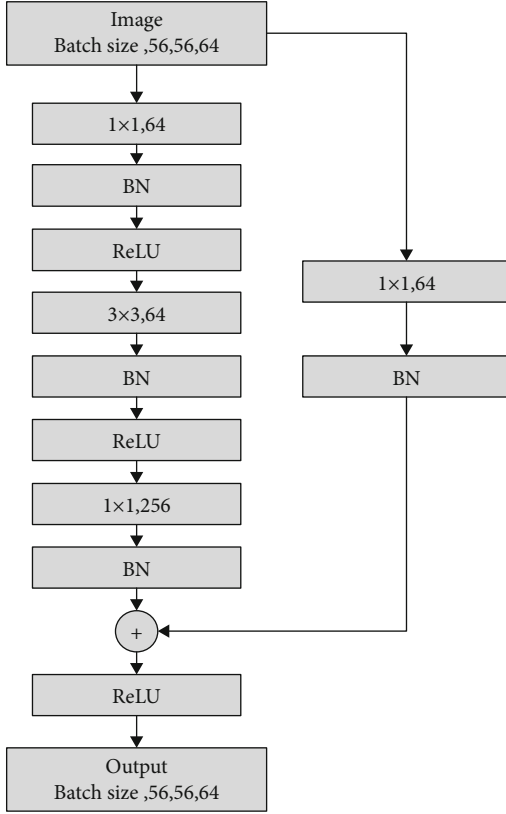


FIGURE 8: Schematic diagram of bottleneck structure.

two fully connected layers is used to remove the correlation between channels. The feature dimension is first reduced, and the dimension is increased to achieve the purpose of reducing the parameter amount and the calculation amount.

**4.3. Improvement of Squeeze-and-Excitation Module.** The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

According to the previous article, we improved the structure in the global pooling layer based on the Squeeze-and-Excitation module, grew the  $1 \times 1$  convolution operation, and substituted the new loss function.

According to the introduction of the CNN pooling layer, the sliding window size of the general pooling layer is fixed. Based on this, RMAC pooling introduces variable sliding windows to pool features. As shown in Figure 10, we use three sliding window sizes. Each sliding window is Max-pooled for the feature map, and 20 local features can be obtained. Max pooling the entire feature map will also get a local feature. In total, we received 21 local features. Then, normalize and add the 21 local features to capture our final features. The advantage of this is that it can better extract the information between each channel and extract more features.

In each dataset, the system is easy to recognize some categories, but difficult for others. Besides, the number of these samples is different, so we choose the CNN classifier suitable

for the moment. The number of samples in each category is prone to uneven proportions.

When a class has a large number of classifiers, the classifier can generally distinguish the class well. Conversely, when the number of classes is small, the performance of the classifier is not so good. In the conventional cross-entropy loss function, there is no distinction between the categories with a larger proportion of classifiers and the fewer categories. This will cause a waste of network resources. Because the system repeatedly learns those samples that have a good discrimination effect without focusing on training those samples with poor discrimination effects.

The focus loss function expression is shown in

$$FL(p_t) = -a_t(1 - p_t)^\gamma \log(p_t). \quad (4)$$

In Equation (4),  $(1 - p_t)$  is the modulation index. It is used for the contribution rate of different sample categories to the loss function. It is not difficult to see that when  $\gamma = 0$ , the focus loss function becomes a conventional cross-entropy loss function [14]. In Equation (4),  $p_t$  is the probability that the model predicts the sample category at the time of output. It divides the work according to the probability that our CNN predicts the sample category as the weight. Therefore, we use the focus loss function as a loss function to improve the network structure to reduce the training time of the improved model.

In the optimization of convolution, we use the idea of grouping convolution to replace all  $1 \times 1$  convolution operations in the structure with the grouping convolution. However, we have not added channel shuffle in this network structure module. This is because the number of channels in the first three convolutions of the SE module itself is different. In the case of an inconsistent number of channels, packet convolution comes with the function of channel shuffle. And blindly adding channel rearrangement itself will increase the memory space occupied by the network structure and significantly reduce the network's applicability.

The improved Se module is shown in Figure 11. It can be seen that when the input feature map of the previous layer is passed to the module, it is divided into two branches in the module. One branch borrowed the idea of shortcut to cascade the input feature map and the output part. The other branch borrows the idea of a bottleneck to compress and change the feature channel.

When the input of the upper layer enters the RSE module, it will be divided into two parts. One part uses the branch of shortcut connects, which takes the input of the previous layer directly as the output. Another branch of the bottleneck part is used to reduce the dimensionality of the previous layer input and then input convolution kernels of different sizes for feature extraction. The RSE module uses a  $1 \times 1$  convolution kernel and a  $3 \times 3$  convolution kernel. This can ensure the diversity of receptive fields between different channels. The features through different convolution kernels are linked. The feature fusion of different receptive field feature channels is performed through a  $1 \times 1$  convolution kernel. Based on

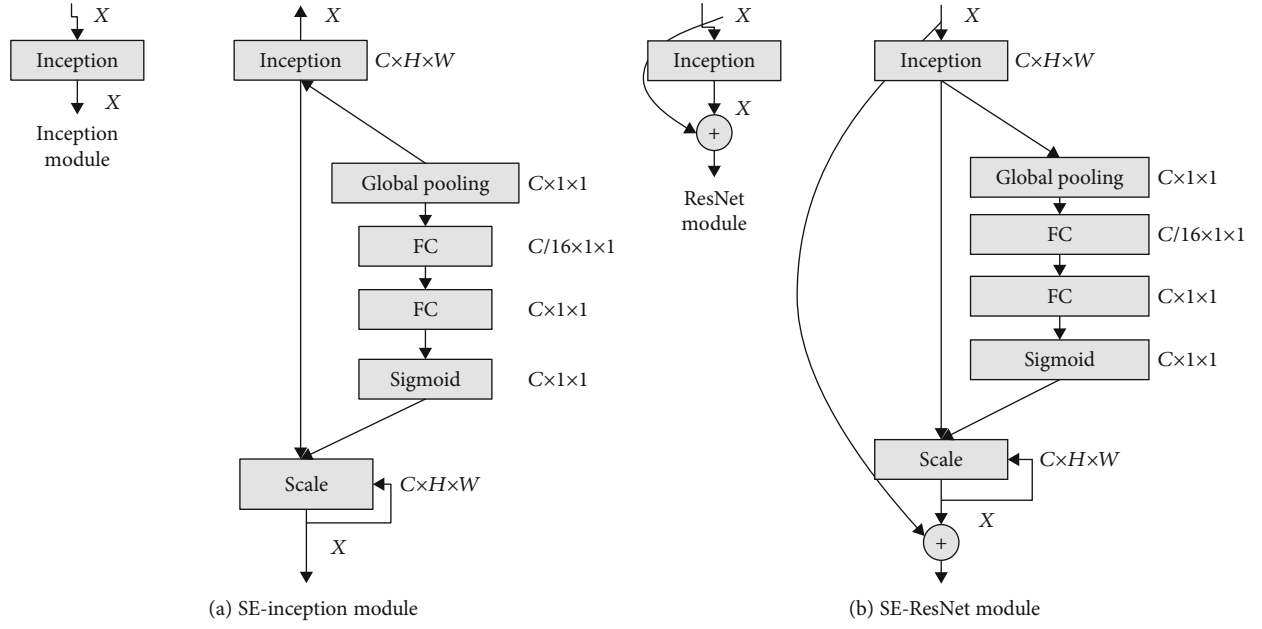


FIGURE 9: Schematic diagram of structure and composition of Squeeze-and-Exception module.

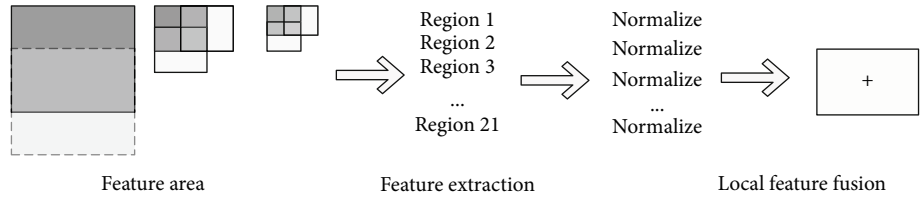


FIGURE 10: Schematic diagram of RMAC pooling process.

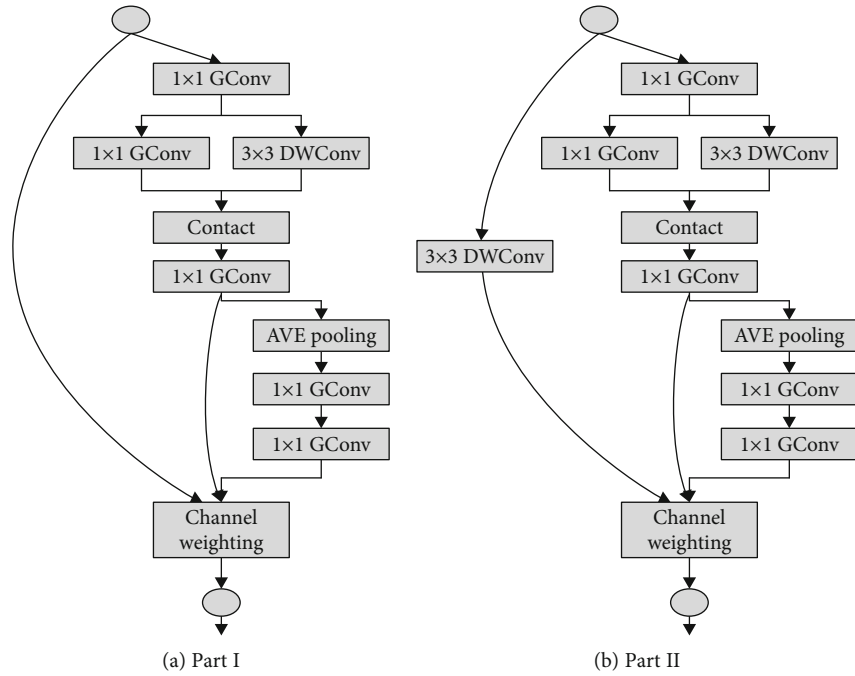


FIGURE 11: Schematic diagram of the improved SE module.

the RSE module, the modules are cascaded. The entire network model is shown in Table 3.

## 5. Results and Analysis

The ORL face dataset [19] was created in 1994 by the Olivetti Lab of the University of Cambridge, UK. There are 40 directories in the dataset. There are 10 facial expressions with different expressions stored in each directory. As shown in Figure 12, the pictures are saved in PGM format. Each image was acquired under different light and shadow, time, and facial features (open eyes, closed eyes, smile, and not smile). This experiment uses 300 of these pictures as the training set and 100 as the test set. The final output category is 40 (training on an Intel Xeon processor, Radeon Pro 580x graphics card, 32G memory MAC based on TensorFlow deep learning framework).

**5.1. Effect of Packet Convolution on Experimental Accuracy and Network Quality.** On the SqueezeNet, the SoftMax classifier at the end was also changed to a combination of SoftMax-Center Loss for monitoring. In the ORL dataset, this paper uses 4 groups on the number of NVM grouping convolutions.

As shown in Table 4, it can be seen that in the face recognition effect using the same classifier, the effect of grouping convolution on the classification accuracy has decreased, but it does not obviously mean that channel shuffle has completed the flow of information between channels. At the same time, after the introduction of grouping convolution, the parameter mount was effectively reduced by 33%. Reducing the model quality under the condition that the recognition effect has a limited impact indicates that the introduction of packet convolution in SqueezeNet is an effective model lightweighting strategy. However, because of the small number of training samples, the experimental accuracy is not very satisfactory.

**5.2. Influence of Different Grouping Numbers on Structural Quality and Accuracy.** Also on the ORL face dataset, the exponent of 2 can completely divide the dimension of the image input, so 2, 4, 8, and 16 grouping convolutions are used for training and testing, respectively. It can be seen from Table 5 that in the process of increasing the number of groups, the recognition accuracy is reduced by 1.2%, and the amount of parameters is correspondingly reduced by 13%. Under the comprehensive comparison, the effect is best when the number of current training sample groups is 4.

**5.3. Realization and Effect Comparison of Lightweight Model.** This experiment trained 40 different epochs on three different network structures on the ORL dataset. As shown in Figure 13, the ResNet convergence speed of NVMNet and residual networks using packet convolution and batch normalization is faster than that of AlexNet using conventional convolution. Convergence can be done in about 15 epochs. The best classification accuracy on the ORL dataset is ResNet, NVMNet, and AlexNet. At the same time, it can be seen in the comparison of the parameter amount and calculation amount in Table 6, because SqueezeNet refers to the deep

TABLE 3: Parameters of each level of the R-SeNet network structure.

Layer	Input size (number of parameters)	Kernel size (number of parameters)	Stride
Input size	$224 \times 224$	N/A	N/A
Convolution layer 1	$112 \times 112$	$7 \times 7$	2
Pooling layer 1	$57 \times 57$	$3 \times 3$	2
RSE.1	$57 \times 57$	$3 \times 3$	1
RSE.2	$57 \times 57$	$3 \times 3$	1
RSE.3	$29 \times 29$	$3 \times 3$	2
RSE.4	$29 \times 29$	$3 \times 3$	1
RSE.5	$15 \times 15$	$3 \times 3$	2
RSE.6	$15 \times 15$	$3 \times 3$	1
RSE.7	$8 \times 8$	$3 \times 3$	2
RSE.8	$8 \times 8$	$3 \times 3$	1
Pooling layer	$1 \times 1$	N/A	Global

compression technology on the basis of the structure, so that the parameter is reduced to 1.24 trillion, NVMNet continues to reduce the number of parameters based on its original 4.6 trillion. In the case of using the same classifier at the same time, the accuracy does not decrease. It shows that the channel shuffle after grouping convolution is more effective.

As shown in Table 6, through comparison with several popular network structures, it can be seen that the AlexNet and VGG models of the conventional convolution mode have larger parameters than other lightweight models. This is related to the many parameters of their fully connected layer. The amount of ResNet parameters using the global average pooling layer is moderate, but because of its deep network structure, the amount of calculation is very large. The advantage of NVMNet in terms of parameter comparison is obvious. This greatly increases the application scenarios of the network model. Save application storage memory and computing costs. Model mobile portability is more excellent.

Compared with that of SqueezeNet, on the ORL dataset, the classification accuracy of NVMNet is decreased by 0.7%. However, the parameter amount was reduced by 33%. Therefore, NVMNet has a certain value in lightweight models.

**5.4. Experimental Results and Data Analysis on the CIFAR-10 Dataset.** The CIFAR-10 [20] dataset contains 60,000 color images with a resolution of  $32 \times 32$ . It contains a total of 10 categories: airplane, car, bird, cat, deer, dog, frog, horse, boat, and truck. There are 6000 pictures in each category. There are 10,000 test set pictures and 50,000 training set pictures. As shown in Figure 14.

We use the CIFAR-10 classic dataset for efficiency comparison and parameter comparison of our CNN model (training on an Intel Xeon processor, Radeon Pro 580x graphics card, 32G memory MAC based on TensorFlow deep learning framework). According to our hardware conditions and the requirements of the network structure, the network



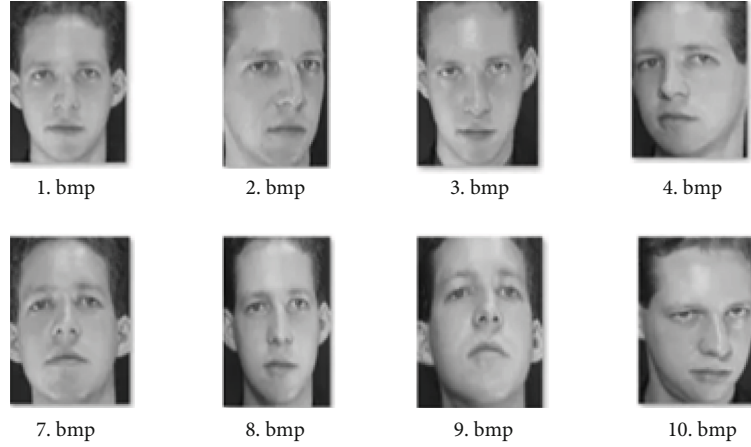


FIGURE 12: ORL dataset.

TABLE 4: Improved network quality nuclear accuracy comparison.

Structure	Parameter quantity (M)	Accuracy	
		Batch normalization	No-batch normalization
SqueezeNet	4.81	N/A	$0.7125 \pm 0.0004$
NVM	3.27	$0.7082 \pm 0.0002$	$0.7016 \pm 0.0013$

TABLE 5: Influence of packet convolution of different groups on quality kernel accuracy.

Number of packets	Accuracy	Parameter quantity (M)
2	$0.7142 \pm 0.0021$	3.31
4	$0.7091 \pm 0.0006$	3.12
8	$0.7069 \pm 0.0009$	3.01
6	$0.7064 \pm 0.0014$	2.88

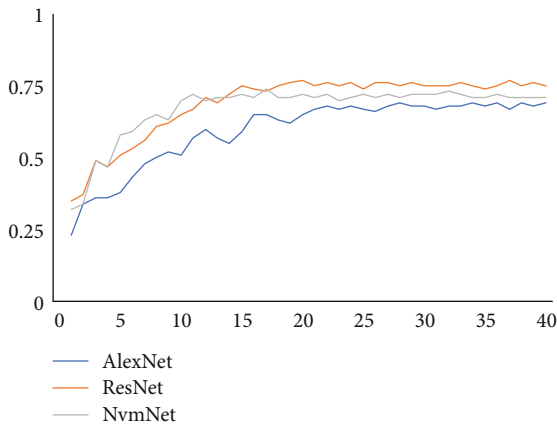


FIGURE 13: The training process of three different network structures on the ORL dataset.

parameter learning rate is 0.01, the optimization strategy of the CNN is stochastic gradient descent (SGD), the learning rate change rate is 0.1, and the maximum number of iterations is 400000.

TABLE 6: Comparison of different network structures.

Structure	Parameter quantity (M)	Amount of computation
SqueezeNet	1.24	0.70
AlexNet	61.18	0.73
ResNet	11.69	3.49
ShuffleNet	1.32	0.32
Vgg-16	138	72
NVMNet	3.21	0.3

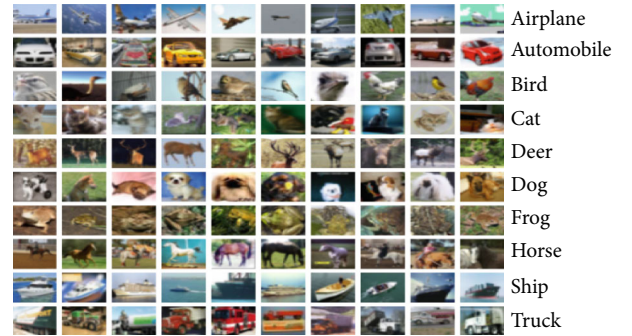


FIGURE 14: Overview of 10 categories and their respective pictures on the CIFAR-10 dataset.

As shown in Table 7, it is obvious that without any other model-specific compression method, the R-SeNet size is 9.6 MB. Compared with the ResNet model, the weight of the network model is nearly 90 MB smaller, but the classification accuracy is almost the same, only 1.8% lower. If the model is further pruned by parameters or channel depth, the size of the network model can be further reduced to about 3 MB while ensuring accuracy.

We compare the convergence of R-SeNet with other popular lightweight CNN networks, such as MobileNet, ShuffleNet, and SqueezeNet for network convergence. The comparative convergence curve is shown in Figure 15. From the convergence curve, ResNet has faster convergence speed and less fluctuation than other CNNs in the initial training. The converged waveform is relatively stable. From the

TABLE 7: Weight and resolution of CNN model on CIFAR-10 classic data.

CNN model	Network model weight (MB)	Network model compression ratio (%)	Network model accuracy (%)	Network model error distribution (%)
ResNet-50	98.1	N/A	95.1	N/A
ResNet-50@2.5	16.5	16.1%	93.1	-2.0
ResNet-50@.5	7.2	7.2%	93.1	-2.0
R-SeNet	9.6	10.1%	93.3	-1.8

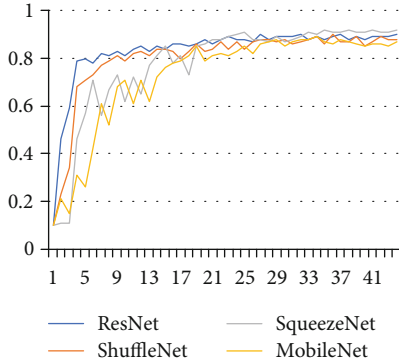


FIGURE 15: Comparison of training convergence of SqueezeNet, ResNet, ShuffleNet, and MobileNet.

perspective of accuracy, the accuracy of the four network models is almost equal: SqueezeNet, ResNet, ShuffleNet, and MobileNet. Among them, ShuffleNet is still unstable after convergence, and the network fluctuates greatly, which may be related to the parameter optimization during CNN group convolution. Through verification, it can be seen that the improved ResNet has a certain improvement in convergence speed and model weight with a precision of only 1.3%.

Based on the comparison of different lightweight CNN structures, we compare the improved parts of ResNet separately to observe the impact of each method on the performance of the network structure. As shown in Table 8, it can be seen that after the RMAC pooling layer is improved, the network extracts more feature information, so the performance of the network structure is slightly improved by 0.8%. The change of the loss function has the greatest impact on structural performance, which has increased by 2.5% before and after the improvement. This may be due to the redundancy of the network during the calculation. The system repeatedly calculates many classifications that are better distinguished without targeted training. Therefore, with the introduction of the focus loss function, the entire CNN can dynamically allocate training resources along with the output probability value, allowing the system to learn more about those indistinguishable classes. After the introduction of packet convolution, network performance decreased slightly by 1.1%. It may be related to the information circulation between the packet convolution channels.

## 6. Conclusions

Berkeley and Stanford proposed SqueezeNet to reduce parameter dimensionality with the AlexNet and VGGNet

TABLE 8: The influence of each compression thought segmentation on the accuracy of CNN.

R-SeNet				
Grouped convolution	✓	✓	✓	
RMAC pooling		✓	✓	
Focus loss function				✓
CNN model accuracy (%)	88.1 ± 0.3	87.0 ± 0.5	87.8 ± 0.1	90.3 ± 0.1

models. The 1 \* 1 convolution kernel has been used to reduce the number of input channels of the network, thereby reducing the number of network parameters. It compresses the CNN, from hundreds of megabytes in original size, to 4.5 megabytes without affecting the accuracy of image classification. When deep compression network compression technology mentioned above is used, the amount of network parameters can be further reduced to about 0.5 trillion. The core of the model is a module called Fire Module. This paper introduces packet convolution to optimize the Fire Module. In order to solve the problem that the channel information does not circulate after the group convolution, a channel shuffle operation is added between the channels, and the classifier is optimized and improved according to the complexity of the face features.

The BCI system based on the convolutional neural network includes functional modules including visual triggering device, EEG acquisition device, EEG preprocessing module, classifier based on convolutional neural network, and classification result display. There are successive dependencies among various modules. The system will first receive the EEG signal data from the EEG collector and then filter and normalize it through EEG signal preprocessing and then use the data as the input of the EEG signal classifier. After these EEG data are recognized and classified by the classifier, the recognition results are displayed in the result output module of this system. Based on the predecessors, this paper makes lightweight improvements based on the Fire Module of the SqueezeNet convolutional network structure. This paper introduces batch normalization and the SoftMax-Center Loss classifier to improve the recognition accuracy and efficiency of the network structure under the face. In the case of refining the overall structure of the network, the classification effect on the ORL dataset has also improved. However, because the ORL dataset has relatively few training samples, data samples can be added for further verification in future experiments. The lightweight model structure has

functional application scenarios. In the future, we plan to explore the feasibility of vision fields other than human faces, including applications in the BCI and BMI fields.

## Data Availability

If necessary, you can contact the author of this article for relevant experimental data.

## Conflicts of Interest

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

## Acknowledgments

This work is supported in part by the Shanghai Academy of Agricultural Sciences for the Program of Excellent Research Team (2017[B-09]). The authors thank the generous support from the Tongji University, particularly the Department of Software Engineering.

## References

- [1] S. Phadikar, J. Sil, and A. K. Das, "Rice diseases classification using feature selection and rule generation techniques," *Computers and Electronics in Agriculture*, vol. 90, pp. 76–85, 2013.
- [2] M. T. Marty, "Eliminating artifacts in electrospray deconvolution with a SoftMax function," *Journal of the American Society for Mass Spectrometry*, vol. 30, no. 10, pp. 2174–2177, 2019.
- [3] M. Khishe and A. Safari, "Classification of sonar targets using an MLP neural network trained by dragonfly algorithm," *Wireless Personal Communications*, vol. 108, no. 4, pp. 2241–2260, 2019.
- [4] D. Zheng, L. Xiangqun, and X. Xu, "Vehicle and pedestrian detection network based on lightweight SSD," *Journal of Nanjing Normal University (Natural Science Edition)*, vol. 42, no. 1, pp. 73–81, 2019.
- [5] J. Wu, Q. Yang, and L. Zhongliang, "Face recognition based on improved squeezeNet," *Science Technology and Engineering*, vol. 19, no. 11, pp. 218–223, 2019.
- [6] Q. Guo, L. Quanli, and W. Wei, "Fast SqueezeNet algorithm with application in metro crowd density estimation," *Control Theory & Applications*, vol. 36, no. 7, pp. 1036–1046, 2019.
- [7] B. Pengchen, J. Luo, and C. Weiwei, "Research on lightweight convolutional neural network technology," *Computer Engineering and Applications*, vol. 55, no. 16, pp. 25–35, 2019.
- [8] Y. Yinhui, Z. Zhengjin, and X. Tao, "Improved SSD model based on SqueezeNet and its application," *China Traffic Informationization*, vol. S1, pp. 189–191, 2018.
- [9] Y. Dong and J. Yu, "Light-weight convolutional neural network SlimNet based on SqueezeNet," *Computer Applications and Software*, vol. 35, no. 11, pp. 226–232, 2018.
- [10] B. Pengcheng, J. Luo, and C. Weiwei, "Lightweight convolutional neural network structure for mobile," *Information Technology and Network Security*, vol. 38, no. 9, pp. 24–29, 2019.
- [11] A. Camargo and J. S. Smith, "Image pattern classification for the identification of disease causing agents in plants," *Computers and Electronics in Agriculture*, vol. 2, no. 2, pp. 121–125, 2009.
- [12] L. Qin, Y. Gong, T. Tang, Y. Wang, and J. Jin, "Training deep nets with progressive batch normalization on multi-GPUs," *International Journal of Parallel Programming*, vol. 47, no. 3, pp. 373–387, 2019.
- [13] T. Kurita, N. Otsu, and T. Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," in *11th IEEE International Conference on Pattern Recognition*, pp. 213–216, The Hague, Netherlands, 1992.
- [14] R. Hong, J. Tang, H.-K. Tan, C.-W. Ngo, S. Yan, and T.-S. Chua, "Beyond search," *ACM transactions on multimedia computing communications and applications*, vol. 7, no. 4, pp. 1–35, 2011.
- [15] R. Hong, M. Wang, Y. Gao et al., "Image annotation by multiple-instance learning with discriminative feature mapping and selection," *IEEE Transactions on Cybernetics*, vol. 5, no. 5, pp. 669–680, 2014.
- [16] C. L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," *Pattern Recognition: The Journal of the Pattern Recognition Society*, vol. 36, no. 10, pp. 2271–2285, 2003.
- [17] C. Yu, T. Tong, and X. Digen, "Joint supervision of center loss and softmax loss for face recognition," *Journal of Chongqing University*, vol. 41, no. 5, pp. 92–100, 2018.
- [18] K. He, "Zhang, Xiangyu, Ren, Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 9, pp. 1904–1916, 2015.

## Research Article

# An Epilepsy Detection Method Using Multiview Clustering Algorithm and Deep Features

Qianyi Zhan<sup>1,2</sup> and Wei Hu<sup>3</sup>

<sup>1</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China

<sup>2</sup>Jiangsu Key Laboratory of Media Design Software Technology, Wuxi, China

<sup>3</sup>Department of Nuclear Medicine, Nanjing Medical University, Affiliated Wuxi People's Hospital, Wuxi, China

Correspondence should be addressed to Wei Hu; 6171610010@stu.jiangnan.edu.cn

Received 15 June 2020; Accepted 8 July 2020; Published 1 August 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Qianyi Zhan and Wei Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The automatic detection of epilepsy is essentially the classification of EEG signals of seizures and nonseizures, and its purpose is to distinguish the different characteristics of seizure brain electrical signals and normal brain electrical signals. In order to improve the effect of automatic detection, this study proposes a new classification method based on unsupervised multiview clustering results. In addition, considering the high-dimensional characteristics of the original data samples, a deep convolutional neural network (DCNN) is introduced to extract the sample features to obtain deep features. The deep feature reduces the sample dimension and increases the sample separability. The main steps of our proposed novel EEG detection method contain the following three steps: first, a multiview FCM clustering algorithm is introduced, and the training samples are used to train the center and weight of each view. Then, the class center and weight of each view obtained by training are used to calculate the view-weighted membership value of the new prediction sample. Finally, the classification label of the new prediction sample is obtained. Experimental results show that the proposed method can effectively detect seizures.

## 1. Introduction

Epilepsy is a common mental disease in neurology. It is a chronic neurological disease caused by sudden and temporary disturbance of brain function due to the paroxysmal abnormal discharge of brain neurons [1]. Worldwide, more than 1% of the population suffers from epilepsy [2]. Among neurological diseases in China, epilepsy has become the second most common disease, with fewer patients than headaches. There are many clinical manifestations of seizures. The more common symptoms are temporary sensory disturbances, behavioral disturbances, limb convulsions, loss of consciousness, or abnormal neurological function. Epilepsy has caused serious harm to patients' health and in some cases has been even life-threatening [3]. It can be said that epilepsy has brought a great burden to human health and social development. Therefore, the study of epilepsy disease has very important significance for medical research and the development of human society.

At present, the diagnosis of epilepsy is mainly through the patient and his relatives and friends, as much as possible to obtain a complete and detailed history of the patient's seizure and then be referred to the EEG for examination. The determination of seizures is mainly based on the patient's clinical manifestations and EEG records. EEG can record the abnormal electrical activity of neurons in patients' brains during epileptic seizures, so it has become an important tool for exploring the characteristics of epilepsy. It has been widely used in various epilepsy studies including epilepsy diagnosis, seizure location, qualitative research, seizure prediction, and seizure control. Until now, electroencephalogram examination is still mainly carried out by medical staff visual observation combined with their own experience. However, relying on manual recognition of epilepsy EEG has many drawbacks. First, with the advancement of medical conditions and the development of EEG acquisition equipment, the amount of EEG data is getting larger and larger, and the duration of EEG of a patient generally exceeds tens



or even hundreds of hours. In the face of a large amount of EEG data, manual detection is time-consuming and inefficient. Second, some features of epilepsy on EEG are very subtle and difficult to recognize by human eyes alone. Third, when medical staffs work for a long time, their judgment will be affected, which will easily lead to an increase in the false detection rate. With the in-depth research and development of computer technology, automatic detection and recognition of epilepsy EEG signals using computers as auxiliary tools have become an important auxiliary detection method. The study of automatic epilepsy detection technology becomes extremely important.

In view of the fact that EEG signals are a type of time series, the time-frequency analysis method is widely used when researching automatic epilepsy detection algorithms. Such methods analyze the time domain or frequency domain of epilepsy EEG signals or combine the two to capture the characteristic information of epilepsy EEG signals. Some time-frequency analysis methods such as wavelet analysis [4, 5], S transform [6], and empirical mode decomposition [7, 8] are used to analyze the characteristics of epilepsy EEG signals. Due to the nonlinear characteristics of EEG signals, nonlinear dynamic analysis has also become a method in the automatic detection of epilepsy. Entropy, as a parameter representing the complexity and randomness of the signal, is used as a feature of epilepsy EEG signals to distinguish it from EEG signals of different periods [9, 10]. With the development of pattern recognition and machine learning research, various learning algorithms [11–20] are also widely used in automatic detection of epilepsy. In recent years, with the development of deep learning, some deep neural networks have also been used in the classification of epilepsy EEG signals [21, 22].

At present, epilepsy detection technology has made great progress and many important research results have been achieved, but there are also some problems in the existing research. For example, (1) in the study of epilepsy detection based on EEG signals, different features such as the time domain and time-frequency domain of EEG are often directly combined with machine learning algorithms to identify different EEG patterns [23–26]. Different EEG feature extraction methods and classification methods are time-consuming and laborious and cannot meet real-time requirements in actual production. (2) Use one of EEG's features for decision-making, ignoring data from other perspectives, and making the decision result less than ideal. (3) The traditional feature extraction method applied to the features obtained by multiview data is easy to lose some important information and is not suitable for the feature extraction of multi-view data. (4) In terms of classifier design, the classifier based on a single perspective does not consider the correlation between data, and the decision result is not comprehensive enough. (5) When many existing epilepsy detection systems are applied to actual clinical data, it is difficult to achieve the performance results obtained on the experimental data set; that is, the robustness of the detection system is relatively low. Based on the above research status and analysis of existing problems, this paper proposes an epilepsy detection method using multiview clustering algorithm and deep

features. The specific contributions of this study are summarized as follows:

- (1) Construct the original multiview data set of epilepsy EEG. Different EEG analysis methods were used to obtain the data set under three perspectives of spectrum signal, time domain signal, and time domain signal of EEG
- (2) Extract deep features. Use deep DCNN to extract deep features of EEG signals from 3 perspectives, respectively, so as to obtain deep features from each perspective. This deep feature has a better ability to recognize epilepsy. And construct a multiview data set with deep features
- (3) Train a multiview classifier. First, a multiview FCM clustering algorithm is introduced, and the training samples are used to train the center and weight of each view. Then, the class center and weight of each view obtained by training are used to calculate the view weighted membership value of the new prediction sample. Finally, the classification label of the sample is obtained

## 2. Related Works

**2.1. Epilepsy Detection Process.** The EEG signals of patients with epilepsy are generally divided into four periods, namely, interseizure period, preseizure period, seizure period, and late seizure period. The interval between seizures refers to the EEG signals collected by epilepsy patients when the seizures are not occurring. The preseizure period refers to the EEG signals that may be collected when epilepsy is about to occur. There is no clear regulation on the length of this period and the starting position. The seizure period refers to the EEG signals recorded during the patient's seizure. Late seizures generally refer to the collection of EEG signals when the brain is in a recovered state after the seizure has ended. The process of predicting seizures can generally be divided into the following steps. The first step is to preprocess the EEG signals. In the second step, feature extraction is performed on the preprocessed samples to reduce the sample dimensions. The third step is to use the training samples to train the classifier. The fourth step is to use the trained classifier to classify the test samples to identify whether they are epilepsy patients. The detection flow chart is shown in Figure 1.

**2.2. EEG Signal and Its Characteristics.** As an important tool for researchers to study brain function, EEG reflects the electrical activity of brain nerve cell groups on the surface of the cerebral cortex or scalp. EEG is obtained by recording continuous, spontaneous, rhythmic potential changes in the brain nerve cell population through electrodes placed in the cortex or scalp. Certain characteristics of EEG signals are usually used as an important basis for the diagnosis of brain diseases. Because it contains a large amount of physiological and pathological information, it has played a huge role in clinical



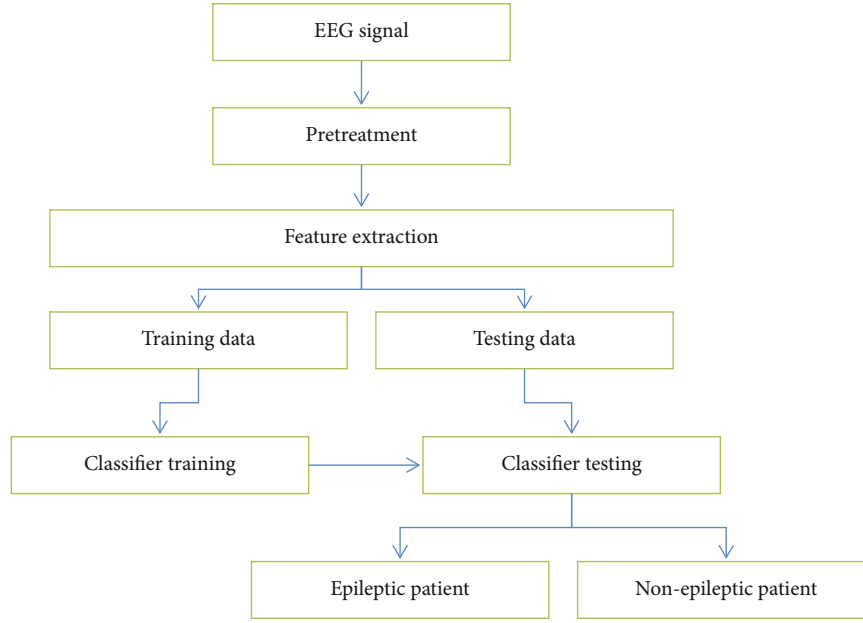


FIGURE 1: Flow chart of epilepsy detection.

medicine, especially for the location and characterization of epilepsy.

The current mainstream EEG analysis methods are the time domain analysis method [27–29], frequency domain analysis method [30–33], time frequency domain analysis method [34, 35], and nonlinear dynamics method [36–39]. The time-domain analysis method mainly analyzes the time-domain waveforms of EEG signals and extracts waveform features such as period and rhythm as the basis for detecting seizures. Its intuitiveness is strong, and its physical meaning is clear. Frequency domain analysis of EEG signals is mainly based on Fourier transform or power spectrum estimation. Analyze the power spectrum distribution of EEG signals and the components of EEG signals in each frequency band. Because EEG signals are random and nonstationary, simple time-domain analysis and frequency-domain analysis cannot provide the joint distribution information of EEG signals in the time domain and frequency domain and cannot describe the change of signal frequency with time. Therefore, a time-frequency analysis method suitable for nonstationary signal processing is used for epilepsy EEG analysis and feature extraction. The EEG data obtained by different analysis methods have different characteristics. Common EEG data features are shown in Table 1.

### 3. Deep Convolutional Neural Network

In the 1960s, when Hubel and Wiesel studied the cat's cerebral cortex, they discovered that the cat's neurons for direction selection and local sensitivity formed a unique network structure. This structure greatly reduces the complexity of the artificial neural network and the number of parameters [40]. DCNN borrows from this network structure to form a unique feedforward neural network. In 1998, Lecun et al.

proposed the LeNet-5 network model [41], but after a long period of time, the development of CNN fell into stagnation. In 2006, Hinton and Salakhutdinov pointed out that the feature learning ability of neural networks with multiple hidden layers is more powerful [42]. Since then, the structure of DCNN has been continuously improved by researchers [43], and models such as VGG [44], Google net [45], and Res net [46] have appeared successively. Compared with general neural networks, DCNN has four important characteristics: local connections, shared weights, pooling layers, and deeper network structures [47]. It is these characteristics that make the DCNN have more advantages than the general neural network when processing multiple types of signals.

In this paper, the DCNN model is used to model the EEG data of patients with epilepsy to achieve automatic learning of epilepsy features. The DCNN structure used in the experiment is shown in Figure 2. The DCNN model is mainly composed of convolutional layers and pooling layers, and ReLu and Dropout layers can be inserted between them to improve network performance.

**3.1. Multiview Clustering.** Current data sets often have high dimensionality and complexity. Complex data sets usually contain many different and reasonable clustering patterns. For example, in document data classification, there are both topic-based models and writing styles. In movie classification, it can be classified according to the type of movie or according to the genre of the movie. In gene classification, clustering can be based on the structure and function of genes.

When traditional clustering algorithms analyze data, they usually only focus on a single clustering pattern in the data. To understand data from a single perspective, we cannot have a comprehensive analysis and understanding of

TABLE 1: Common EEG data feature.

Name	Details
Time domain feature	The EEG signal is a time series signal that changes with time. The discrete points in the signal represent the energy intensity at a certain moment or the voltage value measured at that moment. In this study, the original EEG signal is directly used as the feature data from the time domain perspective.
Spectrum feature	The EEG signal is mainly divided into 6 frequency intervals, namely, $\delta_1$ (0-2 Hz), $\delta_2$ (2-4 Hz), $\theta$ (4-8 Hz), $\alpha$ (8-15 Hz), $\beta$ (15-30 Hz), and $\lambda$ (30-60 Hz). The spectral features of seizures are mainly distributed between 4 Hz and 30 Hz. For the initial EEG signal, the feature data under the spectral view were obtained by Fourier transform.
Time-frequency feature	Wavelet packet decomposition (WPD) is performed on the time-domain signal to obtain feature data from a time-frequency view. The sampling interval of the large frequency domain is set at 2 Hz, and the decomposition level of the wavelet transform is 6 layers. The time-frequency feature selection interval is 4 Hz-30 Hz.

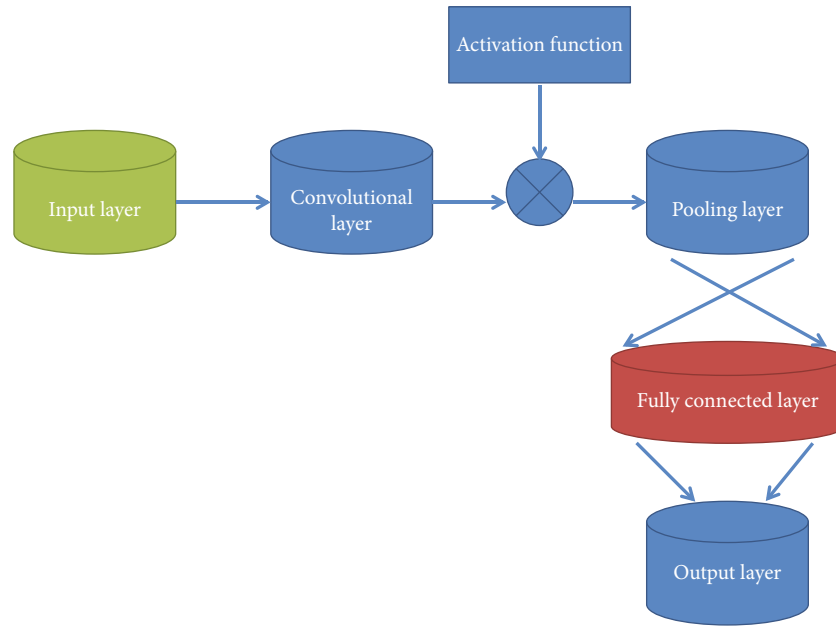


FIGURE 2: DCNN structure.

complex data. This problem has gradually attracted the attention of scholars and promoted the rapid development of multiview clustering. The multiview clustering algorithm hopes to mine multiple different and reasonable clustering structures from the data. There are two basic requirements. (1) The quality of the clustering result is high; that is, it can capture some reasonable internal structure contained in the data set itself. (2) There is no redundancy between different clustering results. Analyze the data from multiple perspectives, and discover multiple different clustering patterns in the data. The comparison between the traditional single-view clustering and multiview clustering schematics is shown in Figure 3 (using the classic FCM algorithm as an example).

#### 4. Introduction of the Proposed Algorithm

**4.1. Algorithm Framework.** The algorithm framework is shown in Figure 4. First, epilepsy EEG data from multiple views is obtained. Fourier transform and wavelet transform

were used to extract the spectrum signal, time domain signal, and time frequency signal of EEG data. Secondly, for each view signal, the classic DCNN is used for automatic feature extraction to reduce the sample dimension of each view. At this stage, the deep features of multiview epilepsy EEG data are obtained. Finally, a new classification method based on unsupervised multiview clustering results is used to obtain the final decision result.

**4.2. Deep Feature Extraction.** DCNN is used to extract the features of the data from various perspectives to obtain the deep spectrum features, deep time domain features, and deep time frequency features. Construct a multiview data set of epilepsy EEG based on deep features. The structure of the DCNN model used when extracting deep features from various perspectives is shown in Table 2.

**4.3. Multiview Fuzzy Clustering Method-Based EEG Classifier.**  $X = \{x_1, x_2, \dots, x_K\}$  is a multiview data set with  $N$

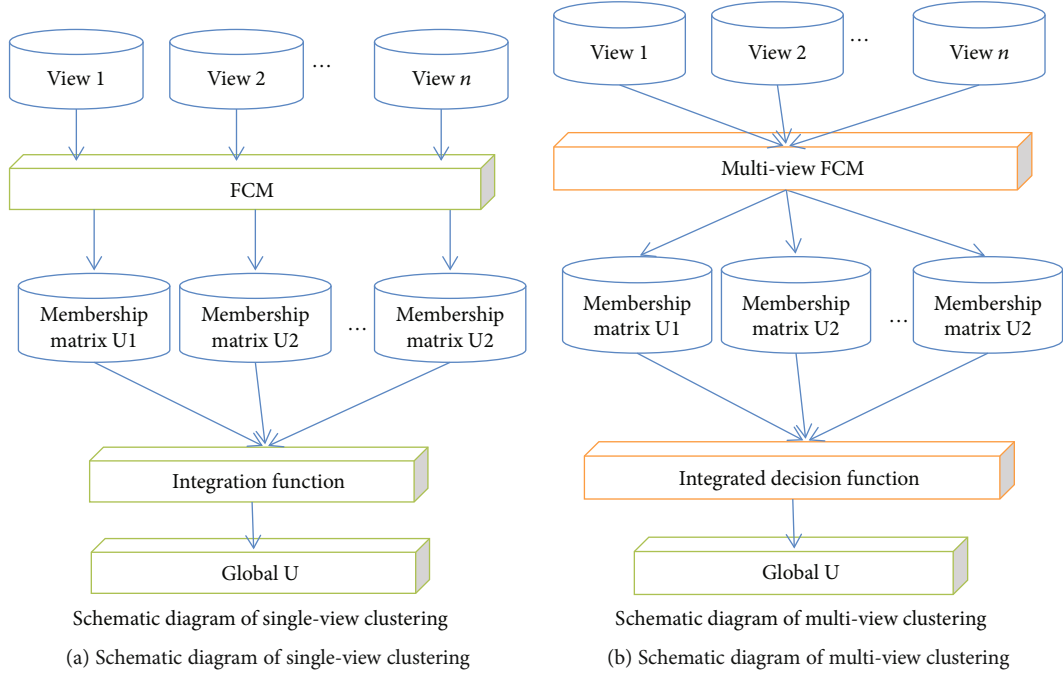


FIGURE 3: Comparison of single-view and multiview clustering algorithm schematics.

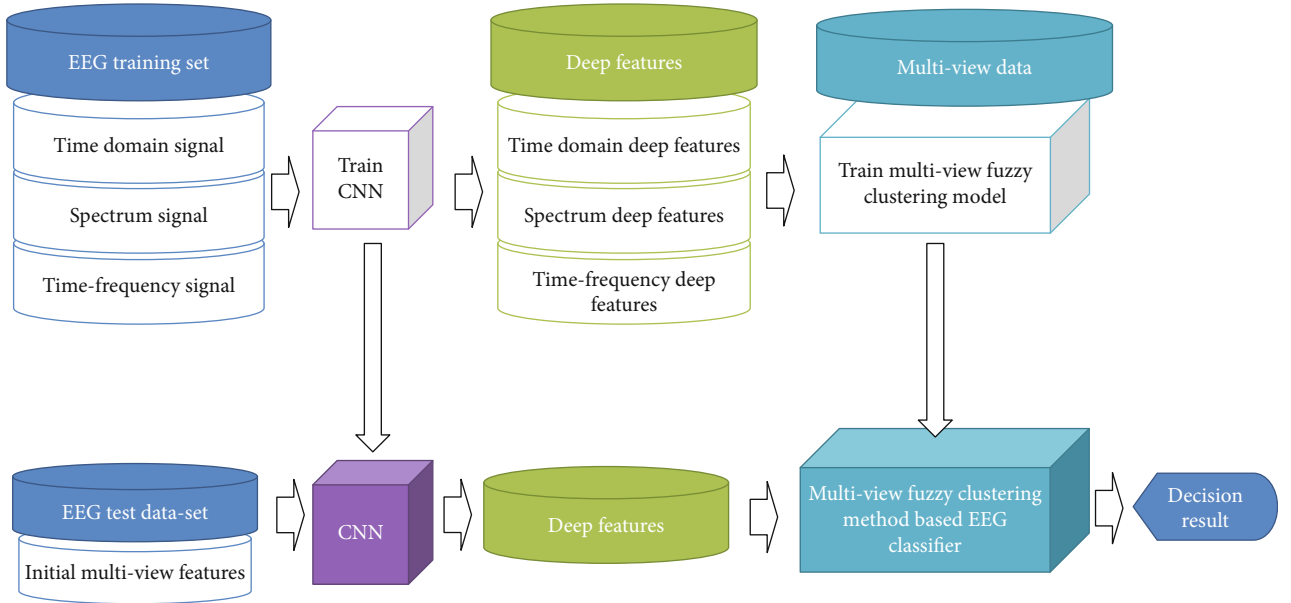


FIGURE 4: The proposed algorithm framework.

views.  $X^i$  represents the  $i$ -th view data,  $1 \leq i \leq N$ . The  $u(1 \leq u \leq K)$  row vector represents the feature vector of the data object  $x_u$  in the  $i$ -th view space, where  $n_i$  is the characteristic dimension of the  $i$ -th perspective. The objective function of multiview FCM is as follows:

$$P(U, O, W) = \sum_{i=1}^N w_i \sum_{c=1}^C \sum_{\mu=1}^K \sum_{p=1}^{n_i} (u_{c\mu})^m d(x_{\mu p}^{(i)}, o_{cp}^{(i)}) + T_w \sum_{i=1}^N w_i \ln(w_i), \quad (1)$$

$$\text{s.t.} \quad \begin{cases} \sum_{c=1}^C u_{c\mu} = 1, & 1 \leq \mu \leq K, \\ \sum_{i=1}^N w_i = 1. \end{cases} \quad (2)$$

Parameter  $m$  is the fuzzy index, which is used to adjust the fuzzy degree of membership.  $U = (u_{c\mu})_{C \times N}$  is the membership matrix, where  $u_{c\mu}$  represents the membership

TABLE 2: Network structure of deep feature extraction from various views.

Parameter/feature	Deep time domain feature	Deep spectrum feature	Deep time-frequency feature
Layer 1 (input layer)	Original matrix: 23*256 Convolution kernel: 1*128 Step size: 1	Original matrix: 23*27 Convolution kernel: 4*4 Step size: 1	Original matrix: 256*23*14 Convolution kernel: 129*1*1 Step size: 1
Layer 2 (convolutional layer)	Feature map: 1@23*129 Convolution kernel: 1*65 Step size: 30	Feature map: 20@20*24 Convolution kernel: 8*8 Step size: 1	Feature map: 1@128*23*24 Convolution kernel: 65*4*4 Step size: 1
Layer 3 (convolutional layer)	Feature map: 30@23*6 Convolution kernel: 4*33 Step size: 1	Feature map: 10@13*17 Convolution kernel: 8*8 Step size: 1	Feature map: 30@64*24*11 Convolution kernel: 30*4*4 Step size: 1
Layer 4 (convolutional layer)	Feature map: 20@20*33 Convolution kernel: 8*18 Step size: 1	—	Feature map: 20@32*17*8 Convolution kernel: 17*8*1 Step size: 1
Layer 5 (convolutional layer)	Feature map: 10@13*16	—	Feature map: 10@16*10*8
Layer 6 (fully connected layer)	The 10 feature maps of size 13*16 output from the convolutional layer of the fifth layer are converted into a vector of size 1*2080, which is used as the input of the fully connected layer. Feature map: 1@1*102	The 10 feature maps of size 13*17 output by the convolutional layer of the third layer are converted into a vector of size 1*2210, which is used as the input of the fully connected layer. Feature map: 1@1*512	The 10 feature maps of size 16*10*8 output by the convolutional layer of the fifth layer are converted into a vector of size 1*12800, which is used as the input of the fully connected layer. Feature map: 1@1*2048
Layer 7 (fully connected layer)	Feature map: 1@1*100	Feature map: 1@1*100	Feature map: 1@1*1024
Layer 8 (fully connected layer)	Output: 1*1024 vector	Output: 1*512 vector	Feature map: 1@1*100
Layer 9 (fully connected layer)	—	—	Output: 1*2048 vector

degree of the object  $x_\mu$  belonging to the  $c$ -th cluster.  $O^{(k)} = \{o_1^{(k)}, o_2^{(k)}, \dots, o_C^{(k)}\}$  is the vector of the  $k$ -th class center.  $W = (w_i)_{1 \times N}$ , where  $w_i$  represents the view  $i$  weights.  $d(x_{\mu p}^{(i)}, o_{cp}^{(i)})$  represents the Euclidean distance between  $x_\mu$  and  $o_c^{(i)}$  in the  $p$ -dimensional space;  $T_w$  stands for regularization parameter. Construct the Lagrange function of Equation (2) as follows:

$$\begin{aligned}
L(U, O, W) = & \sum_{i=1}^N w_i \sum_{c=1}^C \sum_{\mu=1}^K \sum_{p=1}^{\eta_i} \left( u_{c\mu}^{(i)} \right)^m d\left(x_{\mu p}^{(i)}, o_{cp}^{(i)}\right) \\
& + T_w \sum_{i=1}^N w_i \ln(w_i) + \sum_{\mu=1}^K \lambda_\mu \left( \sum_{c=1}^C u_{c\mu}^{(i)} - 1 \right) \\
& + \phi \left( \sum_{i=1}^N w_i - 1 \right).
\end{aligned} \tag{3}$$

Both  $\lambda_\mu$  and  $\phi$  are Lagrange factors of Equation (2). When the minimum value of Equation (3) is taken, the following 5 necessary conditions must be met

$$\frac{\partial L}{\partial \lambda_\mu} = \sum_{c=1}^C u_{c\mu}^{(i)} - 1 = 0, \tag{4}$$

$$\frac{\partial L}{\partial u_{c\mu}^{(i)}} = \sum_{i=1}^N w_i \sum_{p=1}^{\eta_i} m \left( u_{c\mu}^{(i)} \right)^{m-1} d\left(x_{\mu p}^{(i)}, o_{cp}^{(i)}\right) + \lambda_\mu = 0, \tag{5}$$

$$\frac{\partial L}{\partial o_{kp}^{(i)}} = 2 \sum_{\mu=1}^m \left( u_{c\mu}^{(i)} \right)^m \left( x_{\mu p}^{(i)} - o_{cp}^{(i)} \right) = 0, \tag{6}$$

$$\frac{\partial L}{\partial \phi} = \sum_{i=1}^N w_i - 1 = 0, \tag{7}$$

$$\frac{\partial L}{\partial w_i} = \sum_{c=1}^C \sum_{\mu=1}^K \sum_{p=1}^{\eta_i} \left( u_{c\mu}^{(i)} \right)^{m-1} d\left(x_{\mu p}^{(i)}, o_{cp}^{(i)}\right) + T_w (\ln(w_i) + 1) + \phi = 0. \tag{8}$$

By solving Equation (4)-(8), the necessary conditions that need to be met when Equation (1) obtains the minimum value under constraints are as follows:

Input Data set  $X$ , cluster number  $K$ , parameters  $m, T_w$ , initialization iteration parameters  $\tau = 0$ , maximum iteration parameters  $\tau_{\max}$ , termination condition threshold  $\varepsilon$   
Output The membership matrix  $U$  of the multiview data set that minimizes Eq. (1)  
Step1 Randomly generated and normalized membership matrix  $U$  of  $X$ .  
Step2 Repeat the following calculation process until the iteration termination condition is reached.  
For each  $i$   
1) Calculate the clustering center of the  $i$ -th view space according to Eq. (9);  
2) According to Eq. (3)-(4), calculate the distance  $d(x_{\mu p}^{(i)}, o_{cp}^{(i)})$  between the data and the cluster center in the  $p$ -dimensional space of the  $i$ -th view; .....  
3) Calculate the weight  $(w_i)^{\tau+1}$  of the  $i$ -th view space according to Eq. (11)-(12);  
End for  
4) Calculate the membership  $(u_{k\mu})^{\tau+1}$  of the data object according to Eq. (10)  
5)  $\tau = \tau + 1$   
Step3 Using Eqs. (10), (11) and (13) to generate a classifier and output the classification result.

ALGORITHM 1:

TABLE 3: Explanation of evaluation index.

Index	Index calculation formula	Remarks
Accuracy	$TN + TP / TP + TN + FP + FN$	TP is the number of samples that correctly predict the seizure fragments. FN is the number of samples that predict epileptic seizure fragments as nonepileptic seizures.
Sensitivity	$TP / TP + FN$	FP is the number of samples that predict nonseizure fragments as seizures. TN is the number of samples that correctly predict the non-seizure segment.

$$o_{cp}^{(i)} = \frac{\sum_{\mu=1}^K (u_{c\mu}^{(i)})^m x_{\mu p}^{(i)}}{\sum_{\mu=1}^K (u_{c\mu}^{(i)})^m}, \quad (9)$$

$$u_{c\mu}^{(i)} = \left( \frac{\sum_{e=1}^C \left( \frac{\sum_{i=1}^N w_i \sum_{p=1}^{n_i} d(x_{\mu p}^{(i)}, o_{cp}^{(i)})}{\sum_{i=1}^N w_i \sum_{p=1}^{n_i} d(x_{\mu p}^{(i)}, o_{ep}^{(i)})} \right)^{1/(m-1)}}{\sum_{e=1}^C \left( \frac{\sum_{i=1}^N w_i \sum_{p=1}^{n_i} d(x_{\mu p}^{(i)}, o_{cp}^{(i)})}{\sum_{i=1}^N w_i \sum_{p=1}^{n_i} d(x_{\mu p}^{(i)}, o_{ep}^{(i)})} \right)^{1/(m-1)}} \right)^{-1}, \quad (10)$$

$$w_i = \frac{\exp(-D^{(i)}/T_w)}{\sum_{l=1}^N \exp(-D^{(l)}/T_w)}, \quad (11)$$

$$D^{(l)} = \sum_{c=1}^C \sum_{\mu=1}^K \sum_{p=1}^{n_l} (u_{c\mu}^{(i)})^m d(x_{\mu p}^{(l)}, o_{cp}^{(l)}). \quad (12)$$

After the iterative optimization of the above parameters, we can use the following formula to complete the classification of EEG data. This is a new classification method based on the results of unsupervised multi-view clustering. The specific classification model is as follows:

$$f(x) = \max (W_i * U^{(i)}(x)), \quad (13)$$

where  $W_i$  is obtained by Equation (11) and  $U^{(i)}(x)$  is obtained by Equation (10).

The algorithm execution steps are as follows.

## 5. Experiment

**5.1. Experiment Related Settings.** In order to verify the effectiveness of the proposed algorithm, the experiment mainly conducts comparative analysis from two aspects. (1) When the proposed classification method based on unsupervised multiview clustering results is used to analyze whether deep features are used, the trend of classification performance changes. The general feature extraction methods used here are PCA [48] and LDA [49]. (2) In the case of deep features, the performance changes when using single-view classifiers and multiview classifiers. The comparison algorithms here mainly include FCM, SVM, RBF-NN, and multiview FCM. The indexes comparing the performance of epilepsy detection [50] are shown in Table 3. The larger the values of the two evaluation indicators shown in Table 3, the better the performance of the algorithm to be evaluated.

**5.2. Experimental Data Description.** The EEG database used in the experiment came from the Epilepsy Center of the University Hospital of Freiburg, Germany [51]. For 16 patients with refractory focal epilepsy, the epilepsy center recorded the patient's intracranial EEG during the preoperative epilepsy monitoring using invasive electrodes. The data acquisition process uses a Neuro NT digital video monitoring system that can record up to 128 channels of EEG signals. This system uses a 16-bit analog-to-digital converter with a sampling frequency of 256 Hz. Preprocess the collected raw data to obtain the EEG signal with 23 channels. The detailed statistics of the data used in the experiment are shown in Table 4.



TABLE 4: Experimental data.

Number	Signal length during seizures (seconds)	Signal length during the absence of epilepsy (seconds)
1	389	40287
2	202	69882
3	387	132382
4	549	129898
5	364	384984
6	186	220384
7	284	134443
8	583	153433
9	860	130239
10	798	71023
11	453	135668
12	237	174994
13	468	117894
14	346	256768
15	276	67974
16	189	86432

TABLE 5: Performance comparison of different feature extraction methods under multiview FCM classifier.

Feature extraction method\index	Accuracy	Sensitivity
PCA	83.22	45.68
LDA	79.43	33.18
Deep feature	89.75	85.52

**5.3. Analysis of Experimental Results.** In order to analyze the influence of depth features on classification performance, a classifier is fixed, and the multiview FCM is selected here. The classifier is used to process feature data obtained by different feature extraction methods. The experimental data in Table 5 is the clustering performance obtained after processing the feature data extracted by different feature extraction methods using the multiview FCM algorithm. Intuitive feedback from the data in the table shows the depth features obtained based on DCNN feature extraction. After being processed by the multiview FCM classifier, the classification performance is significantly better than the other two traditional feature extraction methods. Comparing the two traditional feature extraction methods of PCA and LDA, the effect of PCA is better than that of LDA, but the effect is much worse than that of deep feature extraction. Figure 5 intuitively shows the comparison of performance under different feature extraction methods.

In order to analyze the classification performance of different classifiers on deep feature data, this paper selected three single-view classifiers and one multiview classifier. The different classifiers are applied to the deep feature data from different perspectives and the deep fea-

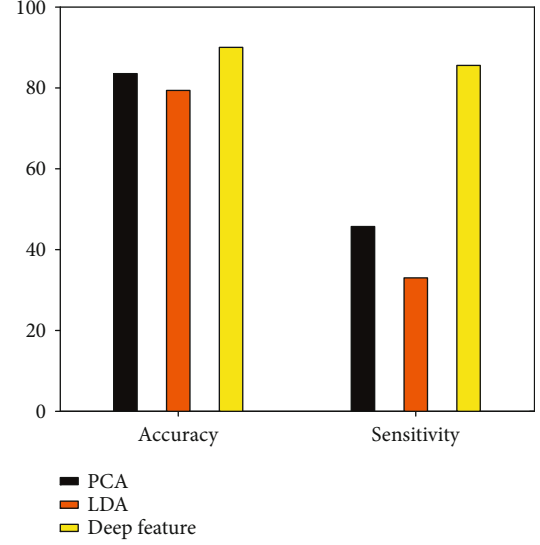


FIGURE 5: Schematic diagram of performance comparison under different feature extraction methods.

ture data from multiple perspectives for comparative analysis. Table 6 compares the classification performance of different algorithms on the deep features of different perspectives. From the experimental data in the table, it can be observed that among the deep features under three vsepate perspectives, the classification performance on the time domain deep feature is the worst. Among them, the classification performance of the deep features under the spectrum perspective is the best. Regardless of the deep feature under which angle of view or the deep feature under multiple angles of view, the multiview FCM algorithm has the best classification performance. The classification performance under a single perspective is worse than that of multiview data, which further shows that multiview data is richer in information and more beneficial to classification processing. Figure 6 gives a visual comparison of the performance of each algorithm under different perspectives.

## 6. Conclusion

This study proposes a new classification method based on unsupervised multiview clustering results. First, a multiview FCM clustering algorithm is introduced, and the training samples are used to train the center and weight of each view. Then the class center and weight of each view obtained by training are used to calculate the view weighted membership value of the new prediction sample. Finally, the classification label of the sample is obtained. The main contributions of the proposed method are as follows: (1) the deep feature can effectively reduce the data dimension while increasing the data separability; (2) the use of multiview data can improve the richness of data information, making decision results more stable and robust; and (3) the use of multiview classifiers improves

TABLE 6: Performance comparison of different algorithms on different depth features.

Feature	Classifier	Accuracy	Sensitivity
Deep features (time domain view)	FCM	73.02	53.57
	SVM	74.38	55.36
	RBF-NN	74.62	56.11
	Multiview FCM	76.98	56.87
Deep features (spectral view)	FCM	92.45	91.56
	SVM	95.03	94.48
	RBF-NN	94.74	94.96
	Multiview FCM	96.42	95.52
Deep features (time-frequency view)	FCM	89.89	85.30
	SVM	91.10	86.85
	RBF-NN	91.54	87.08
	Multiview FCM	92.12	89.83
Multiview deep features	FCM	92.98	92.02
	SVM	95.86	94.78
	RBF-NN	95.25	95.21
	Multiview FCM	<b>97.38</b>	<b>96.26</b>

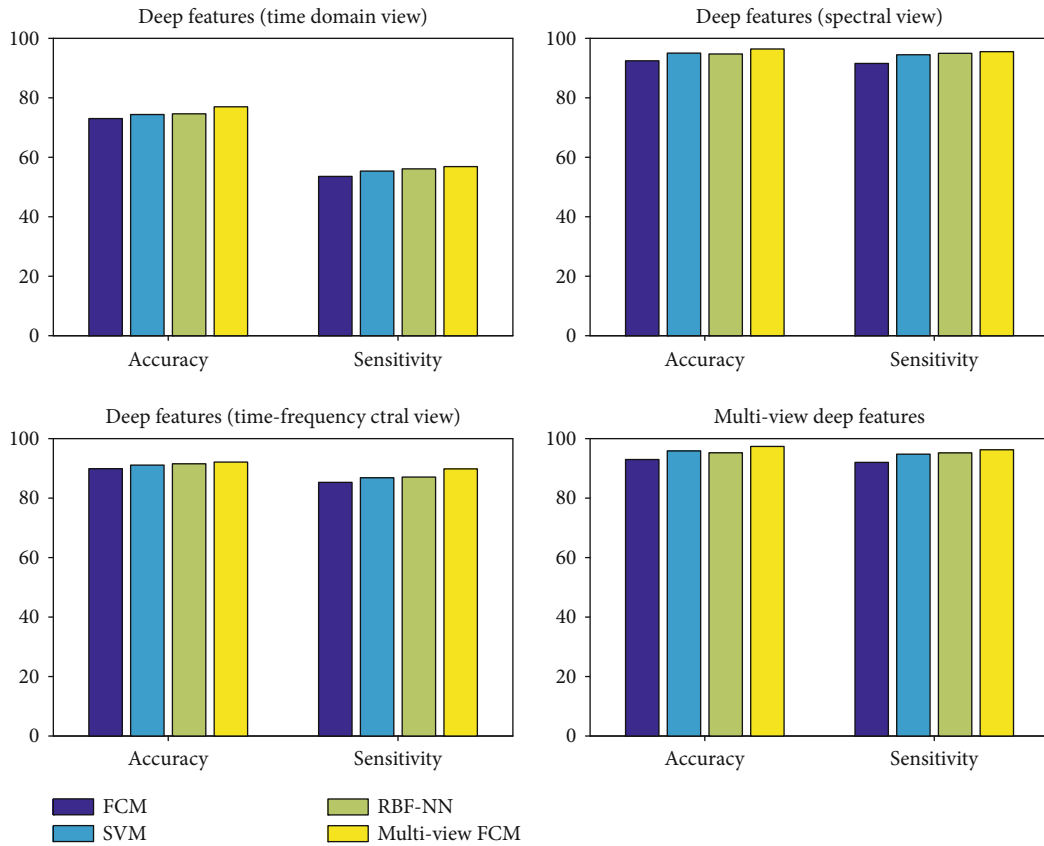


FIGURE 6: Performance comparison of various algorithms under various perspectives.

the classification performance of traditional single-view classifiers. Experiments show that the proposed method is more automated, intelligent, and efficient than traditional EEG detection based on machine learning.

### Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author or upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC, project 61902152) and the Natural Science Foundation of Jiangsu Province (BK20180600).

## References

- [1] W. A. Hauser, J. F. Annegers, and W. A. Rocca, "Descriptive epidemiology of epilepsy: contributions of population-based studies from Rochester, Minnesota," *Mayo Clinic Proceedings*, vol. 71, no. 6, pp. 576–586, 1996.
- [2] T. Gandhi, B. K. Panigrahi, M. Bhatia, and S. Anand, "Expert model for detection of epileptic activity in EEG signature," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3513–3520, 2010.
- [3] R. S. Fisher, W. E. Boas, W. Blume et al., "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, no. 4, pp. 470–472, 2005.
- [4] L. Ayoubian, H. Lacoma, and J. Gotman, "Automatic seizure detection in SEEG using high frequency activities in wavelet domain," *Medical Engineering & Physics*, vol. 35, no. 3, pp. 319–328, 2013.
- [5] S. Xie and S. Krishnan, "Wavelet—based sparse functional linear model with applications to EEGs seizure detection and epilepsy diagnosis," *Medical & Biological Engineering & Computing*, vol. 51, no. 1–2, pp. 49–60, 2013.
- [6] Y. Xia, W. Zhou, C. Li, Q. Yuan, and S. Geng, "Seizure detection approach using S-transform and singular value decomposition," *Epilepsy & Behavior*, vol. 52, Part A, pp. 187–193, 2015.
- [7] S. M. S. Alam and M. I. H. Bhuiyan, "Detection of seizure and epilepsy using higher order statistics in the EMD domain," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 2, pp. 312–318, 2013.
- [8] S. Li, W. Zhou, Q. Yuan, S. Geng, and D. Cai, "Feature extraction and recognition of ictal EEG using EMD and SVM," *Computers in Biology and Medicine*, vol. 43, no. 7, pp. 807–816, 2013.
- [9] N. Kannathal, M. L. Choo, U. R. Acharya, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 3, pp. 187–194, 2005.
- [10] Y. Kumar, M. L. Dewal, and R. S. Anand, "Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine," *Neurocomputing*, vol. 133, no. 8, pp. 271–279, 2014.
- [11] P. Qian, Y. Chen, J.-W. Kuo et al., "mDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 819–832, 2020.
- [12] P. Qian, K. Xu, T. Wang et al., "Estimating CT from MR abdominal images using novel generative adversarial networks," *Journal of Grid Computing*, vol. 18, no. 2, pp. 211–226, 2020.
- [13] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, p. 118, 2019.
- [14] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [15] K. Xia, H.-s. Yin, and Y.-d. Zhang, "Deep semantic segmentation of kidney and space-occupying lesion area based on SCNN and ResNet models combined with SIFT-flow algorithm," *Journal of Medical Systems*, vol. 43, no. 1, p. 2, 2019.
- [16] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [17] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.
- [18] P. Qian, C. Xi, M. Xu et al., "SSC-EKE: semi-supervised classification with extensive knowledge exploitation," *Information Sciences*, vol. 422, pp. 51–76, 2018.
- [19] P. Qian, Y. Jiang, S. Wang et al., "Affinity and penalty jointly constrained spectral clustering with all-compatibility, flexibility, and robustness," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 5, pp. 1123–1138, 2017.
- [20] P. Qian, K. Zhao, Y. Jiang et al., "Knowledge-leveraged transfer fuzzy C-Means for texture image segmentation with self-adaptive cluster prototype matching," *Knowledge-Based Systems*, vol. 130, pp. 33–50, 2017.
- [21] Q. Lin, S. Q. Ye, X. M. Huang et al., "Classification of epileptic EEG signals with stacked sparse autoencoder based on deep learning," in *Intelligent Computing Methodologies*, Springer, Cham, 2016.
- [22] A. Antoniadou, L. Spyrou, D. Martin-Lopez et al., "Detection of interictal discharges with convolutional neural networks using discrete ordered multichannel intracranial EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2285–2294, 2017.
- [23] A. S. M. Murugavel and S. Ramakrishnan, "An optimized extreme learning machine for epileptic seizure detection," *Iaeng International Journal of Computer Science*, vol. 41, no. 4, pp. 212–221, 2014.
- [24] A. Page, C. Sagedy, E. Smith, N. Attaran, T. Oates, and T. Mohsenin, "A flexible multichannel EEG feature extractor and classifier for seizure detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 2, pp. 109–113, 2015.
- [25] M. Z. Parvez and M. Paul, "Epileptic seizure detection by analyzing EEG signals using different transformation techniques," *Neurocomputing*, vol. 145, no. 18, pp. 190–200, 2014.
- [26] M. E. L. Menshaw, A. Benharref, and M. Serhani, "An automatic mobile-health based approach for EEG epileptic seizures detection," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7157–7174, 2015.
- [27] H. Qu and J. Gotman, "Improvement in seizure detection performance by automatic adaptation to the EEG of each patient," *Electroencephalography and Clinical Neurophysiology*, vol. 86, no. 2, pp. 79–87, 1993.

- [28] J. Gotman, "Automatic seizure detection: improvements and evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 76, no. 4, pp. 317–324, 1990.
- [29] R. Sankar and J. Natour, "Automatic computer analysis of transients in EEG," *Computers in Biology and Medicine*, vol. 22, no. 6, pp. 407–422, 1992.
- [30] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: methods and initial evaluation," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 356–362, 1997.
- [31] A. Alkan, E. Koklukaya, and A. Subasi, "Automatic seizure detection in EEG using logistic regression and artificial neural network," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 167–176, 2005.
- [32] H. Khamis, A. Mohamed, and S. Simpson, "Seizure state detection of temporal lobe seizures by autoregressive spectral analysis of scalp EEG," *Clinical Neurophysiology*, vol. 120, no. 8, pp. 1479–1488, 2009.
- [33] O. Faust, U. R. Acharya, L. C. Min, and B. H. C. Spath, "Automatic identification of epileptic and background EEG signals using frequency domain parameters," *International Journal of Neural Systems*, vol. 20, no. 2, pp. 159–176, 2012.
- [34] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.
- [35] R. J. Oweis and E. W. Abdulhay, "Seizure classification in EEG signals utilizing Hilbert–Huang transform," *Biomedical Engineering Online*, vol. 10, no. 1, p. 38, 2011.
- [36] N. Kannathal, U. R. Acharya, C. M. Lim, and P. K. Sadasivan, "Characterization of EEG—A comparative study," *Computer Methods and Programs in Biomedicine*, vol. 80, no. 1, pp. 17–23, 2005.
- [37] D. P. Subha, P. K. Joseph, R. Acharya U, and C. M. Lim, "EEG signal analysis: a survey," *Journal of Medical Systems*, vol. 34, no. 2, pp. 195–212, 2010.
- [38] K. Lehnertz, "Epilepsy and nonlinear dynamics," *Journal of Biological Physics*, vol. 34, no. 3–4, pp. 253–266, 2008.
- [39] U. R. Acharya, S. Vinitha Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: a review," *Knowledge-Based Systems*, vol. 45, no. 3, pp. 147–165, 2013.
- [40] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.
- [41] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient—based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [42] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, pp. 1097–1105, Curran Associates Inc., 2012.
- [44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.
- [45] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, Boston, MA, USA, June 2015.
- [46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, 2015.
- [47] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [48] L. Cheng, D. Li, G. Yu, Z. Zhang, X. Li, and S. Yu, "A motor imagery EEG feature extraction method based on energy principal component analysis and deep belief networks," *IEEE Access*, vol. 8, pp. 21453–21472, 2020.
- [49] X. Tan, L. Deng, Y. Yang, Q. Qu, and L. Wen, "Optimized regularized linear discriminant analysis for feature extraction in face recognition," *Evolutionary Intelligence*, vol. 12, no. 1, pp. 73–82, 2019.
- [50] W. Zhou, Y. Liu, Q. Yuan, and X. Li, "Epileptic seizure detection using lacunarity and bayesian linear discriminant analysis in intracranial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3375–3381, 2013.
- [51] "Freiburg Seizure Prediction Project, Freiburg, Germany," <http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database>.

## Research Article

# Research of Epidemic Big Data Based on Improved Deep Convolutional Neural Network

Wendong Wang 

Yan'an University, College of Mathematics and Computer Science, Yan'an Shaanxi 716000, China

Correspondence should be addressed to Wendong Wang; wangwendong2020@126.com

Received 23 May 2020; Accepted 23 June 2020; Published 22 July 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Wendong Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, with the acceleration of the aging process and the aggravation of life pressure, the proportion of chronic epidemics has gradually increased. A large amount of medical data will be generated during the hospitalization of diabetics. It will have important practical significance and social value to discover potential medical laws and valuable information among medical data. In view of this, an improved deep convolutional neural network (“CNN+” for short) algorithm was proposed to predict the changes of diabetes. Firstly, the bagging integrated classification algorithm was used instead of the output layer function of the deep CNN, which can help the improved deep CNN algorithm constructed for the data set of diabetic patients and improve the accuracy of classification. In this way, the “CNN+” algorithm can take the advantages of both the deep CNN and the bagging algorithm. On the one hand, it can extract the potential features of the data set by using the powerful feature extraction ability of deep CNN. On the other hand, the bagging integrated classification algorithm can be used for feature classification, so as to improve the classification accuracy and obtain better disease prediction effect to assist doctors in diagnosis and treatment. Experimental results show that compared with the traditional convolutional neural network and other classification algorithm, the “CNN+” model can get more reliable prediction results.

## 1. Introduction

With the arrival of the aging age and the acceleration of the pace of life, all kinds of life pressures come one by one. All these factors have caused the incidence rate of epidemic diseases, such as diabetes and cancer, which are increasing year by year. It shows that the prevention and treatment of epidemic diseases have become an urgent problem in the medical and health field. At the same time, with the continuous advancement of medical informatization, China's public health field has accumulated a wealth of data resources, which is in line with the typical characteristics of big data. The massive medical data resources usually contain a large number of valuable information, like patients' diagnosis and treatment laws. It plays an important role for us to better understand the causal relationship of epidemics and health risk factors by fully mining valuable treatment laws. As a typical chronic epidemic, the incidence rate of diabetes has remained high in recent years and shows a rising trend [1, 2]. Nowadays, the number of diabetic patients is large,

and the incidence rate of this disease increases with the incubation period. Therefore, it is an urgent task to establish a reliable prediction model based on the data of diabetic patients and judge the cause of disease as early as possible.

Currently, it is common to use machine learning-related models [3–11] for disease prediction. Researchers have made many explorations in the field of diabetes diagnosis and treatment and have achieved some results [12–21]. By monitoring 16 patients, Abraham of the University of California used statistical methods to analyze the risk factors of diabetes and found that if the glycosylated hemoglobin index of diabetic patients can be obtained from 9.6% down to 7.2%, this can improve the physiological state of patients [22]. At the same time, some researchers used decision tree algorithm and multilayer perceptron to carry out comparative experiments [23]. Sneha and Gangil used machine learning methods such as random forest, SVM,  $k$ -means, and naive Bayes to select early attributes that can be used to predict diabetes. The results show that the decision tree algorithm and random forest can achieve the best prediction effect for diabetes data [24].



Meanwhile, deep convolutional neural network is a hot research field in recent years. Because of its powerful feature extraction ability, it can mine deeper features from a large number of training data with the hierarchical network structure, so as to extract the feature information that cannot be obtained by traditional classifiers. Therefore, it has been widely used in speech recognition, image recognition, text detection, and so on [18, 25–33]. As we know, the medical data set has the characteristics of large amount of data and rich features, so it is helpful to discover potential medical laws and valuable information among medical data by applying deep convolutional neural network to medical data. In a word, it will have important practical significance and social value [12, 34–39]. For example, Swapna et al. applied CNN to realize automatic detection of diabetes mellitus [14]. They used heart rate variability data to obtain heart rate signals and used CNN-LSTM combined network to carry out automatic anomaly detection and fully connected structure, which can realize automatic detection and accurate diagnosis of diabetes [13].

In summary, although researchers have done a lot of research on diabetes, most of them pay attention to the diagnosis method, blood glucose detection method, and complications of diabetes. Few researches are aimed at the inpatients of diabetes. In view of this, a typical epidemic disease, like diabetes mellitus, is taken as an example. And an improved algorithm based on the deep convolutional neural network is proposed to predict the change of diabetes based on the data of inpatient medical records of diabetes patients. The improved algorithm takes advantages of both the bagging integrated classification algorithm and the deep convolutional neural network. It not only has the good data classification ability but also has the strong feature extraction ability, which can effectively improve the classification accuracy and obtain better disease prediction effect to assist doctors in diagnosis and treatment. The innovation of this paper is that the bagging integrated classification algorithm is applied instead of the output layer function of the deep convolutional neural network to build an improved deep CNN algorithm (“CNN+” for short) for the data set of diabetic patients. On the other hand, in order to ensure that the experimental results and analysis are based on good structured data, data preprocessing is also carried out in the process of the experiment, including data cleaning, data balancing, data feature processing, and abnormal data processing. Finally, the structure of the paper is as follows.

Part 1 is the introduction to the research background, status, and significance of the thesis.

Part 2 explains the relevant work of the thesis, so as to provide a theoretical basis for the follow-up research.

Part 3 introduces the principle of the improved algorithm proposed in this paper in detail.

Part 4 verifies the effectiveness of the method proposed in this paper in the detection of diabetes data based on a series of experiments.

## 2. Related Work

Deep convolutional neural network is a special type of neural network. Its super learning ability is mainly achieved by

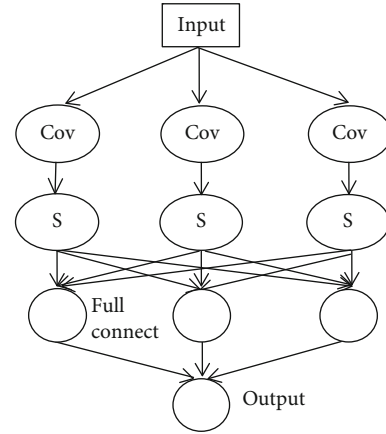


FIGURE 1: The structure of the convolutional neural network.

using multiple nonlinear feature extraction, which can automatically learn hierarchical representation from data. CNN is a deep neural network including input layer, convolution layer, pooling layer, full connection layer, and output layer. Firstly, CNN extracts the features of input data by convolution. Then, by using weight sharing and pooling, the difficulty of training network and the redundant data are greatly reduced, and the features are retained to the maximum extent. Finally, features are transferred from the full connection layer to the output layer for classification [29]. Because of convolution and pooling operation, the deep CNN has the characteristics of sparse connection, parameter sharing, translation, and local translation invariance, which can effectively extract image features while reducing learning parameters and training difficulty, making CNN widely used.

**2.1. The Main Components of Deep CNN.** A typical deep convolutional neural network is generally composed of input layer, convolution layer, pooling layer, full connection layer, and output layer. Its structure is shown in Figure 1. The difference between convolutional neural network and traditional neural network is that the convolutional neural network contains a feature extractor composed of convolution layer and pooling layer. Then, we briefly introduced the function of these components in CNN.

- (1) Deep CNN can directly take many kinds of data as the input data, such as image and audio. But in order to get better results, it is usually necessary to preprocess these data
- (2) The convolution layer is the key section of the convolutional neural network. It can convolute the input data to extract the features and transmit the convoluted results to the lower layer. The essence of convolution is to represent the input in another way. If the convolution layer is regarded as a black box, then we can regard the output as another representation of the input, while the training of the whole network is to train the parameters needed for this representation. Figure 2 is a diagram of the convolution. In Figure 2,  $w$  and  $b$  are the parameters needed for

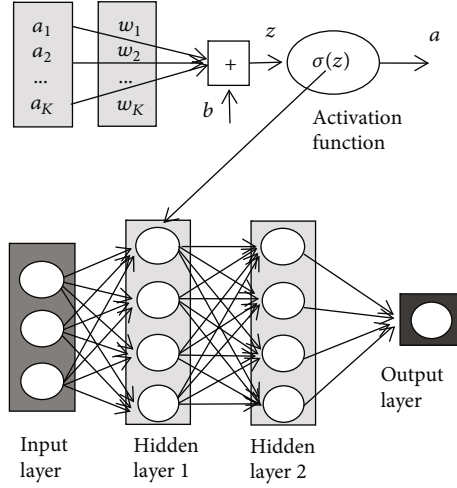


FIGURE 2: The operation diagram of the convolution layer.

network training. In the convolution layer, we need to apply activation function to nonlinear operation. The deep convolution network connects series of the small neural networks so as to form the deep neural network, which is mainly realized by local receptive field and weight sharing. The former refers that the neuron is only connected with its adjacent upper layer neuron, and the final global feature is formed by combining the learned local features. Weight sharing means that the same convolution kernel uses the same weight parameter when it operates on different local receptive fields. In this way, the connection between network layers can be reduced, so as to reduce the amount of parameter calculation in the process of network training

- (3) The pooling layer is also known as subsampling, which is a special data processing operation in the convolutional neural network. After the convolution operation, the data needs to be further processed by pooling to reduce the dimension of the extracted features. In this way, it can effectively remove the problem of large amount of calculation. Figure 3 shows the commonly used pooling method. Because the maximum pooling is achieved by extracting the point with the maximum value of local area, which has the advantage of retaining the factor with the largest influence in the feature area and effectively avoiding the information loss, we use the maximum pooling in the pooling layer
- (4) The full connection layer is usually used for classification at the end of the network. Unlike pooling and convolution, it is a global operation. Each node of the full connection layer is connected with all nodes of the previous layer, which is used to integrate the extracted features and transmit the signals to other full connection layers. Because of its all connected characteristics, the parameters of the general all connected layer are more, which is the layer consuming

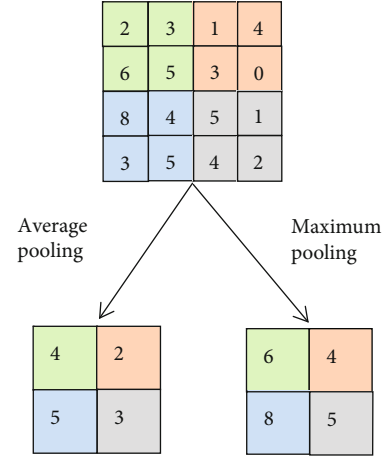


FIGURE 3: The pooling methods.

the most parameters in the network. Therefore, there are many parameters in the full connection layer, which consumes the most energy in the network

- (5) The output layer of the deep convolutional neural network completes different work according to the purpose of research. It usually uses the softmax function to calculate the classification results

**2.2. The Derivation of Deep CCN.** The essence of deep CNN is a mapping between input and output. It can learn a lot of mapping relations between input and output without any precise mathematical expression. As long as the convolution network is trained with known patterns, the network has the mapping ability between input and output pairs. The parameters of the deep convolutional neural network are defined as follows:

Let  $L$  be the number of network layers. In the convolution layer, the size of the convolution kernel is  $K$ . The dimension of the convolution kernel matrix is defined as  $F$ .  $p$  represents the filling size, and the steps of convolution kernel moving are  $S$  [28]. Before the data is transferred to the convolution layer, it is necessary to fill in missing data. Suppose the input data after filling is  $a^l$ . Then, the solution of deep CNN is as follows:

- (1) Initialize parameters

Initialize the weight parameters  $W$ , the bias  $b$  of the network, the maximum number of iterations  $T$ , and the iteration threshold  $\epsilon$ .

- (2) Training phase

*Step 1* (forward propagation). Select the training set and input them into the network. Calculate the corresponding output.

For  $l = 2$  to  $L-1$ :

If  $l$  is a convolution layer, then

$$a^l = \text{RELU}(z^l) = \text{RELU}(a^{l-1} \times W^l + b^l). \quad (1)$$

If  $l$  is a pool layer, then

$$a^l = \text{pool}(a^{l-1}). \quad (2)$$

If  $l$  is a full connection layer, then

$$a^l = \sigma(z^l) = \sigma(W^l a^{l-1} + b^l). \quad (3)$$

End for.

Finally, the output layer  $L$

$$a^L = \text{softmax}(z^L) = \text{softmax}(W^L a^{L-1} + b^L). \quad (4)$$

*Step 2* (backward propagation). Calculate the error between the actual output and the corresponding ideal output.

For  $l = L - 1$  to 2:

If  $l$  is a full connection layer, then

$$\delta^{i,l} = (W^{l+1})^T \delta^{i,l+1} \Theta \sigma(z^{i,l}). \quad (5)$$

If  $l$  is a convolution layer, then

$$\delta^{i,l} = \delta^{i,l+1} * \text{rot180}(W^{l+1}) \Theta \sigma(z^{i,l}). \quad (6)$$

If  $l$  is a pool layer, then

$$\delta^{i,l} = \text{upsample}(\delta^{i,l+1}) \Theta \sigma(z^{i,l}). \quad (7)$$

End for.

*Step 3* (update weights and bias). Adjust the weight matrix and bias according to the method of minimizing error.

For  $l = 2$  to  $L$ :

If  $l$  is a full connection layer, then

$$\begin{aligned} W^l &= W^l - \alpha \sum_{i=1}^m \delta^{i,l} (a^{i,l-1})^T, \\ b^l &= b^l - \alpha \sum_{i=1}^m \delta^{i,l}. \end{aligned} \quad (8)$$

If  $l$  is a convolution layer, then

$$\begin{aligned} W^l &= W^l - \alpha \sum_{i=1}^m \delta^{i,l} * a^{i,l-1}, \\ b^l &= b^l - \alpha \sum_{i=1}^m \sum_{\mu, \nu} (\delta^{i,l})_{\mu, \nu}. \end{aligned} \quad (9)$$

End for.

*Step 4.* If  $\|a^{(t+1)} - a^{(t)}\| < \varepsilon$  or  $t > T$ , the loop ends. Otherwise, return to Step 1.

(3) Output

Output the relation coefficient matrix  $W$  and bias  $b$ .

### 3. The Improved Deep Convolutional Neural Network

The deep convolutional neural network mainly relies on the convolution layer and the pooling layer for feature extraction and feature selection and uses the full connection layer for feature integration and the output layer for feature classification. In order to further improve the classification performance of the deep CNN model, this paper proposes to use the bagging integrated classification algorithm with better classification performance to replace the output layer function of the traditional convolutional neural network, so as to further optimize the classification ability of the convolutional neural network. The principle of the improved algorithm is to use the deep convolutional neural network for training and then extract the features integrated in the full connection layer as a new data set. Then, it will be trained as the input data of the bagging ensemble learning classifier. Finally, output the classification result based on the voting method. In other words, in the improved deep convolutional neural network, the softmax function of the output layer will be replaced by the bagging algorithm. In this way, we can not only use the convolutional neural network to extract the potential features of the data set but also use integrated learning to classify the features, so as to achieve a good effect of disease prediction.

*3.1. The Bagging Integration Classification.* As we know, the bagging method is a kind of parallel integrated classification algorithm which randomly selects training sets. The advantage of the bagging method is that it introduces bootstrap sampling in the training process. In this way, it can ensure that the training subsets are as independent as possible, so as to enhance the differences between the classifiers. Because these basic classifiers can analyze the features of the data set from different angles, they can reduce the generalization error of the algorithm and improve the classification accuracy. Moreover, compared with the ordinary single classifier, they can achieve better classification effect. At the same time, every training in the bagging can run in parallel, which greatly shortens the running time and improves the classification efficiency. What is more, the proportion of noise data

is usually relatively low in the practical application, so the interference of the noise data on the final classification results can be effectively alleviated by means of multiple random sampling of the bagging method [40].

In summary, the structure of the integrated learning layer is shown in Figure 4.

Meanwhile, define  $Y = \{-1, +1\}$  as a classification label set.  $T$  represents the number of the base classifiers. Then, the assumed function of the bagging method is

$$H(x) = \text{sign} \left( \sum_{i=1}^T h_i(x) \right). \quad (10)$$

The specific steps of the bagging algorithm are described in Algorithm 1.

**3.2. The Deep CNN Based on the Bagging Algorithm.** In the training stage of the deep CNN, the amount of data used is relatively large compared with the number of features, so there is no need to worry about the fitting problem of the network and no need to drop out the network. Firstly, input the data into the deep CNN model for training, and save the trained model which can achieve the best classification effect. The features integrated by the full connection layer of the saved model are the best features. Then, the best features saved are used as the input data of the integrated learning layer for classification. Figure 5 shows the flow of the improved deep CNN algorithm.

As shown in Figure 5, data cleaning, data balancing, data feature processing, and exception data processing can be collectively referred to as data preprocessing, so that the whole algorithm process can be divided into data preprocessing and training model. Because the data set used in this experiment is the original data recorded by the hospital, the data resources cannot be trained directly due to some irresistible factors. Firstly, we need to analyze the data and then clean the data set properly. Then, according to the fact that the distribution of medical data is unbalanced in real life, we need to balance the data after cleaning. After the data is balanced, the next step is to analyze the data features in the data set, obtain high-quality feature attributes, and process the abnormal data. Finally, put the preprocessed data input into the deep DNN for training, and finally, get the classification result of the trained data based on the integrated classification layer. The steps of the improved deep CNN are described in detail as follows.

- (1) *Data Cleaning.* This deals with the missing values in the original data set. It mainly solves the problems of data format inconsistency, data incompleteness, data record error, and so on.
- (2) *Data Balancing.* The data set chooses from the medical data in real life. Due to abnormal data, data value missing, and other reasons, the data after cleaning still has the problem of unbalanced distribution. In view of this, we use a few kinds of oversampling technology to process the data set, so that the sample data

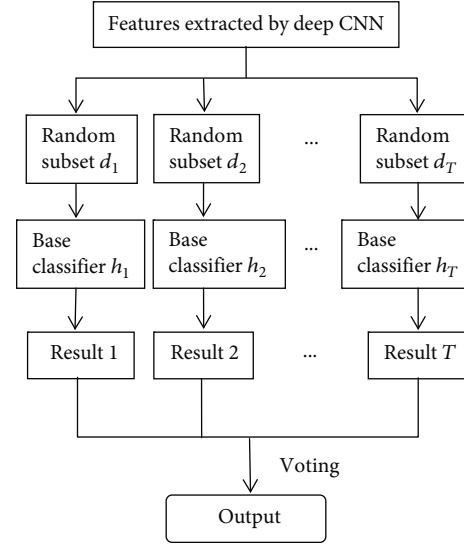


FIGURE 4: The structure of the integrated learning layer.

The description of the bagging algorithm.

**Input:**

The data set  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , the base classifiers  $L$ , and the number of the base classifiers  $T$ .

**Output:**

$$H(x) = \arg \max_{y \in Y} \sum_{i=1}^T \text{sign}(h_i(x) = y).$$

**Training:**

For  $i = 1, \dots, T$

$h_i = L(D, D_{bs}) / * D_{bs}$  is the bootstrap distribution/

End for

ALGORITHM 1

can reach equilibrium and ensure the accuracy of the experiment and the value of the experimental results.

- (3) *Data Feature Processing.* In this study, the embedding method is used to select features. We mainly use the xgboost algorithm to sort each feature according to its importance. At the same time, considering the opinions of diabetes experts, we screened out the important characteristics finally.
- (4) *Abnormal Data Processing.* Because the data set is manually recorded by the medical staff in the hospital, there are some noise data. In order to avoid the impact of these noise data on the experimental results, it is necessary to detect the outliers of these data and screen out the abnormal data. In the paper, the isolated forest algorithm is used to detect outliers, so as to obtain a smoother data set.
- (5) *Model Training.* Through the above processing, we can get the data set with balanced, smooth, and high feature contribution and use it as the input data of DNN for model training.
- (6) *Classification Training.* The trained data feature is used as the new input data of the bagging algorithm

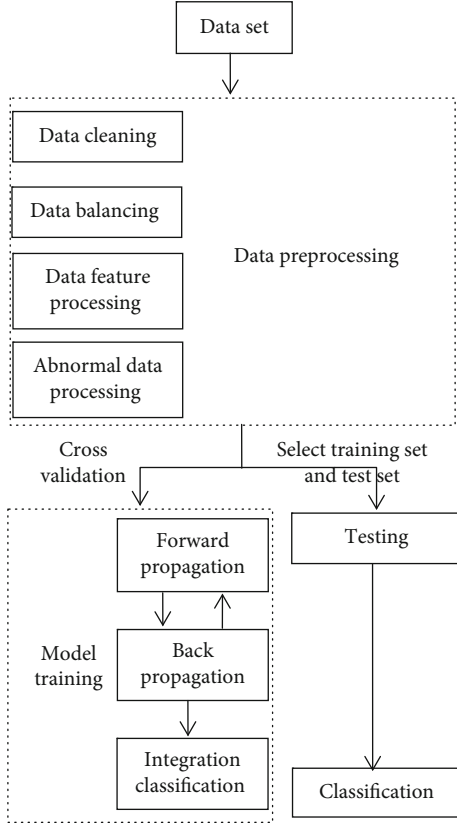


FIGURE 5: The flow of the improved deep CNN algorithm.

in the integrated classification layer to train the base classifiers.

- (7) *Output Results.* Finally, we will have tests on the improved deep DNN and output the classification results.

Algorithm 2 describes the detailed description of the deep CNN based on the bagging algorithm.

## 4. Experimental Studies

**4.1. The Experimental Design.** The experimental environment adopts Intel (R) core (TM) i7-9700u processor, 3.0GHz main frequency, 8G memory, and Windows 10 operating system. We use PyCharm 2018.1.4 (Professional Edition) as the experimental platform and python 3.6 as the coding tool. The algorithm is realized by building the tensor flow machine learning library for programming.

**4.2. The Experimental Data and Parameter Setting.** In this study, we adopt the data of diabetes patients admitted to 130 hospitals in the United States for 10 years as the experimental data [41, 42]. There are 10766 samples in the data set. And each sample contains 49 attribute columns and 1 label variable. The label variable 1 indicates that the patient is readmitted within 30 days, and 0 indicates that the patient is not readmitted within 30 days after discharge. The 49 attributes include personal information (such as patient number, race, gender, and age), diagnosis details (such as admission

The description of deep CNN based on the bagging algorithm.

**Input:**

The data set  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , where  $x_i$  is the  $i$ -th sample, the classification label is  $y_i \in \{0, 1\}$ ,  $T$  is the number of deep CNN iterations,  $L$  represents the number of network layers,  $\xi$  is base classifiers, and the number of the base classifiers is  $K$ .

**Output:**

$$H(f) = \arg \max_{y \in Y} \sum_{i=1}^K \text{sign}(h_i(f) = y).$$

**Training:**

Loop: for iterator to  $T$ .

Step 1 (forward propagation).

For  $i$  to  $n$

$$a^L = \text{soft max}(z^L) = \text{soft max}(W^L a^{L-1} + b^L).$$

Step 2 (back propagation).

For  $l = L - 1$  to 2:

$$\text{Update } W^l = W^l - \alpha \sum_{i=1}^m \delta^{i,l} * a^{i,l-1}.$$

$$\text{Update } b^l = b^l - \alpha \sum_{i=1}^m \sum_{\mu, \nu} (\delta^{i,l})_{\mu, \nu}.$$

$$\text{Save } S = \{(f_1, y_1), (f_2, y_2), \dots, (f_m, y_m)\}.$$

End loop.

Step 3 (the integration classification).

$$\text{Input } S = \{(f_1, y_1), (f_2, y_2), \dots, (f_m, y_m)\}.$$

For  $i = 1, \dots, K$

$$h_i = \xi(S, S_{bs})$$

End for

$$\text{Output } H(f) = \arg \max_{y \in Y} \sum_{i=1}^K \text{sign}(h_i(f) = y).$$

ALGORITHM 2

TABLE 1: The parameter settings of different network layers.

Network layer	Input	Filter	Step	Padding
Input layer	6 * 6 * 1	/	/	/
Convolution layer 1	6 * 6 * 1	3 * 3 * 32	1	Same
Pool layer 1	6 * 6 * 32	2 * 2	1	Valid
Convolution layer 2	5 * 5 * 32	3 * 3 * 64	1	Same
Pool layer 2	5 * 5 * 64	2 * 2	1	Valid
Full connection layer 1	128	/	/	/
Full connection layer 2	128	/	/	/

type, outpatient diagnosis record), medication records (such as medication quantity and diabetes drugs), and examination records. Therefore, the data set has rich samples and complete characteristics on the diagnosis and treatment of diabetes. As the data of diabetes used in this study have only 36 features after data preprocessing, we limit the depth of convolutional neural network. This is different from the traditional convolutional neural network with deep network structure for image processing. In addition, we have also made customization in the design of convolution core and pool filter, and its structural parameters are shown in Table 1.

**4.3. Experimental Results and Analysis.** In the process of experimental analysis, we usually judge the performance of classifier based on the accuracy of classification results.



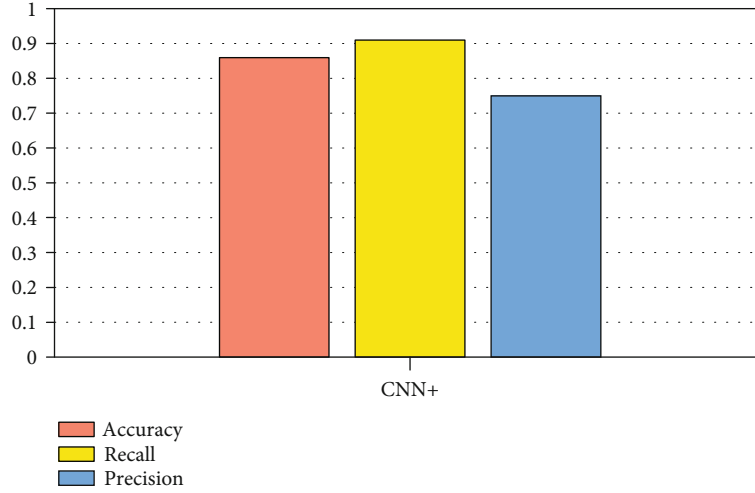


FIGURE 6: The application of CNN+ based on the data of diabetic medical records.

However, for some application scenarios, accuracy as an evaluation index has certain limitations, and it cannot always effectively evaluate the work of a classifier. In view of this, we need to introduce other indicators to evaluate the experimental results. Therefore, this paper uses accuracy, recall, precision, and F-measure to evaluate the performance of the improved deep CNN. Then, each indicator is defined as follows:

$$\begin{aligned}
 \text{accuracy} &= \frac{(TP + TN)}{(P + N)}, \\
 \text{recall} &= \frac{TP}{P}, \\
 \text{precision} &= \frac{TP}{(TP + FP)}, \\
 \text{F-measure} &= \frac{2TP}{(2TP + FP + FN)},
 \end{aligned} \tag{11}$$

where  $P$  is the number of samples identified as positive in the test set,  $TP$  is the number of samples correctly predicted as positive by the classifier, and  $FP$  is the number of samples incorrectly predicted as negative by the classifier. F-measure is the harmonic average of recall rate and precision, and it is a comprehensive index that can better reflect the classification performance. In the experiments based on the medical data, researchers pay more attention to the recall rate than the accuracy. In this experiment, if the recall rate is higher, it can show that the possibility of judging patients' bad recovery as good recovery is lower. And if the prediction is more accurate, it is more helpful to achieve the experimental goal.

After completing the above experimental parameter settings and experimental index settings, firstly, we trained the data set with the deep CNN. And in the training, we set the number of iterations as 1000 and then save the parameters of the model. Then, we extract the features of CNN in the full connection layer and segmented the data. Then, we use the

bagging integrated classification algorithm to train the feature data extracted from the full connection layer and use the voting method to vote the experimental results, so as to obtain the classification results. As in the paper, we use recall, precision, and accuracy to measure the effect of classification; Figure 6 shows the variance and average of the above three indicators.

In order to verify the effectiveness of the algorithm in this paper, we have carried out 4000 iterations of CNN+, and Figure 7 shows the curve of the accuracy and the loss of network training with the number of iterations.

As seen in Figure 7(a), the training accuracy gradually increases with the number of iterations and then tends to be stable. It can be seen from Figure 7(b) that the loss value gradually decreases with the increase of the number of iterations and then starts to stabilize near the smaller value, indicating that the trained network model has better stability.

Finally, in order to verify the necessity of data balance processing for the original data in the process of the experiment, we carry out the experiment based on the unbalanced data and balanced data. Meanwhile, the compared experiments were conducted among the traditional convolutional neural network and decision tree, random forest, and naive Bayes. The experimental results are illustrated in Tables 2 and 3, respectively.

As seen from Tables 2 and 3, we can find the following information:

- (1) Whether in the balanced data set or in the nonbalanced data set, CNN+ can achieve better experimental results in general. Compared with the experimental results of CNN, decision tree, random forest, and naive Bayes classifier, the CNN+ accuracy, recall rate, and precision are significantly improved. It shows that the CNN+ algorithm has higher classification accuracy and better stability
- (2) In addition, comparing the experimental results of Tables 2 and 3, it can be found that using the data

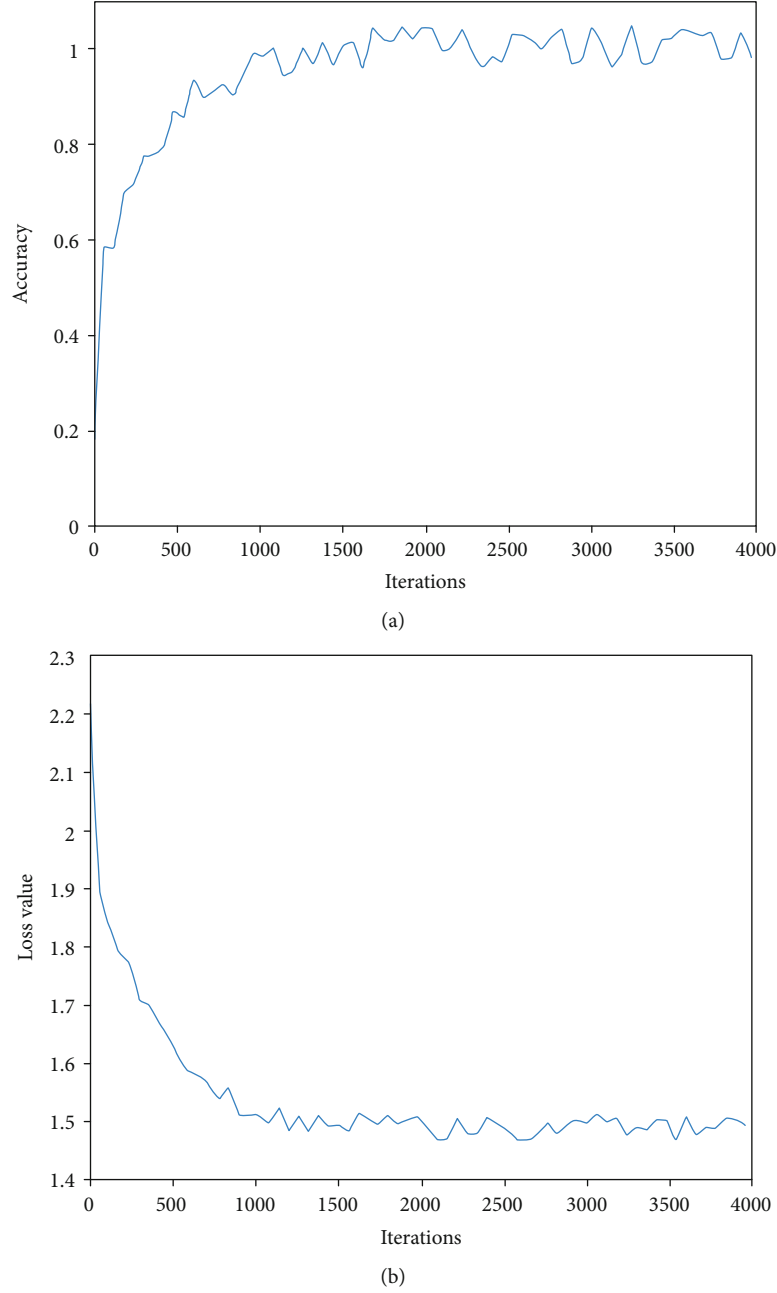


FIGURE 7: The curve of accuracy and loss value with iteration times.

TABLE 2: The segmentation accuracy of brain tissue based on various algorithms.

Models	Accuracy	Recall	Precision	F-measure
Decision tree	0.81	0.85	0.70	0.77
Random tree	0.80	0.83	0.71	0.77
Naive Bayes	0.78	0.81	0.70	0.75
CNN	0.81	0.84	0.73	0.78
CNN+	0.83	0.88	0.73	0.79

set after data equalization processing is indeed helpful to improve the classification accuracy of the classifier. Comparing the experimental results of the

TABLE 3: The segmentation accuracy of brain tissue based on various algorithms.

Models	Accuracy	Recall	Precision	F-measure
Decision tree	0.82	0.86	0.72	0.78
Random tree	0.82	0.81	0.72	0.77
Naive Bayes	0.80	0.82	0.73	0.78
CNN	0.83	0.90	0.72	0.80
CNN+	0.84	0.91	0.74	0.81

above algorithms, we can see that the experimental results on the balanced data set are obviously better than those on the unbalanced data set

- (3) In summary, the prediction stability of the algorithm is as important as the prediction accuracy in the field of medical research, which is related to the patient's condition recovery. The improved deep convolutional neural network can achieve good accuracy and precision in the detection of diabetes data. And it is easy to learn and train, has high stability, and shows certain advantages compared with other classifiers. Therefore, it provides the possibility for its application in practice

At the same time, we also compared the research results of other researchers based on this data set. For example, Strack et al. also conducted experiments on the same data set [42]. They used HbA1c, or glycosylated hemoglobin, as an indicator to predict the probability of readmission of patients with diabetes after treatment. In the study, the researchers also noticed that few researchers studied diabetes care during hospitalization, so they proposed to use the data set of inpatient care records to predict the relationship between the patients' readmission in the short term. The key of their study was to use multivariate logistic regression to fit the relationship between HbA1c value and the short-term readmission of patients. However, due to the small sample size of HbA1c value in this data set, the results of this study only prove that there is a relationship between HbA1c value and the probability of readmission. Compared with the study of Strack et al., we considered a number of admission records. Firstly, we use the deep convolutional neural network to extract features and make full use of the useful features in the data set. At the same time, we combine with the bagging integrated method for classification, and the reliable prediction effect is achieved.

## 5. Conclusion

With the increasing rate of epidemic disease (like cancer and diabetes), the human health has been seriously threatened, so it is necessary to study the prevention and treatment of epidemic diseases. In view of this, the deep learning technology was used to analyze the data of diabetes inpatient cases, which can help mine valuable treatment rules from them and assist doctors in diagnosis and treatment and improve treatment efficiency. Due to the large number of features and large amount of data in this data set, an improved deep convolutional neural network algorithm is proposed to predict the condition of diabetes, so as to improve the classification accuracy of deep CNN. The improved deep CNN can take the advantages of the bagging integrated classification algorithm based on the deep CNN. In order to demonstrate the effectiveness of the improved algorithm, the comparative experiments were conducted among the traditional neural network and classifiers. The experimental results prove that the improved algorithm does improve the classification accuracy and stability. Because of the large sample size of the data set, good experimental results can be achieved in the convolutional neural network. However, the convolutional neural network requires a large amount of data and data characteristics; the experimental results for the small sample data set

are not ideal. The classification of small samples is worth further study.

## Data Availability

All relevant data are within the paper.

## Conflicts of Interest

The author declares that there are no conflicts of interest.

## References

- [1] T. F. Blaschke, M. Lumpkin, and D. Hartman, "The World Health Organization prequalification program and clinical pharmacology in 2030," *Clinical Pharmacology and Therapeutics*, vol. 107, no. 1, pp. 68–71, 2019.
- [2] A. M. Carracher, P. H. Marathe, and K. L. Close, "International Diabetes Federation 2017," *Journal of Diabetes*, vol. 10, no. 5, pp. 353–356, 2018.
- [3] P. Qian, H. Friel, M. S. Traugher et al., "Transforming UTE-mDixon MR abdomen-pelvis images into CT by jointly leveraging prior knowledge and partial supervision," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, p. 1, 2020.
- [4] K. Xia, H. Yin, P. Qian, Y. Jiang, and S. Wang, "Liver semantic segmentation algorithm based on improved deep adversarial networks in combination of weighted loss function on abdominal CT images," *IEEE Access*, vol. 7, pp. 96349–96358, 2019.
- [5] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, 2019.
- [6] P. Qian, Y. Chen, J.-W. Kuo et al., "mDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 819–832, 2020.
- [7] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [8] P. Qian, K. Xu, T. Wang et al., "Estimating CT from MR abdominal images using novel generative adversarial networks," *Journal of Grid Computing*, vol. 18, no. 2, pp. 211–226, 2020.
- [9] P. Qian, C. Xi, M. Xu et al., "SSC-EKE: semi-supervised classification with extensive knowledge exploitation," *Information Sciences*, vol. 422, pp. 51–76, 2018.
- [10] K. Xia, H.-s. Yin, and Y.-d. Zhang, "Deep semantic segmentation of kidney and space-occupying lesion area based on SCNN and ResNet models combined with SIFT-Flow algorithm," *Journal of Medical Systems*, vol. 43, no. 1, 2019:1–2:12.
- [11] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [12] V. Miljanović, M. Tjepkema-Cloostermans, and M. V. Putten, "O20 Deep learning with convolutional neural networks for detection of interictal epileptiform discharges," *Clinical Neurophysiology*, vol. 128, no. 9, p. e186, 2017.

- [13] G. Swapna, R. Vinayakumar, and K. P. Soman, "Diabetes detection using deep learning algorithms," *ICT Express*, vol. 4, no. 4, pp. 243–246, 2018.
- [14] G. Swapna, K. Soman, and R. Vinayakumar, "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals," *Procedia Computer Science*, vol. 132, pp. 1253–1262, 2018.
- [15] A. E. Kitabchi, G. E. Umpierrez, J. M. Miles, and J. N. Fisher, "Hyp-erglycemic crises in adult patients with diabetes," *Diabetes Care*, vol. 32, no. 7, pp. 1335–1343, 2009.
- [16] T. Cukierman, H. C. Gerstein, and J. D. Williamson, "Cognitive decline and dementia in diabetes—systematic overview of prospective observational studies," *Diabetologia*, vol. 48, no. 12, pp. 2460–2469, 2005.
- [17] T. K. Motawi, H. A. Darwish, M. A. Hamed, N. S. el-Rigal, and A. F. A. Naser, "A therapeutic insight of niacin and coenzyme Q10 against diabetic encephalopathy in rats," *Molecular Neurobiology*, vol. 54, no. 3, pp. 1601–1611, 2017.
- [18] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, and P. Hufnagel, "Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology," *Computerized Medical Imaging and Graphics*, vol. 61, pp. 2–13, 2017.
- [19] J. Li, B. Zhang, G. Lu, J. You, and D. Zhang, "Body surface feature-based multi-modal learning for diabetes mellitus detection," *Information Sciences*, vol. 472, pp. 1–14, 2019.
- [20] V. Vijayanv and A. Ravikumar, "Study of data mining algorithms for prediction and diagnosis of diabetes mellitus," *International Journal of Computer Applications*, vol. 95, no. 17, pp. 12–16, 2014.
- [21] Z. Alhassan, A. S. McGough, R. Alshammari, T. Daghestani, D. Budgen, and N. Al Moubayed, "Type 2 diabetes mellitus diagnosis from time series clinical data using deep learning models," in *Artificial Neural Networks and Machine Learning – ICANN 2018*, pp. 468–478, Springer, Cham, 2018.
- [22] A. Chao, C. F. Sum, B. Lam, A. Cheng, S. Low, and S. C. Lim, "Effect of bariatric surgery on diabetic nephropathy in obese type 2 diabetes patients in a local retrospective 2 year study," *Diabetes Research and Clinical Practice*, vol. 120, pp. S190–S190, 2016.
- [23] T. Araki, N. Ikeda, D. Shukla et al., "PCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: a link between carotid and coronary grayscale plaque morphology," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 137–158, 2016.
- [24] N. Sneha and T. Gangil, "Analysis of diabetes mellitus for early prediction using optimal features selection," *Journal of Big Data*, vol. 6, no. 1, 2019.
- [25] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of deep belief networks for natural language understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 778–784, 2014.
- [26] L. P. Zhao, H. Bolouri, M. Zhao, D. E. Geraghty, Å. Lernmark, and The Better Diabetes Diagnosis Study Group, "An object-oriented regression for building disease predictive models with multiallelic HLA genes," *Genetic Epidemiology*, vol. 40, no. 4, pp. 315–332, 2016.
- [27] H. Y. Khaw, F. C. Soon, J. H. Chuah, and C. O. Chow, "Image noise types recognition using convolutional neural network with principal components analysis," *IET Image Processing*, vol. 11, no. 12, pp. 1238–1245, 2017.
- [28] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.
- [29] D. Ravi, C. Wong, B. Lo, and G. Z. Yang, "A deep learning approach to onnode sensor data analytics for mobile or wearable devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 56–64, 2017.
- [30] A.-r. Mohamed, T. N. Sainath, G. Dahl, B. Ramabhadran, G. E. Hinton, and M. A. Picheny, "Deep belief networks using discriminative features for phone recognition," in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 5060–5063, Prague, Czech Republic, May 2011.
- [31] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," *Neural Networks*, vol. 16, no. 5-6, pp. 555–559, 2003.
- [32] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequ-ential deep learning for human action recognition," in *Human Behavior Understanding, Lecture Notes in Computer Science*, pp. 29–39, Springer, Berlin Heidelberg, 2011.
- [33] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2011.
- [34] K. Stone and J. M. Keller, "Convolutional neural network approach for buried target recognition in FL-LWIR imagery," *Proceedings of SPIE*, vol. 5, pp. 112–115, 2014.
- [35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016.
- [36] O. Anya and H. Tawfik, "Designing for practice-based context-awareness in ubiquitous E-health environments," *Computers & Electrical Engineering*, vol. 61, pp. 312–326, 2017.
- [37] L. Mansn, Y. Zhang, and K. Fredrik, "Robust abdominal organ segmentation using regional convolutional neural networks," *Applied Soft Computing*, vol. 70, pp. 465–471, 2018.
- [38] X. Liu, F. Hou, H. Qin, and A. Hao, "Multi-view multi-scale CNNs for lung nodule type classification from CT images," *Pattern Recognition*, vol. 77, pp. 262–275, 2018.
- [39] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60–88, 2017.
- [40] R. W. Johnson, "An introduction to the bootstrap," *Teaching Statistics*, vol. 23, no. 2, pp. 49–54, 2001.
- [41] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Computers in Biology and Medicine*, vol. 89, pp. 135–143, 2017.
- [42] B. Strack, J. P. DeShazo, C. Gennings et al., "Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records," *BioMed Research International*, vol. 2014, 11 pages, 2014.

## Research Article

# Spatial-Frequency Feature Learning and Classification of Motor Imagery EEG Based on Deep Convolution Neural Network

Minmin Miao <sup>1,2</sup>, Wenjun Hu,<sup>1,2</sup> Hongwei Yin,<sup>1,2</sup> and Ke Zhang <sup>1,2</sup>

<sup>1</sup>School of Information Engineering, Huzhou University, Huzhou 313000, China

<sup>2</sup>Zhejiang Province Key Laboratory of Smart Management & Application of Modern Agricultural Resources, Huzhou University, Huzhou 313000, China

Correspondence should be addressed to Minmin Miao; 15005142754@163.com

Received 26 November 2019; Revised 30 January 2020; Accepted 20 February 2020; Published 20 July 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Minmin Miao 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.

EEG pattern recognition is an important part of motor imagery- (MI-) based brain computer interface (BCI) system. Traditional EEG pattern recognition algorithm usually includes two steps, namely, feature extraction and feature classification. In feature extraction, common spatial pattern (CSP) is one of the most frequently used algorithms. However, in order to extract the optimal CSP features, prior knowledge and complex parameter adjustment are often required. Convolutional neural network (CNN) is one of the most popular deep learning models at present. Within CNN, feature learning and pattern classification are carried out simultaneously during the procedure of iterative updating of network parameters; thus, it can remove the complicated manual feature engineering. In this paper, we propose a novel deep learning methodology which can be used for spatial-frequency feature learning and classification of motor imagery EEG. Specifically, a multilayer CNN model is designed according to the spatial-frequency characteristics of MI EEG signals. An experimental study is carried out on two MI EEG datasets (BCI competition III dataset IVa and a self-collected right index finger MI dataset) to validate the effectiveness of our algorithm in comparison with several closely related competing methods. Superior classification performance indicates that our proposed method is a promising pattern recognition algorithm for MI-based BCI system.

## 1. Introduction

Brain computer interface (BCI) technology [1–3] uses multiple brain function signals, including scalp Electroencephalogram (EEG) [4], Local Field Potentials (LFPs) [5], and Electrocorticography (ECoG) [6], to establish a direct communication channel between human brain and external devices. This characteristic of BCI is extremely important for patients with severe brain nerve damage, since the normal communication channel for such patients has been blocked [7]. Considering the convenience, safety, and cost, scalp EEG is most frequently used in BCI fields. Among various BCI control paradigms, motor imagery- (MI-) based BCI system is a very important branch. Via MI-based BCI system, users can control robots or external machines merely by movement imagination, without the intervention of peripheral nerve. Due to its great potential application value in motor

function rehabilitation [8], motor function assistance, and so forth, MI-based BCI system has been widely concerned.

EEG pattern recognition is an important part of MI-based BCI system; traditional EEG pattern recognition algorithm mainly includes two steps, namely, feature extraction and feature classification. In feature extraction stage, common spatial pattern (CSP) algorithm [9–11] is the most commonly used algorithm, but several factors would affect the performance of CSP algorithm, such as the spatial channels, frequency bands of sensorimotor rhythm signal, and time windows. It is worth noticing that most of the research efforts have been dedicated to optimizing the frequency bands for significant CSP features extraction. Filter band common spatial pattern (FBCSP) algorithm [12] is a benchmark for spatial-frequency feature learning and has been widely applied to MI EEG analysis. More recently, a sparse filter band common spatial pattern (SFBCSP)



algorithm [13] has been proposed to select most significant CSP features in multiple frequency bands via sparse regression. In the feature classification stage, many machine learning algorithms, such as linear discriminant analysis (LDA) [14], support vector machine (SVM) [13], and logistic regression (LR) [15], are used to classify different EEG patterns of motor imageries. We notice that some more sophisticated algorithms have been also proposed for MI EEG classification in recent years. Jiao et al. [16] developed a sparse group representation model (SGRM) in which a test sample can be estimated as a linear combination of samples in a composite dictionary matrix composed by CSP features from multiple subjects. Jin et al. [17] introduced a sparse Bayesian extreme learning machine (SBELM) method for MI-related EEG classification by combining the advantages of both extreme learning machine (ELM) and sparse Bayesian learning. As we can see that, for most traditional MI EEG pattern recognition algorithms, feature extraction and feature classification are separated; however, these two stages usually have different objective functions; hence it is easy to cause information loss [18].

The convolutional neural network (CNN) is based on deep learning theory and has been widely used in image recognition [19], speech recognition [20], and other fields. Its main characteristics are weight sharing and local perception so that the number of weight parameters is greatly reduced compared with the ordinary deep neural network. In addition, CNN implements feature learning and classification in the network simultaneously, which is simpler and clearer than the traditional pattern recognition method. Furthermore, less information is lost in this procedure.

In the past few years, deep learning techniques, i.e., deep neural networks, have been investigated to deal with complex brain function signals [21]. In terms of the research on MI EEG, the independent CNN and CNN-based hybrid models are widely used. Lee and Choi [22] proposed obtaining time-frequency representations of EEGs using continuous wavelet transform (CWT) as the input of CNN model. Tabar and Halici [23] applied short time Fourier transform (STFT) method to convert EEG temporal series into 2D images and used CNN and stacked autoencoders (SAE) for MI EEG classification. Uktveris and Jusas [24] used Fast Fourier Transform (FFT) energy map method for CNN and achieved satisfactory results for four-class MI EEG classification. Liu et al. [25] introduced a multiscale deep CNN method to deal with the representation for MI EEG signals. Hartmann et al. [26] investigated how CNN represented spectral features through the sequence of intermediate stages of the network. Wang et al. [27] devised a CNN-based method for MI EEG feature extraction and adopted weak classifier for feature selection. Tan et al. [28] trained a deep neural network with CNN and recurrent neural network (RNN) for the EEG classification task. Yang et al. [29] investigated the classification of multiclass MI EEG signals by augmented CSP (ACSP) features and CNN model. Tang et al. [30] constructed a 5-layer CNN model based on the spatiotemporal characteristics of EEG for MI tasks classification. In this paper, a novel deep learning approach is proposed for classification of MI EEG signal. Unlike all

these above works, we do not use any complex algorithm, such as CWT, STFT, FFT, and ACSP, for two-dimensional feature map generation. Besides, different from the work in [30] which studied the spatiotemporal characteristics of MI EEG, this work offers an insight into the spatial-frequency features of MI EEG. Furthermore, rather than exploiting the spatial-frequency characteristics of EEG by FBCSP or SFBCSP, we herein propose learning and classifying the spatial-frequency features of MI EEG simultaneously in a unified CNN framework. Specifically, we convert raw EEG data to image representation by computing the energies of multichannel EEG signals in multiple frequency bands at first. Afterward, a novel multilayer CNN model is designed, and the spatial characteristics of MI EEG are analyzed according to the obtained weight parameters of convolution layers. Finally, with a public dataset and a self-collected right index finger motion imagination dataset, extensive experimental comparisons are carried out between our method and several closely related machine learning algorithms.

## 2. Materials and Methods

*2.1. Dataset Description.* In this study, in order to better evaluate the effectiveness of the proposed algorithm, we used two different datasets for analysis. The first dataset is public BCI competition III dataset IVa, in which EEG signals are collected from five subjects (denoted by aa, al, av, aw, and ay) using 118 electrode amplifier. There are two types of MI tasks in the dataset, namely, right hand MI and right foot MI. The time duration of each MI trial is 3.5 seconds, and the sampling rate is set to 1000 Hz. For each subject, there are 280 MI trials (140 trials for hand MI and 140 trials for foot MI) in total. More details about this dataset can be found at <http://www.bbci.de/competition/iii/>. The second dataset is right index finger motion imagination dataset (denoted by Finger Dataset) which was collected by us. In this dataset, EEG signals are collected from five subjects (denoted by S1, S2, S3, S4, and S5) using 21 electrode amplifier. There are two types of MI tasks in the dataset, namely, finger movement imagination and rest state. The time duration of each MI trial is 4 seconds, and the sampling rate is also set to 1000 Hz. It should be noted that different subjects in this dataset have different numbers of MI trials. For each subject, the exact numbers of finger movement imagination and rest state trials are denoted by the following: subject (finger movement imagination, rest state) S1 (58, 58), S2 (59, 48), S3 (52, 63), S4 (56, 57), and S5 (62, 39). For the BCI competition III dataset IVa, we selected the first 260 trials as the training set, the following 10 trials as the testing set, and the last 10 trials as the validation set. For this public dataset, to more reliably evaluate the classification performance, we also implemented 10-fold cross-validation to compute the average classification accuracy. In each fold of 10-fold cross-validation, each part was for testing and the remaining nine parts were for training (90%) and validation (10%). For Finger Dataset with small sample size, we randomly selected about 80% of the samples as the training set and the remaining samples for testing and validation. The numbers of training samples, validation samples, and testing samples

are denoted by S1 (92, 12, 12), S2 (85, 11, 11), S3 (93, 11, 11), S4 (89, 12, 12), and S5 (81, 10, 10).

## 2.2. Methods

**2.2.1. EEG Data Representation Transform.** Before applying CNN, we firstly convert raw EEG data to image representation. For Finger Dataset, since the sensorimotor rhythm usually appears in the frequency bands of 8–14 Hz ( $\mu$  rhythm) and 18–26 Hz ( $\beta$  rhythm), we use bandpass filtering to obtain 8–30 Hz signal component from the original EEG signal. In addition, EEG signals from 0.5 seconds to 2.5 seconds after the appearance of visual cue are usually used for pattern analysis. In this study, the signals in this period are also extracted for subsequent processing. After the above processing, each sample can be represented as a matrix of size  $21 \times 2000$ , in which 2000 is the number of sampling points and 21 is the number of electrodes. It should be noted that although the sensorimotor rhythm usually appears in the frequency range of 8–30 Hz, the frequency bands related to MI tasks vary among different subjects, and the optimal frequency band is mostly a local narrow band. In order to learn more precise frequency information, we decompose the EEG signal in the range of 8–30 Hz into 10 subbands, the width of each subband is 4 Hz, and the overlap between adjacent subbands is 2 Hz. After that, each sample can be represented as a matrix of size  $21 \times 10 \times 2000$ , where 10 is the number of frequency subbands. Then, for the EEG signal in each frequency subband of each spatial electrode, we calculate its signal energy as follows:

$$p = \log(\text{var}(\mathbf{x})), \quad (1)$$

where  $\text{var}(\mathbf{x})$  is the variance of EEG signal sequence  $\mathbf{x}$ . Thus, each sample can be represented as a matrix of size  $21 \times 10$ , and each element of the matrix represents the energy of EEG signal in a certain subband of a certain EEG electrode. For each subject in the dataset, we then normalize the EEG energy as follows:

$$\mathbf{P}_{i,j}^r = \frac{\mathbf{P}_{i,j}^r - m_{i,j}}{\delta_{i,j}}, \quad (2)$$

where  $\mathbf{P}$  is the energy matrix of each sample,  $r$  denotes the index of each sample,  $m_{i,j}$  denotes the average energy of all samples at this location, and  $\delta_{i,j}$  is the corresponding standard deviation. After the above steps of signal processing, the original EEG signal is transformed into image representation, in which the EEG electrodes are distributed along the vertical axis and the frequency subbands are distributed along the horizontal axis.

For BCI competition III dataset IVa, the raw EEG data is also bandpass filtered within the range of 8–30 Hz. It should be noted that there are 118 electrodes in this dataset. In order to reduce the burden of subsequent calculation and remove the influence of redundant channels, according to the recommendation of literature [18], we extract the EEG signals from 49 channels for subsequent analysis. After the above processing, each sample can be represented as a matrix of size  $49 \times 3500$ , in which 3500 is the number of sampling

points and 49 is the number of electrodes. Afterward, frequency domain decomposition, 0.5 to 2.5 seconds' time period segmentation, energy extraction, and normalization processing are also carried out. Then, each sample can be represented as a matrix of size  $49 \times 10$ .

In order to observe the spatial-frequency characteristics of EEG after preprocessing, for each subject in the two datasets, we divided all samples into two groups according to the categories of MI and calculated the mean value of each group for comparison. Figure 1 and Figure 2 show the distributions of EEG spatial-frequency energy characteristics of subjects in Finger Dataset and the BCI competition III dataset IVa under different MI states, respectively.

From Figure 1, it can be seen that, compared with the rest state, the energy at the specific electrode decays when finger MI is conducted, so there is a pattern difference in this area, and this phenomenon mainly occurs on the opposite side of the motion imagining limb, such as electrode C3 (index 10); all of the above phenomena conform to the ERD theory [31]. Further observation shows that this pattern difference has obvious frequency domain characteristics; for example, it is most obvious in the 8–14 Hz frequency band. From Figure 2, it can be observed that the energy attenuation of hand motion imagination is more intense than foot motion imagination, so this relatively stable pattern difference appears in several local spatial-frequency blocks. It should be noted that, from Figure 1 and Figure 2, there are different forms of spatial-frequency pattern differences among different subjects. Specifically, the locations of significant spatial-frequency blocks and degrees of differences are different. All these factors will affect the final pattern recognition results. At the same time, we need to focus on whether CNN model can learn spatial-frequency features adaptively.

**2.2.2. CNN Structure Design.** Through the descriptions in the previous section, we can see that the MI EEG signal has a very obvious spatial-frequency characteristic, which is also consistent with the basic ERD theory, and has a relatively stable pattern difference after being converted into image form. In order to learn the MI EEG characteristics adaptively and carry out pattern recognition, this section designs a novel multilayer CNN structure, as shown in Figure 3. The proposed CNN model consists of five layers; the first layer is the input layer which is specially designed to capture the spatial-frequency characteristics of MI EEG signals. Note that we do not use any complex algorithm and retain the energy change information in spatial-frequency domain more completely. The next two layers are the convolution layers, mainly for spatial-frequency feature extraction. It should be noted that we use one-dimensional convolutional filters in the first convolution layer for better analysis of spatial features. The last two layers are the fully connected layer and the softmax layer; these two layers mainly complete the classification task. To prevent overfitting, we implemented dropout regularization before the output layer. The specific description of each layer is as follows.

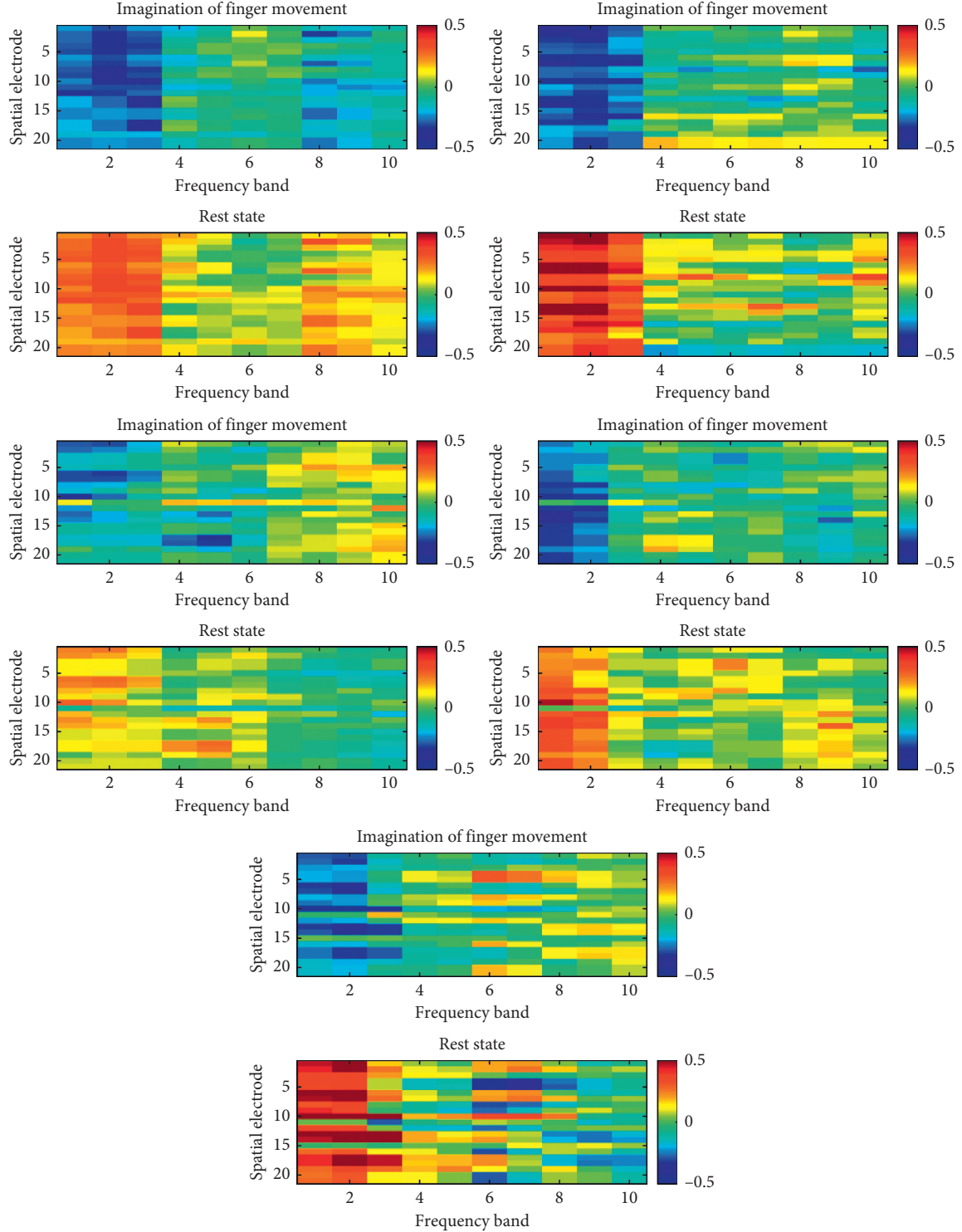


FIGURE 1: Spatial-frequency energy distributions of EEG in different motor imagery states (Finger Dataset).

- (1) Input layer: the input is the image form of the preprocessed motor imagery EEG sample. For the Finger Dataset, the input is the matrix of size  $21 \times 10$ , where 21 is the number of electrodes and 10 is the number of frequency subbands. For BCI competition III dataset IVa, the input is the matrix of size  $49 \times 10$ , where 49 is the number of electrodes and 10 is the number of frequency subbands.
- (2) Convolution layer C1: the main function of this layer is to filter the input signal in the spatial domain (i.e., different electrodes are assigned with different weight values). Therefore, one-dimensional convolution operation (convolution of spatial electrodes on the vertical axis) is carried out in this layer, and the convolution kernel slides along the horizontal axis. This shows the characteristics of CNN, namely,

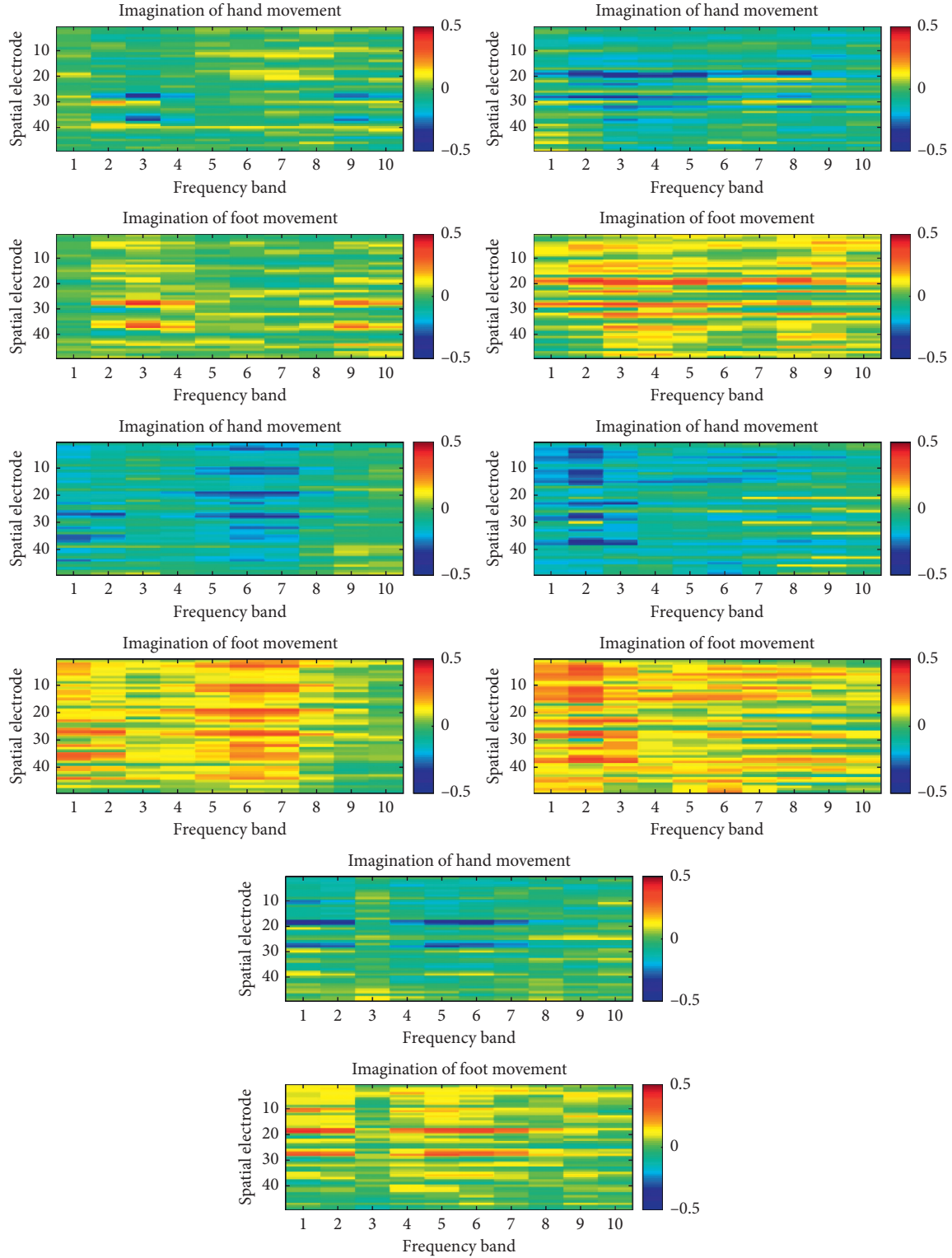


FIGURE 2: Spatial-frequency energy distributions of EEG in different motor imagery states (BCI competition III dataset IVa).

weight sharing and local perception. It should be noted that, since there is no hybrid of any time or frequency domain information, by observing the spatial filter obtained from this layer, we can understand part of the spatial characteristics of motor imagery EEG. In this study, we use 6 spatial filters;

thus we can get 6 feature maps after spatial convolution. For the Finger Dataset, the convolution kernel size is set to  $[21 \times 1]$ . For the BCI competition III dataset IVa, the convolution kernel size is set to  $[49 \times 1]$ . Since there are 10 frequency subbands, the size of the feature map is  $[1 \times 10]$ .

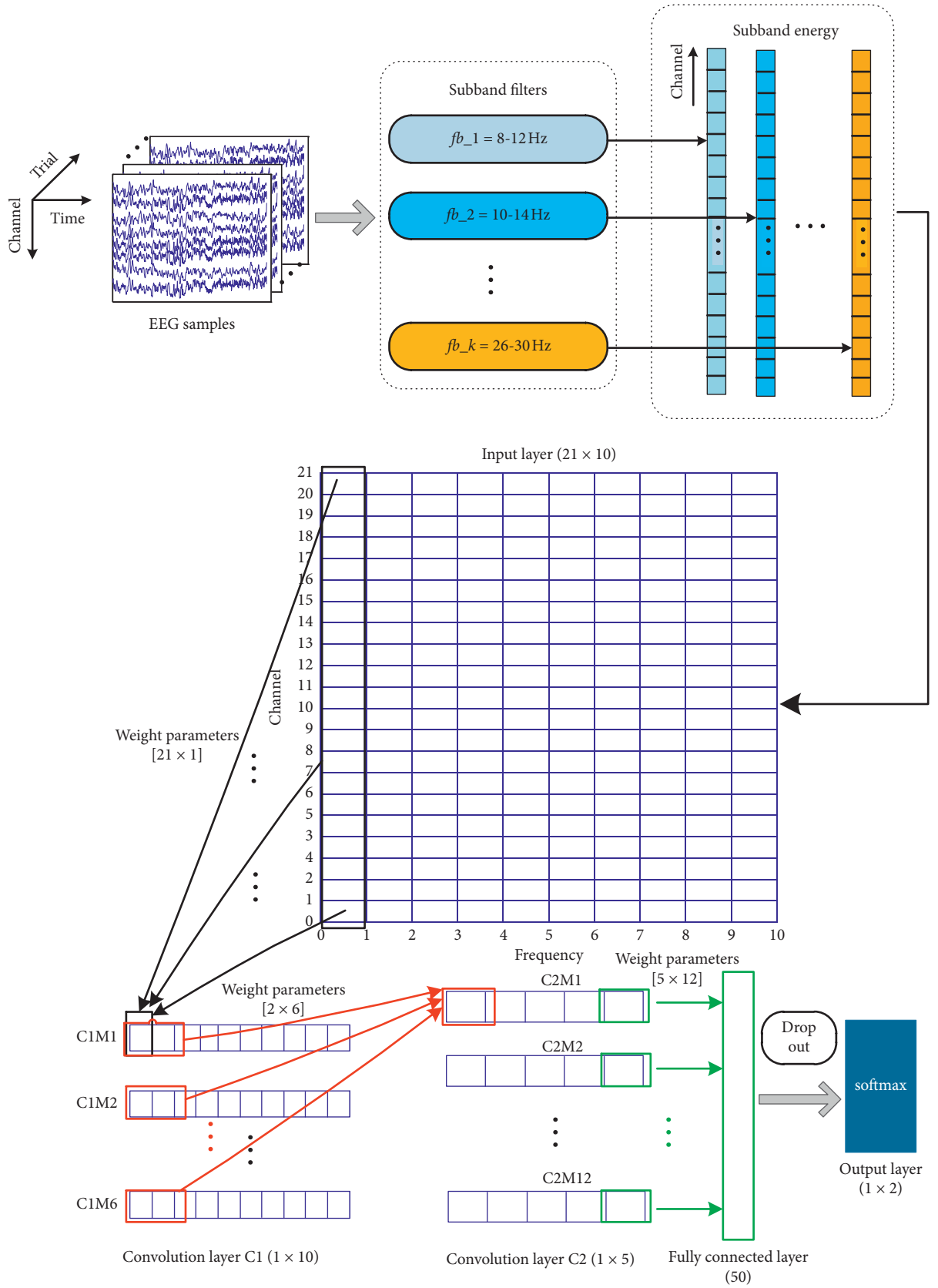


FIGURE 3: The structure diagram of convolutional neural network for motor imagery EEG pattern recognition (Finger Dataset).



- (3) Convolution layer C2: this layer mainly combines the features of the previous layer and learns more complex and abstract spatial-frequency features. The convolution operation of this layer still adopts the concept of local connection and weight sharing. Each convolution operation uses two feature values from all the feature maps of the previous layer and slides along the horizontal axis. The sliding step is also set to two to reduce the number of parameters. In this layer, we set up 12 filters in total, the size of the convolution kernel is  $[2 \times 6]$ , and finally we obtain 12 feature maps of size  $[1 \times 5]$ .
- (4) Full connection layer: this layer mainly connects the convolution layer C2 and the final output layer. We have set up 50 neurons in this layer. Because of the full connection, each neuron is connected with all the feature values of the previous layer.
- (5) Output layer: the function of this layer is to output the predicted MI category. Since this study only involves two types of motor imagination, it only contains two neurons, and each neuron is connected with all neurons in the previous layer.

The training of CNN model is realized by back propagation algorithm. Given the input data, all neurons in the network structure produce the activation values according to the initial weight values, bias values, and activation function. According to the output of the softmax layer, the loss function value of the model is calculated. Then, the gradients of the weight and bias terms are computed based on the value of loss function. Afterward, the parameters of the network are updated according to the gradient values. In this study, we initialize the weight value of the network to a random value in the normal distribution with the mean value of 0 and the standard deviation of 0.1 and uniformly initialize the bias value to 0.1. All neurons in front of the output layer use the Rectified Linear Unit (ReLU) as the activation function, and its operation is as follows:

$$y = \text{ReLU}(\mathbf{w}^T \mathbf{x} + b) = \max(0, \mathbf{w}^T \mathbf{x} + b), \quad (3)$$

where  $\mathbf{x}$  is the feature vector,  $\mathbf{w}$  is the weight value vector,  $b$  is the bias value, and  $y$  is the output activation value. Many studies have shown that the traditional sigmoid function has the problem of gradient vanishing. However, using the ReLU activation function, which is similar to the neuron response mechanism in the biological neural system, we can usually achieve satisfactory training effect [32]. The output layer adopts the softmax model, and the specific operation of the model is as follows:

$$z_i = \mathbf{w}^i \mathbf{x} + b^i, \quad (i = 1, 2),$$

$$y_i = \frac{e^{z_i}}{\sum_{j=1}^2 e^{z_j}}. \quad (4)$$

Each output value of softmax model represents the probability that the sample belongs to a certain category, and the category with the maximum value is the category of the final output. The loss function uses cross entropy (CE):

$$H(p, q) = - \sum_x p(x) \ln(q(x)), \quad (5)$$

where  $p$  and  $q$  are the probability distributions of the predicted and original categories, respectively.

Gradient descent method is used to update weight and bias terms, and the optimizer adopts adaptive Adam algorithm. Note that, in order to prevent overfitting and achieve better results in the test set, the dropout [33] operation is performed before the output layer in this study. When training the neural network, some neurons are discarded according to a certain probability  $P(d)$ , and the model is trained according to the sparse network structure. In the testing stage, all the neurons are used, but all the weights are corrected to  $1-P(d)$  times of the original weights. In this study, we set  $P(d)$  to 0.5. Table 1 provides an overview of the CNN model hyperparameters.

### 3. Experiments

As mentioned in Section 2, the function of the first convolution layer is mainly to filter the original input signal in the spatial domain. According to the theory of ERD, the brain area most related to motor imagery is usually located on the opposite side of the motor imagery limb. In order to intuitively understand the characteristics of the spatial filter learned by CNN, for the Finger Dataset, we extract a filter from each subject's six filters for brain topographic map illustration. Figure 4 shows the brain pattern distribution of each spatial filter for five subjects in this dataset. It can be seen from the figure that the electrode with the largest weight is usually distributed near the "C3" channel; in another word, the EEG signal in this area has the greatest impact on pattern recognition. This area is just on the opposite side of the right index finger and is in the primary motor cortex. Therefore, the first convolution layer of CNN can learn the spatial characteristics of motor imagination EEG well. It should be noted that the learning of the above parameters is completely without manual operation, which is the result of network parameter updating and iteration, so CNN shows strong adaptive learning ability.

In this study, we train CNN model according to training set and validation set. After training, the models of all the subjects can converge effectively. Take subject S1 in the Finger Dataset as an example, the change curve of classification accuracy is shown in Figure 5, the blue solid line shows the classification accuracy of the training set, the red dotted line shows the classification accuracy of the validation set, the abscissa shows the number of iterations, and the ordinate shows the classification accuracy. It can be seen from Figure 5 that, after 1600 iterations, the classification accuracy of training set reaches the highest value and then remains stable, while the classification accuracy of validation set has been kept at the highest value, so it can be considered that the model achieves the best training effect after 1600 iterations, and the trained model is considered as the optimal classification model of subject S1.

In this study, we evaluate the classification performance of CNN model according to the classification accuracy of test

TABLE 1: Model hyperparameter.

Parameter	Value
Padding	Valid
Optimizer	Adam
Activation function	Relu
Regularization	Dropout
Cost function	Cross_entropy
Batch size	Size of training set

set. The software and hardware platforms of the proposed CNN model are Intel (R) Core (TM) i5-8500 3.00 GHz CPU, 8.0GB RAM, Spyder, Python 3.6, and TensorFlow 2.0 (CPU). With two EEG datasets, extensive experimental comparisons are carried out between our method and other closely related approaches.

For BCI competition III dataset IVa, we firstly used the fixed sample set segmentation for classification performance evaluation, namely, 260 samples for training, 10 samples for testing, and another 10 samples for validation. To evaluate the classification performance of our method objectively, three other algorithms including CSP [9], FBCSP [12], and SFBCSP [13] were adopted for performance comparison. Note that CSP is the baseline method for MI EEG pattern recognition; FBCSP and SFBCSP are state-of-the-art methods for MI EEG spatial-frequency feature learning, which are closely related to our method. The experimental settings for these competing methods are listed as follows.

- (i) CSP: EEG signals on 49 channels were extracted as suggested in [18]. EEG data between 0.5 and 2.5 s after the visual cue were used for feature extraction. A bandpass filter with passband of 8 to 16 Hz has been applied to capture the  $\mu$  rhythm [17]. The number of CSP filters was set to 2 [13]. Three widely used machine learning algorithms, including LDA, SVM, and LR, were used for classification.
- (ii) FBCSP: EEG signals on the same 49 channels as in CSP were extracted. EEG data between 0.5 and 2.5 s after the visual cue were used. 6 bandpass filters having bandwidth of 6 Hz in the range of 4 to 40 Hz with no overlap have been used, as described in [34]; these settings gave optimal results. Mutual information-based feature selection has been performed as it gave the best results in [35]; SVM was used for classification.
- (iii) SFBCSP: EEG signals on the same 49 channels as in CSP were extracted. EEG data between 0.5 and 2.5 s after the visual cue were used. 17 bandpass filters having bandwidth of 4 Hz with an overlap of 2 Hz in the range of 4 to 40 Hz have been used, as in [13]. The regularization parameter  $\lambda$  was determined by 10-fold cross-validation, and linear kernel SVM was adopted for classification as in [13].

Table 2 lists the classification accuracies of CSP + LDA, CSP + LDA, CSP + LR, FBCSP, SFBCSP, and our method on BCI competition III dataset IVa. From this table, we can observe that our method achieves the same results or

performs better than all competitors for subjects aa, al, and av and gives the highest average accuracy of 90%.

For the public BCI competition III dataset IVa, to more reliably evaluate the classification performance of our method, we further implemented 10-fold cross-validation to compute the average classification accuracy and also compared it with CSP, FBCSP, and SFBCSP. The settings of these three competing methods remained unchanged; however for CSP only SVM was adopted since it gave the best results in Table 2. We also adjusted our CNN model; the changes are mainly in three aspects: (1) 6 bandpass filters having bandwidth of 6 Hz in the range of 4 to 40 Hz with no overlap were used; thus the size of the input matrix was  $49 \times 6$ , where 49 was the number of electrodes and 6 was the number of frequency subbands. (2) Two more fully connected layers were added, and the activation functions were changed. (3) The optimizer was changed to Adadelta. Table 3 summaries the CNN model architecture for 10-fold cross-validation; the adjusted CNN model contains 2 convolution layers, a flatten layer, 3 fully connected layers, and an output layer. Elu and Relu are chosen as activation function for convolution layers and fully connected layers, respectively. Figure 6 presents classification accuracies derived by CSP, FBCSP, SFBCSP, and our method. The two modified CSP algorithms and our method outperformed the baseline CSP algorithm. Our proposed CNN model further yielded higher average classification accuracy than those of the FBCSP and SFBCSP methods. The average classification accuracy improvements achieved by our method were 3.66%, 1.44%, and 1.59% in comparison with CSP, FBCSP, and SFBCSP methods, respectively.

For the Finger Dataset, CSP and FBCSP were adopted for performance comparison. Sparse learning methods usually need more training samples to ensure the performance; thus, we did not use SFBCSP method for this small dataset. Note that, for this dataset, LDA classification algorithm was adopted since it gave superior performance compared to LR and SVM. Figure 7 presents classification accuracies derived by CSP, FBCSP, and our method for this dataset. Our method yielded higher classification accuracies for most subjects than those of the other two competing methods. The average classification accuracy improvements achieved by our method were 4.84% and 10.63% in comparison with CSP and FBCSP methods, respectively.

To further understand the computational cost of the proposed CNN model, for BCI competition III dataset IVa with fixed sample set segmentation, we recorded the training time of each subject (1600 iterations), and Figure 8 summaries the time durations of CNN training for all subjects in this dataset. From this figure we observe that all CNN models can be trained within about 5 seconds, which is an acceptable time cost for real application.

#### 4. Discussion

As we can see from Table 2, Figure 6, and Figure 7, overall our algorithm yielded superior classification performance than other competing algorithms. However, we notice that for some subjects, especially the subjects aw

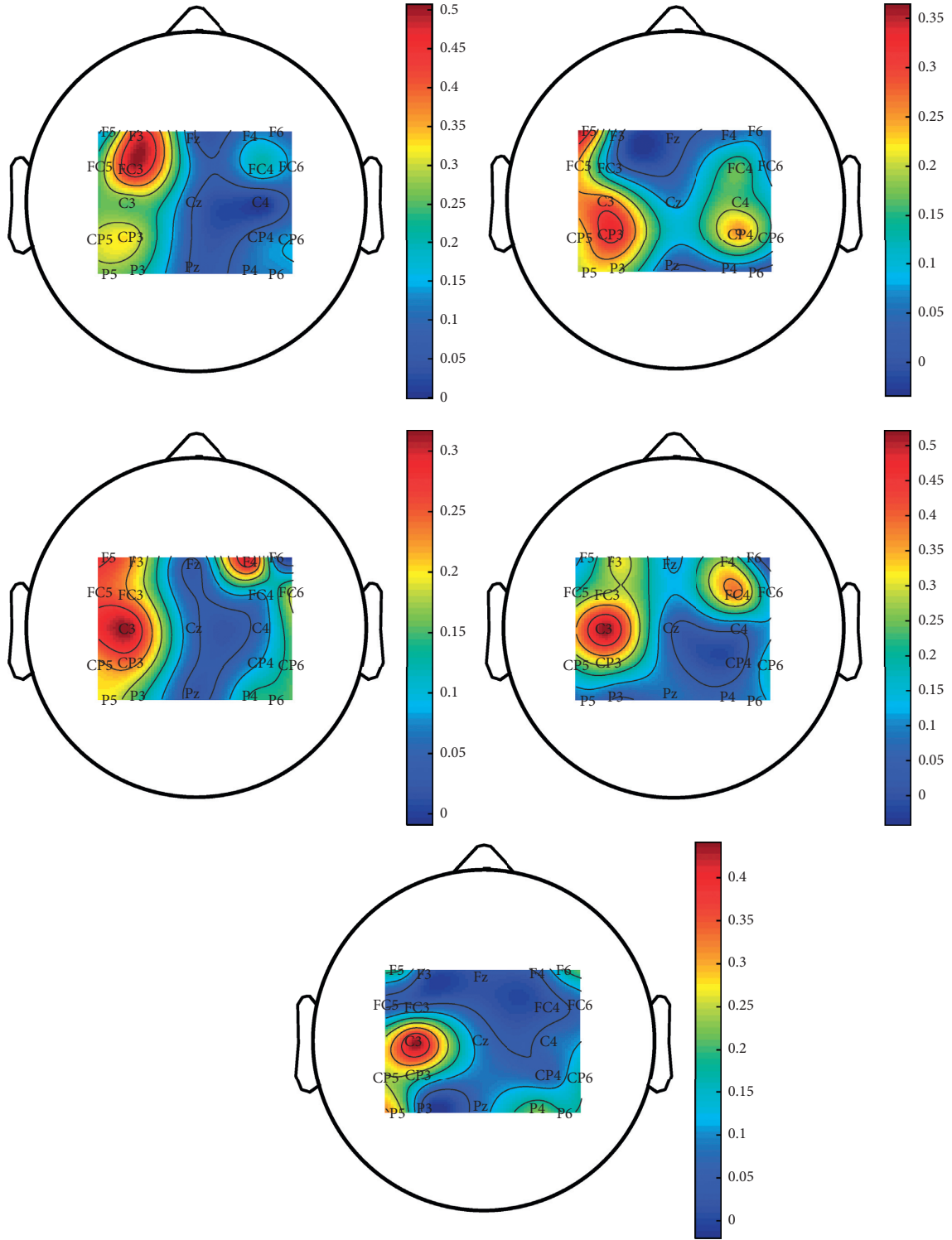


FIGURE 4: Spatial filter brain pattern distribution (Finger Dataset).

and  $a_y$  in Table 2, the performance of our proposed CNN model is even worse than that of the baseline CSP method. The reasons can be mainly concluded as follows. (1) For all the subjects in a certain dataset, the structures and hyperparameters of the CNN models are the same. However, different users have different characteristics. Therefore, for some subjects the uniform parameter

setting may lead to suboptimal solution. (2) The performance of deep learning models highly depends on the size of training data [21]. However, in BCI competition III dataset IVa with fixed sample set segmentation, 260 samples were used for CNN model training; the size of the training set was relatively small, which may affect the stability of the model to some extent. This can be further

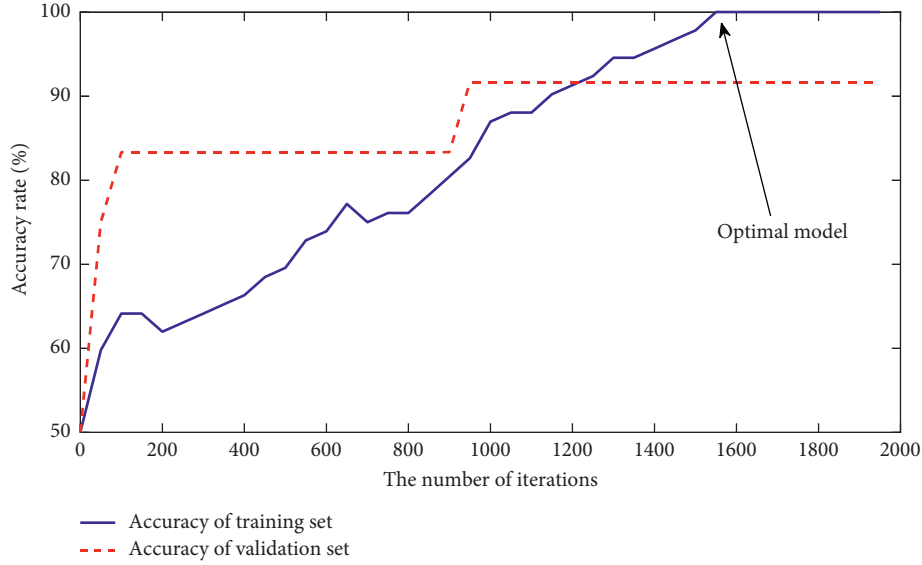


FIGURE 5: Classification accuracy of validation set and training set during CNN training (subject S1 of Finger Dataset).

TABLE 2: Comparison of classification accuracy for our method and 5 other competing methods; the highest accuracy is marked in boldface (BCI competition III dataset IVa).

Subject	aa (%)	al (%)	av (%)	aw (%)	ay (%)	Mean (%)
CSP [9] + LR	60	<b>90</b>	50	<b>100</b>	<b>100</b>	80
CSP [9] + SVM	60	<b>90</b>	60	<b>100</b>	<b>100</b>	82
CSP [9] + LDA	60	<b>90</b>	50	<b>100</b>	<b>100</b>	80
FBCSP [12]	60	<b>90</b>	60	90	<b>100</b>	80
SFBCSP [13]	60	<b>90</b>	70	90	<b>100</b>	82
Our method	<b>100</b>	<b>90</b>	<b>90</b>	90	80	<b>90</b>

TABLE 3: CNN model architecture for 10-fold cross-validation (BCI competition III dataset IVa). Conv refers to convolution layer, Flatten refers to flatten layer, and FC refers to fully connected layer.

	Kernel size	Kernel number	Padding	Activation	Output shape
Conv_1	$49 \times 1$	6	Valid	Elu	$1 \times 6 \times 6$
Conv_2	$1 \times 3$	12	Valid	Elu	$1 \times 2 \times 12$
Flatten	—	—	—	—	24
FC_1	—	—	—	Relu	50
FC_2	—	—	—	Relu	100
FC_3	—	—	—	Relu	200
Softmax	—	—	—	—	2

verified by the results in Figure 6; the classification performances of our method steadily surpass CSP algorithm for subjects aw and ay when 10-fold cross-validations were adopted.

To address the aforementioned limitations and further improve the classification performance of our proposed CNN method, the following three aspects are worthy of our future investigations. (1) Instead of using uniform CNN parameter setting, we plan to study the characteristics of different subjects and apply user specific CNN structure and parameters for MI EEG pattern recognition. (2) To cope with the problem of small training set and improve the stability of deep CNN model, the

combination of deep learning and subject to subject transfer learning is an important research direction in our plan. (3) In this study, the spatial-frequency features of MI EEG were automatically learned by the proposed deep CNN model; however 0.5 seconds to 2.5 seconds after the appearance of visual cue were manually selected. In another word, the features on temporal domain have not been fully studied. As pointed out in [36], the use of a fixed time window could hardly capture discriminative features for all subjects. Based on this, we consider further extending the proposed deep CNN model for MI EEG feature learning on the entire spatial-temporal-frequency domains.



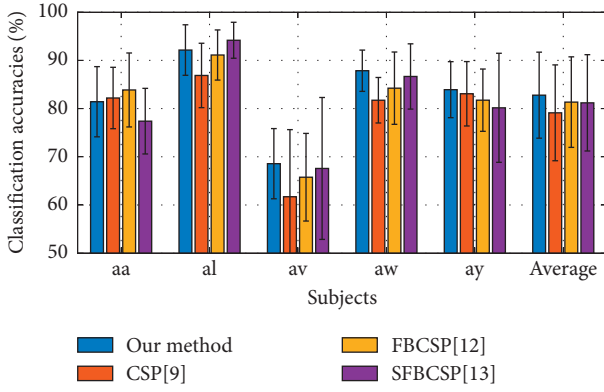


FIGURE 6: Classification accuracies of 10-fold cross-validations performed by our method and three other competing methods (BCI competition III dataset IVa).

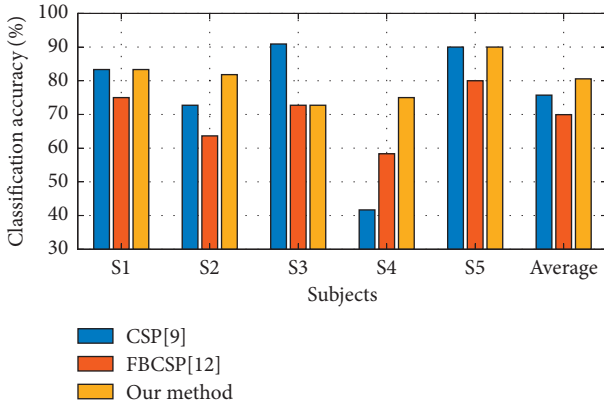


FIGURE 7: Classification accuracies derived by CSP, FBCSP, and our method (Finger Dataset).

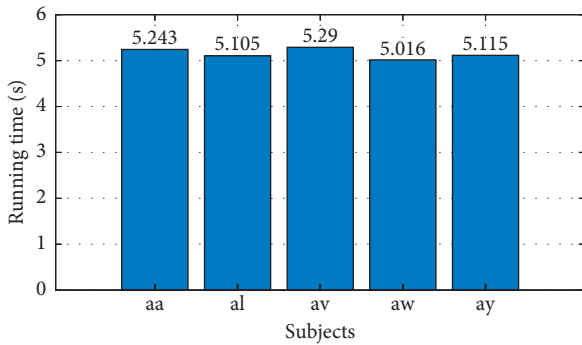


FIGURE 8: Running times of CNN training for all subjects in BCI competition III dataset IVa.

## 5. Conclusion

In this paper, a deep learning algorithm for limb motor imagery EEG pattern classification is proposed. A multilayer CNN model is designed for motor imagery EEG classification, and the spatial-frequency characteristics of motor imagery EEG signals are analyzed according to the obtained parameters of convolution layers in the neural network.

Finally, the proposed CNN model is compared with several state-of-the-art machine learning algorithms. In the experiment, the public BCI competition III dataset IVa and a self-collected right index finger movement imagination EEG dataset are used to verify the proposed algorithm. The experimental results demonstrate that the proposed CNN method outperforms all competitors in terms of the mean classification precision for both datasets. In addition, the training time of the proposed model is relatively short. Therefore, we would like to note that the proposed classification method is of great interest for real-life BCI systems.

## Data Availability

In this study, we used two different datasets for analysis. The first dataset is public BCI competition III dataset IVa and the second dataset is right index finger motion imagination dataset (denoted by Finger Dataset) which was collected by us. For BCI competition III dataset IVa: the BCI competition III dataset IVa used to support the findings of this study has been deposited in the website <http://www.bbci.de/competition/iii/>. For Finger Dataset: the Finger Dataset used to support the findings of this study is available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Supplementary Materials

The supplementary files S1, S2, S3, S4 and S5 corresponds to the five subjects of Finger Dataset, here DynamicSelect corresponds to finger movement imagination, and StaticSelect corresponds to the rest state and all these mat files are the collected MI EEG datasets. (*Supplementary Materials*)

## Acknowledgments

This work was partially supported by grants from the National Natural Science Foundation of China (nos. 61772198 and 61772199) and Zhejiang Province Basic Public Welfare Research Project of China (no. LGN18F020002).

## References

- [1] M. X. Dai, D. Z. Zheng, S. C. Liu, and P. J. Zhang, "Transfer kernel common spatial patterns for motor imagery brain-computer interface classification," *Computational and Mathematical Methods in Medicine*, vol. 2018, Article ID 9871603, 9 pages, 2018.
- [2] H. L. Chang and J. M. Yang, "Genetic-based feature selection for efficient motion imaging of a brain-computer interface framework," *Journal of Neural Engineering*, vol. 15, no. 5, 2018.
- [3] S.-B. Lee, H.-J. Kim, H. Kim, J.-H. Jeong, S.-W. Lee, and D.-J. Kim, "Comparative analysis of features extracted from EEG spatial, spectral and temporal domains for binary and multiclass motor imagery classification," *Information Sciences*, vol. 502, pp. 190–200, 2019.



- [4] Y. Li, X. R. Zhang, B. Zhang, M. Y. Lei, W. G. Cui, and Y. Z. Guo, "A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 6, pp. 1170–1180, 2019.
- [5] S. A. Shah, H. L. Tan, G. Tinkhauser, and P. Brown, "Towards real-time, continuous decoding of gripping force from deep brain local Field potentials," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 7, pp. 1460–1468, 2018.
- [6] M. Angrick, C. Herff, E. Mugler et al., "Speech synthesis from ECoG using densely connected 3D convolutional neural networks," *Journal of Neural Engineering*, vol. 16, no. 3, Article ID 036019, 2019.
- [7] K. M. Sonkin, L. A. Stankevich, J. G. Khomenko, Z. V. Nagornova, and N. V. Shemyakina, "Development of electroencephalographic pattern classifiers for real and imaginary thumb and index finger movements of one hand," *Artificial Intelligence in Medicine*, vol. 63, no. 2, pp. 107–117, 2015.
- [8] S. L. Norman, M. Dennison, E. Wolbrecht, S. C. Cramer, R. Srinivasan, and D. J. Reinkensmeyer, "Movement anticipation and EEG: implications for BCI-contingent robot therapy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 8, pp. 911–919, 2016.
- [9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.
- [10] R. Zhang, P. Xu, T. J. Liu et al., "Local temporal correlation common spatial patterns for single trial EEG classification during motor imagery," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 591216, pp. 1–7, 2013.
- [11] X. Song and S.-C. Yoon, "Improving brain-computer interface classification using adaptive common spatial patterns," *Computers in Biology and Medicine*, vol. 61, pp. 150–160, 2015.
- [12] K. K. Ang, Z. Y. Chin, C. C. Wang, C. T. Guan, and H. H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV Datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, 2012.
- [13] Y. Zhang, G. X. Zhou, J. Jin, X. Y. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface," *Journal of Neuroscience Methods*, vol. 255, pp. 85–91, 2015.
- [14] M. H. Bhatti, J. Khan, M. U. G. Khan et al., "Soft computing-based EEG classification by optimal feature selection and neural networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5747–5754, 2019.
- [15] A. M. Azab, L. Mihaylova, K. K. Ang, and M. Arvaneh, "Weighted transfer learning for improving motor imagery-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.
- [16] Y. Jiao, Y. Zhang, X. Chen et al., "Sparse group representation model for motor imagery EEG classification," *Ieee Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 631–641, 2019.
- [17] Z. Jin, G. Zhou, D. Gao, and Y. Zhang, "EEG classification using sparse Bayesian extreme learning machine for brain-computer interface," *Neural Computing and Application*, vol. 32, no. 11, pp. 6601–6609, 2020.
- [18] H. Zeng and A. G. Song, "Optimizing single-trial EEG classification by stationary matrix logistic regression in brain-computer interface," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2301–2313, 2016.
- [19] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [20] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE-ACM Transactions on Audio Speech and Language Processing*, vol. 22, no. 10, pp. 1533–1545, 2014.
- [21] X. Zhang, L. Yao, X. Wang, J. Monaghan, and D. McAlpine, "A survey on deep learning based brain computer interface," *Recent Advances and New Frontiers*, vol. 1, no. 1, p. 66, 2018.
- [22] H. K. Lee and Y. S. Choi, "A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequency image," in *Proceedings of 32nd International Conference on Information Networking*, Cracow, Poland, May.2018.
- [23] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *Journal of Neural Engineering*, vol. 14, no. 1, 2017.
- [24] T. Uktveris and V. Jusas, "Application of convolutional neural networks to four-class motor imagery classification problem," *Information Technology and Control*, vol. 46, no. 2, pp. 260–273, 2017.
- [25] J. W. Liu, Y. Cheng, and W. D. Zhang, "Deep learning EEG response representation for brain computer interface," in *34th Chinese Control Conference, Chinese Control Conference Q*, Zhao and S. Liu, Eds., pp. 3518–3523, 2015.
- [26] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "Hierarchical internal representation of spectral features in deep convolutional networks trained for EEG decoding," in *Proceedings of 6th International Conference on Brain-Computer Interface*, Gangwon, Korea, January 2018.
- [27] Q. Wang, Y. J. Hu, and H. Chen, "Multi-channel EEG classification based on Fast convolutional feature extraction," in *Advances in Neural Networks*, F. Cong, A. Leung, and Q. Wei, Eds., vol. 10262, , pp. 533–540, Lecture Notes in Computer Science, 2017.
- [28] C. Tan, F. Sun, W. Zhang, J. Chen, and C. Liu, "Multimodal classification with deep convolutional-recurrent neural networks for electroencephalography," in *Neural Information Processing*, Springer International Publishing, New York, NY, USA, 2017.
- [29] H. J. Yang, S. Sakhavi, K. K. Ang, and C. T. Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," in *Proceedings of 37th Annual International Conference of the Ieee Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society Conference Proceedings*, Milano, Italy, pp. 2620–2623, August 2015.
- [30] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11–18, 2017.
- [31] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.
- [32] H. Z. Zhao, F. X. Liu, L. Y. Li, and C. Luo, "A novel softplus linear unit for deep convolutional neural networks," *Applied Intelligence*, vol. 48, no. 7, pp. 1707–1720, 2018.
- [33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural

- networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [34] S. Kumar and A. Sharma, “A new parameter tuning approach for enhanced motor imagery EEG signal classification,” *Medical & Biological Engineering & Computing*, vol. 56, no. 10, pp. 1861–1874, 2018.
- [35] K. K. Ang, Z. Y. Chin, H. H. Zhang, and C. T. Guan, “Filter bank common spatial pattern (FBCSP) in brain-computer interface,” in *Proceedings of IEEE International Joint Conference on Neural Networks*, vol. 1-8, IEEE International Joint Conference on Neural Networks (IJCNN), Hong Kong, China, pp. 2390–2397, June 2008.
- [36] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, “Temporally constrained sparse group spatial patterns for motor imagery BCI,” *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3322–3332, Sep, 2019.

## Research Article

# An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm

Wentao Wu<sup>1</sup>,<sup>✉</sup> Daning Li,<sup>2</sup> Jiaoyang Du,<sup>1</sup> Xiangyu Gao,<sup>2</sup> Wen Gu,<sup>3</sup> Fanfan Zhao,<sup>2</sup> Xiaojie Feng,<sup>2</sup> and Hong Yan<sup>1</sup>

<sup>1</sup>Department of Epidemiology and Health Statistics School of Public Health, Xi'an Jiaotong University, Xi'an, China

<sup>2</sup>School of Public Health, Xi'an Jiaotong University, Xi'an, China

<sup>3</sup>The First Affiliated Hospital, Xi'an Jiaotong University Health Science Center, Xi'an, Shaanxi 710061, China

Correspondence should be addressed to Wentao Wu; [qrr22222@stu.xjtu.edu.cn](mailto:qrr22222@stu.xjtu.edu.cn)

Received 1 June 2020; Accepted 1 July 2020; Published 14 July 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Wentao Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Among the currently proposed brain segmentation methods, brain tumor segmentation methods based on traditional image processing and machine learning are not ideal enough. Therefore, deep learning-based brain segmentation methods are widely used. In the brain tumor segmentation method based on deep learning, the convolutional network model has a good brain segmentation effect. The deep convolutional network model has the problems of a large number of parameters and large loss of information in the encoding and decoding process. This paper proposes a deep convolutional neural network fusion support vector machine algorithm (DCNN-F-SVM). The proposed brain tumor segmentation model is mainly divided into three stages. In the first stage, a deep convolutional neural network is trained to learn the mapping from image space to tumor marker space. In the second stage, the predicted labels obtained from the deep convolutional neural network training are input into the integrated support vector machine classifier together with the test images. In the third stage, a deep convolutional neural network and an integrated support vector machine are connected in series to train a deep classifier. Run each model on the BraTS dataset and the self-made dataset to segment brain tumors. The segmentation results show that the performance of the proposed model is significantly better than the deep convolutional neural network and the integrated SVM classifier.

## 1. Introduction

The incidence of brain tumors increases with age [1]. This article focuses on gliomas in brain tumors. According to the location of the glioma, the cell type, and the severity of the tumor, the World Health Organization classifies the glioma into I~IV grades. Among them, Classes I and II are low-grade gliomas, and Classes III and IV are high-grade gliomas [2]. In order to facilitate doctors to accurately remove gliomas during surgery, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Computed Tomography (PET) and other imaging techniques are commonly used in clinical treatment to brain image segmentation of the glioma area which helps the doctor to safely remove the tumor within the maximum range.

At the same time, MRI has the characteristics of significant soft tissue contrast and can provide abundant physiological tissue information. In the clinical treatment of gliomas, MRI is usually used to diagnose gliomas preoperatively, intraoperatively, and postoperatively.

Glioma is a tumor composed of a necrotic core, a margin of tumor activity, and edema tissue. Multiple MRI sequences can be used to image different tumor tissues [3], as shown in Figure 1. At present, MRI imaging of gliomas generally has four modal sequences: T1-weighted, post-contrast T1-weighted, T2-weighted, and FLAIR. Different sequences reflect different glioma tissues [4]. The general FLAIR sequence is suitable for observing edema tissues, and the T1ce sequence is suitable for observing the active components of the tumor core.

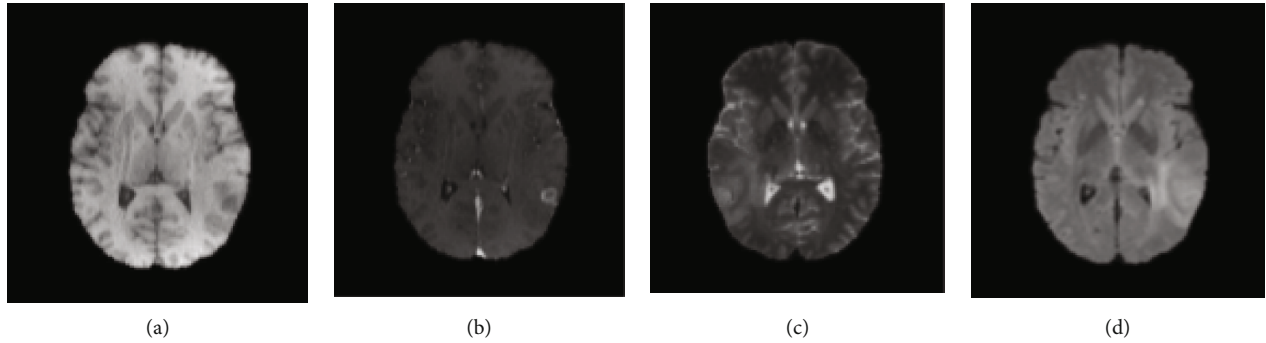


FIGURE 1: MRI of glioma: (a) T1-weighted, (b) postcontrast T1-weighted, (c) T2-weighted, and (d) FLAIR.

MRI-based segmentation of gliomas and their surrounding abnormal tissues facilitates the doctor to observe the external morphology of each tumor tissue of the patient's glioma and also facilitates the doctor's imaging-based analysis and further treatment. Therefore, the segmentation of glioma is considered to be a first step in the MRI analysis of glioma patients. Because gliomas have different degrees of deterioration and contain multiple tumor tissue regions and brain MRI is a multimodal and many-layer three-dimensional scan image, manual segmentation of glioma regions requires a lot of time and manpower. In addition, manual segmentation is often based on the brightness of the image observed by the human eye for area segmentation, which is easily affected by the quality of the image generation and the personal factors of the tagger. It is prone to erroneous segmentation and segmentation of redundant areas. Therefore, in clinical practice, a fully automatic segmentation method with good segmentation accuracy for gliomas is needed. However, the problems in the study of automatic glioma segmentation methods are summarized as follows: (1) glioma is often distinguished in the image by the change in pixel intensity between the lesion area and surrounding normal tissues. Due to the presence of a gray-scale offset field, the intensity gradient between adjacent tumor tissues will be smoothed, resulting in blurred tumor tissue boundaries. (2) The structure of gliomas differs in size, shape, and position, making segmentation algorithms difficult to model. And because the growth position of glioma is not fixed, it is often accompanied by a tumor mass effect. This will cause the surrounding normal brain tissue to be compressed and change its shape, thereby generating irregular background information and increasing the difficulty of segmentation.

At present, computer-aided diagnosis technology based on machine learning has been widely used in medical image analysis in recent years [5–14]. Since the algorithm based on machine learning can train model parameters through various features of medical images and use the trained model to predict the extracted features, it can well solve the classification, regression, and aggregation in medical images. At the same time, the deep learning technology in machine learning can directly obtain high-dimensional features directly from the data and automatically adjust the model parameters through forward propagation and back-regulation algorithms, so that the performance of the model in related tasks

can be optimized. Therefore, medical data processing of deep learning technology has developed into a research hotspot.

Brain tumor segmentation methods can be roughly divided into three categories: based on traditional image algorithms [15–20], based on machine learning [21–24], and based on deep learning [25–30]. In recent years, deep learning has become the method of choice for complex tasks due to its high accuracy. The convolutional neural network (CNN) proposed in [25] has made tremendous progress in the field of image processing. Therefore, the segmentation method based on the convolutional neural network is widely used in segmentation of lung nodules, retinal segmentation, liver cancer segmentation, and glioma segmentation [26]. Many scholars have begun to apply CNN in deep learning to segmentation of gliomas. Reference [31] proposes a brain cancer segmentation method based on dual-path CNN. Reference [32] trained two CNNs to segment high-grade gliomas and low-grade gliomas. Reference [33] proposed a two-channel three-dimensional CNN for glioma segmentation.

This paper mainly studies the segmentation method of glioma based on the deep learning method, aiming at automatically and accurately segmenting the glioma region from the brain MRI through the deep learning algorithm. For the task of glioma segmentation, this paper proposes a DCNN-F-SVM deep classifier. The main research contents of this article are as follows:

- (1) A new depth classifier is proposed. The classifier is composed of a deep convolutional neural network and an integrated SVM algorithm. First, CNN was trained to learn the mapping from image space to tumor label space. The predicted labels in CNN together with the test images were input into an integrated SVM classifier. In order to make the results more accurate, we deepened the classification process and iterated these two steps again to form the framework of the next CNN-SVM in series
- (2) The traditional segmentation method is to use the training set to train a suitable classifier, and then test the set for verification. The method proposed in this study is completely different from the traditional method. The proposed model mainly includes three stages: one is preprocessing, feature extraction, and

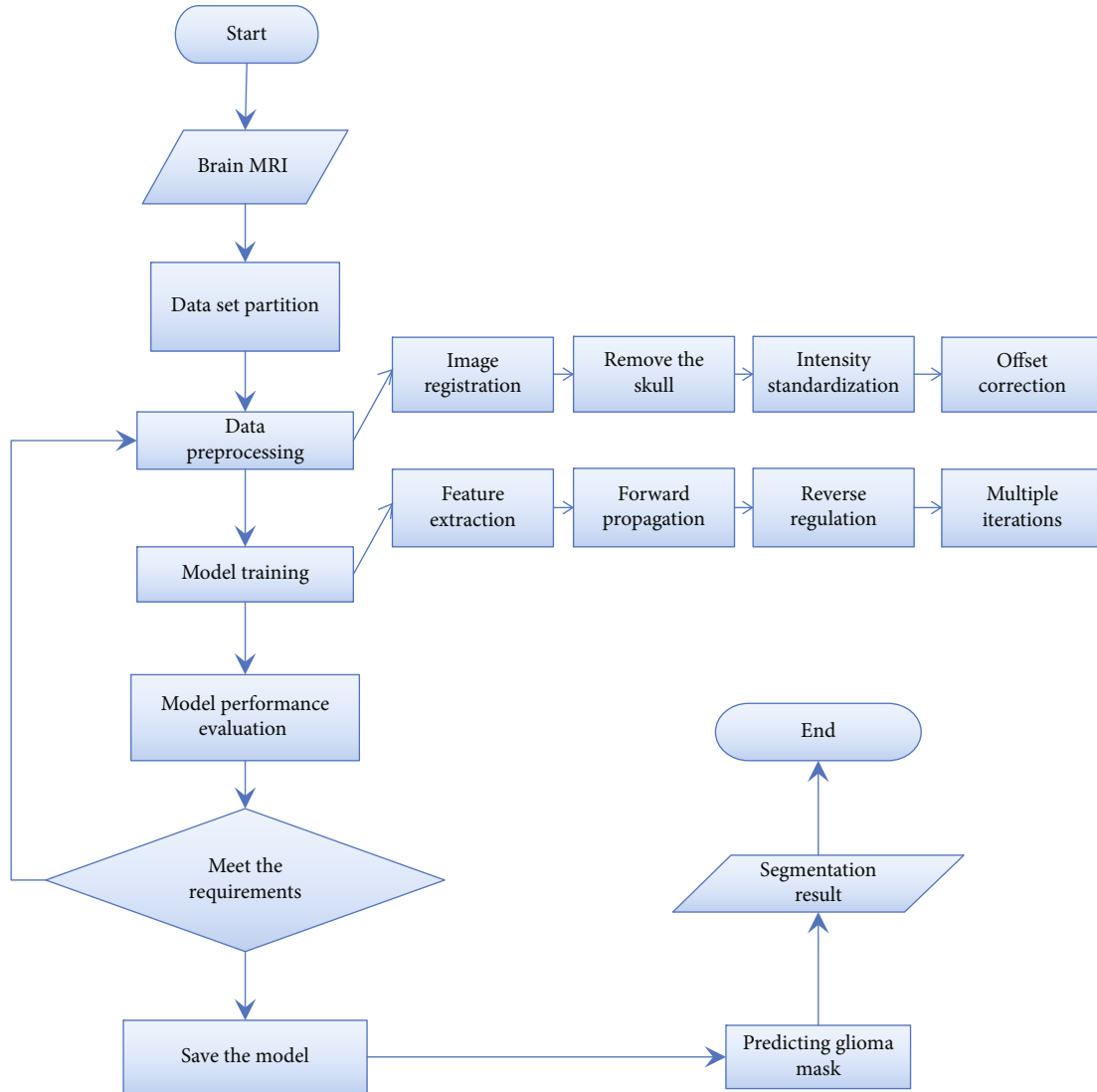


FIGURE 2: Flow chart of glioma segmentation algorithm based on deep learning.

training CNN and SVM. The second is to test and generate the final segmentation results. The third is to deepen the order of our CNN-SVM cascade classifier through an iterative step

- (3) Apply the proposed model to public datasets and self-made datasets for evaluation. Compared with the segmentation performance of CNN and SVM alone, the superiority of the proposed model can be reflected in various evaluation indexes

## 2. Related Works

**2.1. Process of Brain Tumor Segmentation Algorithm Based on Deep Learning.** In the currently proposed glioma segmentation method, the segmentation results of traditional image processing algorithms rely heavily on manual intervention, and a priori constraints are required to ensure the segmentation effect, resulting in poor robustness and low efficiency of the method. The glioma segmentation method based on

machine learning needs to manually select the features of the image, so that the segmentation effect of this type of method depends on the artificial features, and the generalization ability of the segmentation algorithm itself is weak.

The glioma segmentation method based on deep learning can automatically extract image features through the neural network model and segment the glioma region. Therefore, the shortcomings of strong prior constraints and manual intervention in the above method are overcome. The automation and robustness of the segmentation algorithm are improved, and good segmentation results can be achieved in large-scale complex glioma segmentation scenarios. Figure 2 is the flow of glioma segmentation algorithm based on deep learning. The process can be described as follows: first, obtain the MRI of the patient's brain and use it as the input data of the algorithm; then, divide the input data into the training set, verify the set, and test the set. At the same time, due to factors such as noise and uneven intensity in the original brain MRI, the divided data needs to be preprocessed. Commonly used glioma image preprocessing methods include



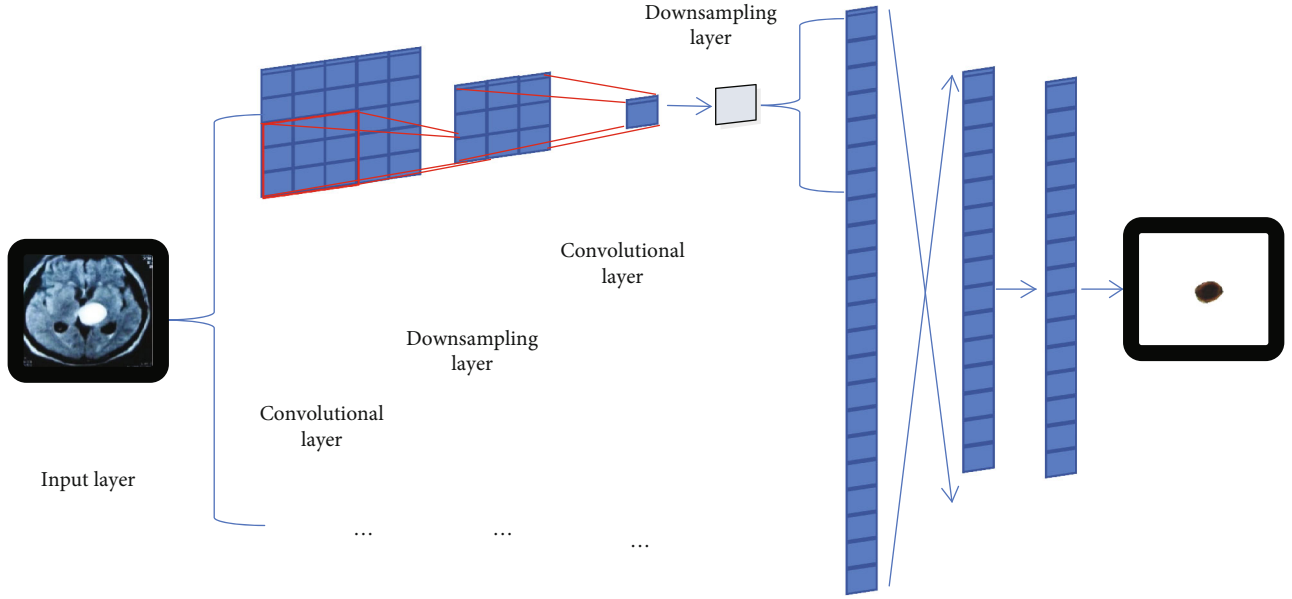


FIGURE 3: LeNet convolutional neural network structure.

image registration, skull removal, intensity standardization, and offset correction. Next, use the preprocessed input data to train the deep learning model. During the training process, the deep model will automatically perform feature extraction, and add the extracted features to the designed model structure for forward propagation. At the same time, the multiregion mask of glioma is used as a label to calculate the loss value, so that the model parameters are reversely adjusted in multiple iterations to achieve the purpose of optimal model performance. Then, at the end of each iteration, different evaluation indicators are used to evaluate the performance of the model, and the models that meet the conditions of the indicators are saved. Finally, the highly evaluated model is used to segment the test set data to obtain the final glioma segmentation results.

**2.2. A Deep Brain Tumor Feature Generation Method.** CNNs are well-known practical models in the field of deep learning, and their innovative ideas stem from the processing of human brain nerves. The perceptron model proposed in 1980 is considered to be the original model of convolutional neural networks. The perceptron model is a classic model in the field of machine learning, but this model also has great shortcomings and cannot solve XOR problems well. On this basis, reference [34] proposed the LeNet model, which has multiple convolutional layers, and each layer is a fully connected model trained using the back propagation algorithm [35]. Reference [36] proposed an artificial neural network called displacement invariance and studied the parallel structure of the convolutional neural network. However, these models are limited by experimental data and hardware conditions. Therefore, it is not suitable for complex tasks such as object detection and scene classification. In order to solve some problems in the training process of convolutional neural networks, Krizhevsky et al. proposed the AlexNet model [37]. In order to solve the overfitting problem of convolu-

tional neural networks, the model proposes local convolution and Relu technologies, and the overfitting problem is well solved.

CNN is essentially a multilayer perceptron and a multilayer neural network, and there is an obvious sequence between these layers, which is composed of an input layer, a hidden layer, and an output layer. There can be multiple hidden layers, and each layer is composed of multiple two-dimensional planes. Each plane contains multiple neurons, and the hidden layer consists of a convolution layer, a downsampling layer, and a fully connected layer. The convolution layer and the downsampling layer appear alternately and can have multiple layers, and the fully connected layer can also have multiple layers. The network structure of the traditional convolutional neural network LeNet is shown in Figure 3.

In the convolution layer, the feature maps output by the previous layer are convolved by the learned convolution kernel, and the corresponding partial derivatives are input into the activation function together to form an output feature maps. The downsampling layer is used for feature selection to select representative features. The fully connected layer is a neural network layer whose role is to map two-dimensional distributed features into feature vectors for better classification. The output layer is a simple classification layer, usually using logistic regression for classification. Here, we use the Softmax classifier for classification.

The activation function usually selects a nonlinear function to better fit the nonlinear model. Selecting the activation function needs to consider its monotonicity and derivability. Common activation functions are shown as follows:

- (1) Relu function:  $f(x) = \max(0, x)$
- (2) Softplus function:  $f(x) = \log(1 + e^x)$

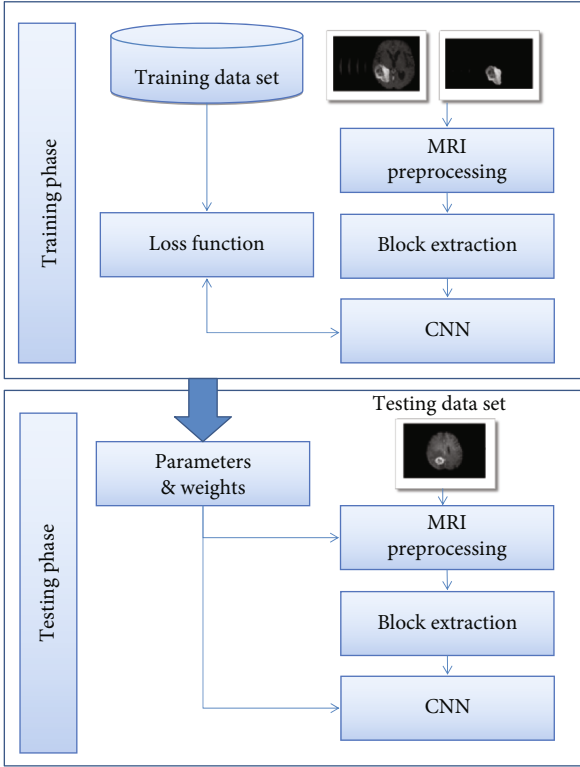


FIGURE 4: CNN flow chart.

The CNN model structure is simpler and easier to expand than the neurocognitive machine. In the neurocognitive machine, the downsampling layer and the convolutional layer alternate to form the function of feature extraction and abstraction, while in the convolutional neural network, the convolutional layer and the downsampling layer alternate, and their functions are similar. The convolution operation simplifies feature extraction, the excitation function replaces multiple nonlinear functions of the neurocognitive machine, and the pooling operation is also simpler. The CNN algorithm flow is shown in Figure 4.

**2.3. Introduction of Brain Tumor Dataset.** The BraTS Challenge held in 2012 provided a brain MRI dataset with both low-grade gliomas and high-grade gliomas. The dataset provides MRI of multiple patients and provides a multiregion glioma segmentation ground truth for each patient. Among them, ground truth is the result of fusion of 20 segmentation algorithms and then manually labeled by multiple human experts. Every BraTS competition will provide a public dataset of gliomas. However, the glioma dataset provided since BraTS17 has been significantly different from the dataset provided before 2016. The dataset used between BraTS14 and BraTS16 contains images of gliomas before and after surgery, which leads to confusing glioma segmentation criteria in the dataset and does not have the conditions to be true segmentation criteria. Therefore, the datasets between BraTS14 and BraTS16 are no longer used in the games after BraTS17. The BraTS18 dataset is based on the BraTS17 dataset with the addition of the TCIA glioma dataset. The TCIA glioma data-

TABLE 1: Introduction of BraTS dataset over the years.

Dataset	Date	Total number of samples			
		Training set	Validation set	Test set	Total
BraTS12	2012	30	10	25	65
BraTS13	2013	30	10	25	65
BraTS14	2014	40	10	25	65
BraTS15	2015	274	—	110	384
BraTS16	2016	274	—	191	465
BraTS17	2017	210	46	146	412
BraTS18	2018	285	67	191	543

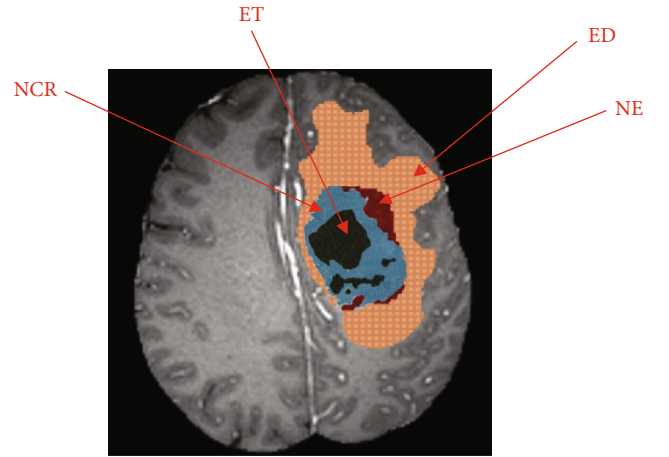


FIGURE 5: Tumor area division of glioma.

set includes 262 high-grade glioma patient images and 199 low-grade glioma patient images. This dataset contains the MRI and ground truth of 543 glioma patients and is currently the most standard glioma segmentation dataset. The details of the datasets in the BraTS competition datasets over the years are shown in Table 1.

As shown in Figure 5, gliomas are generally divided into four tumor regions, namely, edema around the tumor (ED), nonenhanced tumor core (NET), enhanced tumor core (ET), and necrotic core (NCR). Among them, ED, NET, and NCR are real glioma tumor tissues. The enhancement of the tumor core is to facilitate the observation of the tumor core.

**2.4. Evaluation Method of Segmentation Result.** The common evaluation methods for evaluating the performance of each model in the field of image segmentation are shown in Table 2.

In addition to the above evaluation indicators, there are indicators such as Hausdorff Li and positive predictive value. The most commonly used are DSC and sensitivity.

### 3. Introduction of DCNN-F-SVM Model

This study proposes a brain tumor segmentation model based on convolutional neural network fusion SVM. Figure 6 is the model flow chart.

TABLE 2: The description of the adopted indices.

Index	Expression/description
True Positive (TP)	TP indicates that the model predicts a glioma region, and the doctor marks pixels that are also glioma regions
False Positive (FP)	FP means pixels predicted by the model as the glioma area are actually the background area
True Negative (TN)	TN indicates that the model predicted as the background area is actually the pixel of the background area
True Negative (TN)	FN means pixels predicted by the model as the background area are actually as the tumor area
Dice Similarity Coefficient (DSC)	$DSC = 2TP / (FP + 2TP + FN)$
Sensitivity	$Sens = TP / (TP + FN)$
Specificity	$Spec = TN / (TN + FP)$

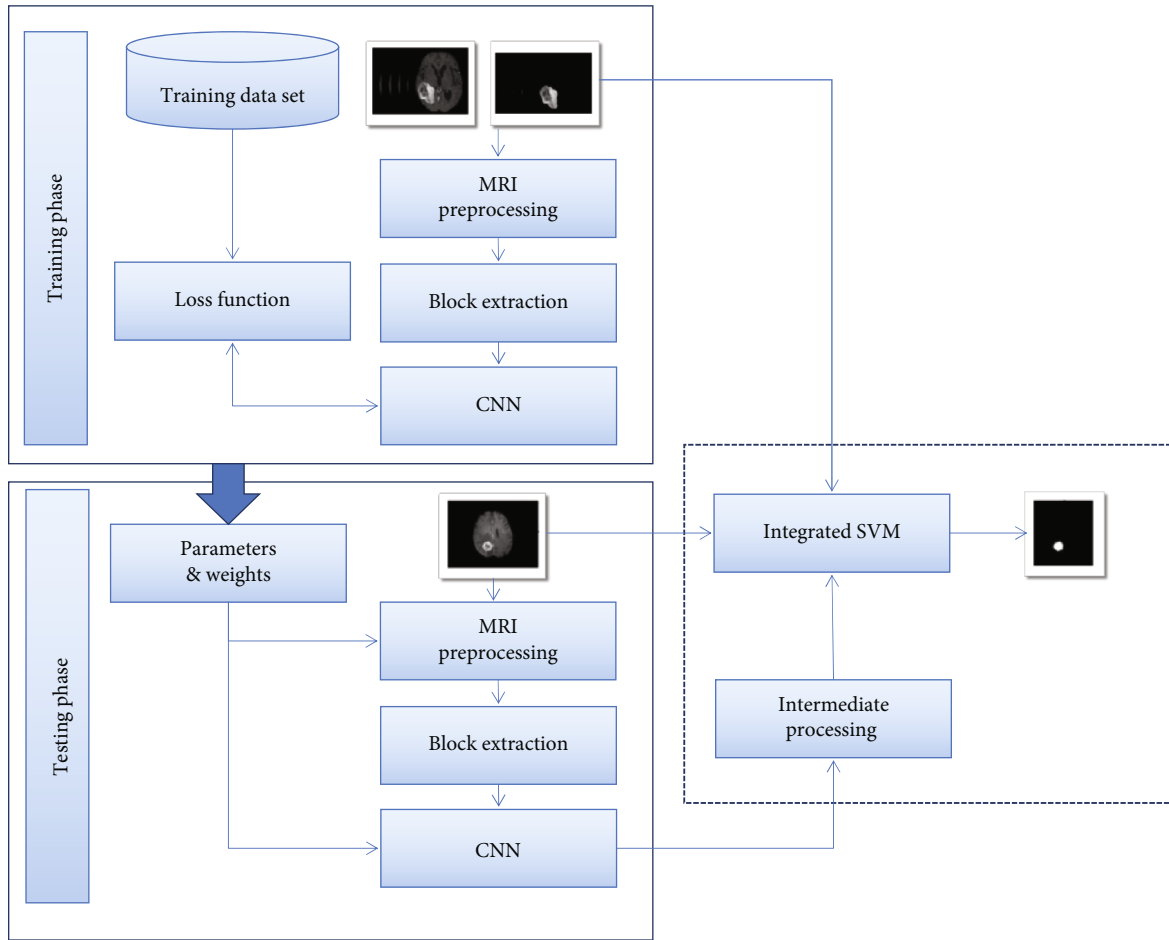


FIGURE 6: The proposed model flow chart.

The proposed model segmentation of brain tumor images can be divided into two parts: one is preprocessing, feature extraction, and training CNN and SVM; the other is testing and generating the final segmentation results. It can be divided into 3 stages. In the first stage, CNN and integrated SVM are trained to obtain the mapping from the gray image domain to the tumor label domain. In the second stage, the labeled output of CNN and the test image are input into the integrated SVM classifier. In the third stage, an iterative step is used to connect the CNN and

the integrated SVM classifier, which increases the number of layers. In order to select the optimal feature, an intermediate processing step is added to the model, as shown in Figure 7.

Grayscale, mean, and median are used to represent each pixel. These features are used to train CNN to obtain a non-linear mapping between input features and labels. In the testing stage, an aggregated SVM classifier is independently trained using the aggregated CNN label map and the same features as before.

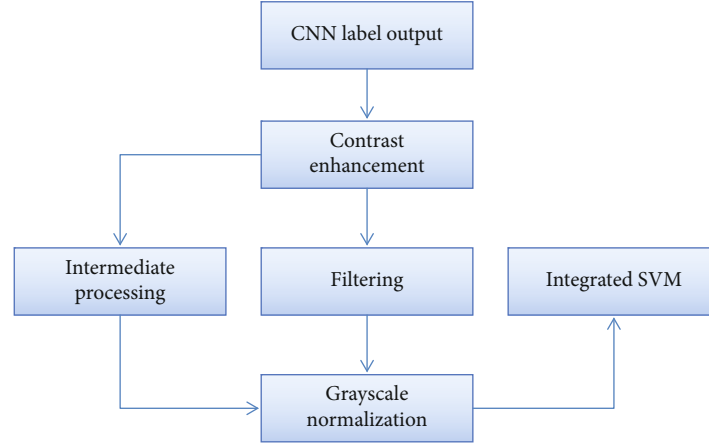


FIGURE 7: Schematic diagram of intermediate processing.

TABLE 3: Experimental environment description.

Hardware configuration		Software configuration	
Configuration item	Configuration parameter	Configuration item	Configuration parameter
Operating system	Ubuntu 14.04	Development environment	PyCharm
CPU	AMD A8-5600K	Programming language	Python
RAM	16.0GB	Image algorithm library	OpenCV
Video memory	479 MB	Deep learning algorithm library	TensorFlow

An iterative classification process is applied to the pre-processed input image. First, CNN classifies the pixels in the key area, thus generating a kind of presegmentation, which will be sent to the integrated SVM classifier. Then, a Region Of Interest (ROI) on presegmentation will be generated. In addition to presegmentation, classification based on integrated SVM will be performed on this ROI. After that, the integrated SVM explores the neighborhood of the CNN output. Use CNN to classify the marked ROI again. Repeat the above steps to further refine the segmentation results.

#### 4. Simulation Experiment

**4.1. Experiment-Related Instructions.** The experimental dataset used in this study includes the public dataset and the self-made dataset. The comparison models are SVM, CNN, and DCNN-F-SVM. In the setting of experimental parameters, set the window size to 5,  $\sigma=0.1$ , and  $C=1000$ . The public dataset used is the BraTS18 dataset. The self-made dataset is the clinical MRI images of 26 patients. The evaluation index used in the experiment is DSC, sensitivity, and specificity. The description of the experimental software and hardware environment is shown in Table 3.

**4.2. Public Dataset Experiment.** After the model training is completed, the test set can be predicted by the model to obtain the glioma segmentation result obtained by the model segmentation. In the test set divided by three-fold cross-validation, the evaluation index pair of each model on the BraTS18 dataset is shown in Table 4. The data in the table fully shows that the proposed model has better tumor

TABLE 4: Evaluation index of each model.

Model	DSC	Sensitivity	Specificity
SVM	0.8268	0.8306	0.9845
CNN	0.8556	0.8876	0.9962
DCNN-F-SVM	0.8958	0.9110	0.9982

segmentation performance than SVM and CNN. Compared with SVM, the proposed algorithm has improved by 8.3%, 9.7%, and 1.4% on the three indicators: DSC, sensitivity, and specificity; compared with CNN, the proposed algorithm has three indicators: DSC, sensitivity, and specificity, increased by 4.7%, 2.6% and 0.2%, respectively.

**4.3. Self-Made Data Experiment.** In this section, clinical MRI images of 26 patients were collected, and brain tumors were trained and segmented using three models, and the experimental results were given. Tables 5 and 6 show the segmentation results of CNN and DCNN-F-SVM for 26 patients, respectively.

Among the index values shown in Table 5, the DSC values are generally distributed around 0.86 and have an up and down floating error of about 0.18. The sensitivity values are generally distributed around 0.89 and have a floating error of about 0.14. The specificity values are generally distributed around 0.95 and have an up and down floating error of about 0.11.

Among the index values shown in Table 6, the DSC value is generally distributed around 0.89, and there is an upward and downward floating error of about 0.15. The sensitivity

TABLE 5: Evaluation data of 26 patients with brain tumor segmentation using the SVM model.

Number	DSC	Sensitivity	Specificity	Number	DSC	Sensitivity	Specificity
1	0.8801	0.9020	0.9563	14	0.8695	0.8896	0.9411
2	0.8768	0.8963	0.9368	15	0.8753	0.8976	0.9520
3	0.8893	0.9158	0.9605	16	0.8536	0.8729	0.9264
4	0.8682	0.8910	0.9482	17	0.8463	0.8667	0.9118
5	0.8926	0.9089	0.9795	18	0.8831	0.9053	0.9786
6	0.8796	0.8998	0.9385	19	0.8920	0.9107	0.9632
7	0.8859	0.9096	0.9543	20	0.8697	0.8896	0.9408
8	0.8633	0.8859	0.9386	21	0.8787	0.9006	0.9602
9	0.8828	0.9010	0.9715	22	0.8811	0.9120	0.9632
10	0.8989	0.9157	0.9634	23	0.8980	0.9234	0.9728
11	0.9003	0.9236	0.9726	24	0.8479	0.8752	0.9388
12	0.8429	0.8695	0.9367	25	0.8256	0.8610	0.9286
13	0.8396	0.8600	0.9302	26	0.8694	0.8887	0.9385

TABLE 6: Evaluation data of 26 patients with brain tumor segmentation using the DCNN-F-SVM model.

Number	DSC	Sensitivity	Specificity	Number	DSC	Sensitivity	Specificity
1	0.8923	0.9220	0.9663	14	0.8956	0.9222	0.9785
2	0.8867	0.9063	0.9368	15	0.8896	0.9185	0.9669
3	0.9091	0.9193	0.9702	16	0.8876	0.9104	0.9678
4	0.8782	0.9014	0.9588	17	0.8782	0.9086	0.9585
5	0.9026	0.9289	0.9795	18	0.9020	0.9103	0.9786
6	0.8998	0.9098	0.9405	19	0.9023	0.9123	0.9752
7	0.9056	0.9196	0.9743	20	0.8885	0.9116	0.9600
8	0.9030	0.9229	0.9696	21	0.8963	0.9205	0.9696
9	0.8927	0.9110	0.9711	22	0.9004	0.9287	0.9745
10	0.9126	0.9289	0.9806	23	0.9102	0.9258	0.9798
11	0.9185	0.9298	0.9885	24	0.8763	0.9115	0.9598
12	0.8789	0.9110	0.9605	25	0.8689	0.9088	0.9469
13	0.8825	0.9168	0.9693	26	0.8996	0.9305	0.9797

TABLE 7: Evaluation indexes of the segmentation results of the three models.

Method	DSC	Sensitivity	Specificity
SVM	0.8705	0.9001	0.9586
CNN	0.8869	0.9152	0.9657
DCNN-F-SVM	0.9010	0.9236	0.9889

values are generally distributed around 0.91, and there is about 0.12 up and down floating error. The specificity value is generally distributed around 0.96, and there is about 0.09 up and down floating error.

Table 7 shows the DSC, specificity, and sensitivity values of the three methods. The proposed DCNN-F-SVM has increased in comparison with CNN and SVM used independently, in which the three indicators in the table (DSC, sensitivity, and specificity) are 3.5%, 2.6%, and 3.2% higher compared to those of SVM and 1.6%, 0.9%, and 2.4% higher compared to those of CNN. The proposed model can indeed improve the segmentation performance.

## 5. Conclusion

The diagnosis of brain diseases requires accurate diagnosis without deviation. Any misdiagnosis will cause irreparable losses. The incidence of brain tumors in brain diseases has been high, and the number of patients has increased year by year. This has also increased the workload of medical personnel in this field to a certain extent. An accurate and efficient method of brain tumor image segmentation needs to be urgently proposed, which has solved the increasing demand. Based on this background, this paper proposes a depth classifier to improve the segmentation accuracy and achieve automatic segmentation without manual intervention. The classifier is mainly composed of DCNN and integrated SVM connected in series. The implementation of the model is divided into three stages. In the first stage, a deep convolutional neural network is trained to learn the mapping from the image space to the tumor marker space. In the second stage, the predicted labels obtained from the deep convolutional neural network training are input into the integrated support vector machine classifier together with



the test images. In the third stage, a deep convolutional neural network and an integrated support vector machine are connected in series to train a deep classifier. The simulation implementation verified the superiority and effectiveness of the proposed model. However, the proposed model still has shortcomings such as long calculation time. How to optimize the algorithm and shorten the running time will be the next research content.

## Data Availability

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

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This research was supported by the Program for China Northwest Cohort Study of the National Key Research and Development Program of China (Grant numbers 2017YFC0907200 and 2017YFC0907201) and Project of Birth Defect Control and Prevention in Shaanxi of the Shaanxi Health and Family Planning Commission (Grant number Sxwsjswzfcg2016-013).

## References

- [1] C. S. Muir, H. H. Storm, and A. Polednak, "Brain and other nervous system tumours," *Cancer surveys*, vol. 19, no. 20, pp. 369–392, 1994.
- [2] D. N. Louis, H. Ohgaki, O. D. Wiestler et al., "The 2007 WHO classification of tumours of the central nervous system," *Acta Neuropathologica*, vol. 114, no. 5, pp. 547–547, 2007.
- [3] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Physics in Medicine and Biology*, vol. 58, no. 13, pp. R97–R129, 2013.
- [4] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
- [5] S. Wang and R. M. Summers, "Machine learning and radiology," *Medical Image Analysis*, vol. 16, no. 5, pp. 933–951, 2012.
- [6] P. Qian, H. Friel, M. S. Traugher et al., "Transforming UTE-mDixon MR abdomen-pelvis images into CT by jointly leveraging prior knowledge and partial supervision," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.
- [7] P. Qian, Y. Chen, J.-W. Kuo et al., "mDixon-Based Synthetic CT Generation for PET Attenuation Correction on Abdomen and Pelvis Jointly Using Transfer Fuzzy Clustering and Active Learning-Based Classification," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 819–832, 2020.
- [8] Y. Jiang, K. Zhao, K. Xia et al., "A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation," *Journal of Medical Systems*, vol. 43, no. 5, pp. 118:1–118:9, 2019.
- [9] P. Qian, K. Xu, T. Wang et al., "Estimating CT from MR Abdominal Images Using Novel Generative Adversarial Networks," *Journal of Grid Computing*, vol. 18, pp. 211–226, 2020.
- [10] K. Xia, X. Zhong, L. Zhang, and J. Wang, "Optimization of diagnosis and treatment of chronic diseases based on association analysis under the background of regional integration," *Journal of Medical Systems*, vol. 43, no. 3, pp. 46:1–46:8, 2019.
- [11] P. Qian, C. Xi, M. Xu et al., "SSC-EKE: semi-supervised classification with extensive knowledge exploitation," *Information Sciences*, vol. 422, pp. 51–76, 2018.
- [12] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [13] P. Qian, J. Zhou, Y. Jiang et al., "Multi-view maximum entropy clustering by jointly leveraging inter-view collaborations and intra-view-weighted attributes," *IEEE Access*, vol. 6, pp. 28594–28610, 2018.
- [14] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [15] A. Stadlbauer, E. Moser, S. Gruber et al., "Improved delineation of brain tumors: an automated method for segmentation based on pathologic changes of <sup>1</sup>H-MRSI metabolites in gliomas," *NeuroImage*, vol. 23, no. 2, pp. 454–461, 2004.
- [16] W. Deng, W. Xiao, H. Deng, and J. Liu, "MRI brain tumor segmentation with region growing method based on the gradients and variances along and inside of the boundary curve," in *2010 3rd International Conference on Biomedical Engineering and Informatics*, vol. 1, pp. 393–396, Yantai, China, 2010.
- [17] D. Jayadevappa, S. S. Kumar, and D. S. Murty, "A hybrid segmentation model based on watershed and gradient vector flow for the detection of brain tumor," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 2, no. 3, pp. 29–42, 2009.
- [18] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Medical Image Analysis*, vol. 8, no. 3, pp. 275–283, 2004.
- [19] A. Gooya, K. M. Pohl, M. Bilello et al., "GLISTR: glioma image segmentation and registration," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1941–1954, 2012.
- [20] D. Kwon, R. T. Shinohara, H. Akbari, and C. Davatzikos, "Combining generative models for multifocal glioma segmentation and registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 763–770, Cham, Germany: Springer, 2014.
- [21] H. Khotanlou, O. Colliot, J. Atif, and I. Bloch, "3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy Sets and Systems*, vol. 160, no. 10, pp. 1457–1473, 2009.
- [22] S. Bauer, L. P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 354–361, Berlin, Germany: Springer, 2011.
- [23] E. Geremia, B. H. Menze, and N. Ayache, *Spatial decision forests for glioma segmentation in multi-channel MR images*, MICCAI Challenge on Multimodal Brain Tumor Segmentation, Springer, Germany, 2012.

- [24] L. Le Folgoc, A. V. Nori, S. Ancha, and A. Criminisi, "Lifted auto-context forests for brain tumour segmentation," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 171–183, Cham, Germany: Springer, 2016.
- [25] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [26] Z. Zhang and E. Sejdić, "Radiological images and machine learning: trends, perspectives, and prospects," *Computers in Biology and Medicine*, vol. 108, no. 6, pp. 354–370, 2019.
- [27] D. Zikic, Y. Ioannou, M. Brown, and A. Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," *Proceedings MICCAI-BRATS*, vol. 36, pp. 36–39, 2014.
- [28] P. Dvořák and B. Menze, "Local structure prediction with convolutional neural networks for multimodal brain tumor segmentation," in *International MICCAI Workshop on Medical Computer Vision*, pp. 59–71, Cham, Germany: Springer, 2015.
- [29] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
- [30] S. Pereira, A. Pinto, V. Alves et al., "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI," in *BrainLes 2015*, pp. 131–143, Cham, Germany: Springer, 2015.
- [31] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.
- [32] K. Kamnitsas, C. Ledig, V. F. J. Newcombe et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
- [33] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Annual Conference on Medical Image Understanding and Analysis*, pp. 506–517, Cham, Germany: Springer, 2017.
- [34] Y. LeCun, B. E. Boser, J. S. Denker et al., *Handwritten digit recognition with a back-propagation network*, Advances in neural information processing systems, Morgan Kaufmann Publishers, Inc, 1990.
- [35] R. Hechtmielsen, "Theory of the backpropagation neural network," in *International 1989 Joint Conference on Neural Networks*, vol. 1, pp. 593–605, Washington, DC, USA, 1989.
- [36] W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, "Parallel distributed processing model with local space-invariant interconnections and its optical architecture," *Applied Optics*, vol. 29, no. 32, pp. 4790–4797, 1990.
- [37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097–1105, 2012.

## Research Article

# Multimodal MRI Brain Tumor Image Segmentation Using Sparse Subspace Clustering Algorithm

Li Liu <sup>1</sup>, Liang Kuang<sup>1,2</sup> and Yunfeng Ji<sup>1</sup>

<sup>1</sup>School of IoT Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China

<sup>2</sup>School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China

Correspondence should be addressed to Li Liu; [liul2@jsit.edu.cn](mailto:liul2@jsit.edu.cn)

Received 22 April 2020; Revised 24 May 2020; Accepted 8 June 2020; Published 4 July 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Li Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Brain tumors are one of the most deadly diseases with a high mortality rate. The shape and size of the tumor are random during the growth process. Brain tumor segmentation is a brain tumor assisted diagnosis technology that separates different brain tumor structures such as edema and active and tumor necrosis tissues from normal brain tissue. Magnetic resonance imaging (MRI) technology has the advantages of no radiation impact on the human body, good imaging effect on structural tissues, and an ability to realize tomographic imaging of any orientation. Therefore, doctors often use MRI brain tumor images to analyze and process brain tumors. In these images, the tumor structure is only characterized by grayscale changes, and the developed images obtained by different equipment and different conditions may also be different. This makes it difficult for traditional image segmentation methods to deal well with the segmentation of brain tumor images. Considering that the traditional single-mode MRI brain tumor images contain incomplete brain tumor information, it is difficult to segment the single-mode brain tumor images to meet clinical needs. In this paper, a sparse subspace clustering (SSC) algorithm is introduced to process the diagnosis of multimodal MRI brain tumor images. In the absence of added noise, the proposed algorithm has better advantages than traditional methods. Compared with the top 15 in the Brats 2015 competition, the accuracy is not much different, being basically stable between 10 and 15. In order to verify the noise resistance of the proposed algorithm, this paper adds 5%, 10%, 15%, and 20% Gaussian noise to the test image. Experimental results show that the proposed algorithm has better noise immunity than a comparable algorithm.

## 1. Introduction

Tumor is one of the common malignant diseases that endanger human health. According to origin, tumors are generally divided into primary and secondary. Compared with breast, lung, and esophageal tumors, the incidence of brain tumors is relatively low. Compared with the overall incidence of human tumors, it accounts for about 1.4%; however, the mortality rate reaches 2.4% of human tumors [1]. Glioma is the most common primary brain tumor in adults. It is mainly distributed in glial cells and the tissues it infiltrates, and it is the most common malignant brain tumor. According to the nature of tumors, gliomas are generally divided into benign and malignant. Benign gliomas generally grow relatively slowly, patients have a longer survival period, and the long course of disease is the main manifestation of benign gliomas.

Malignant gliomas generally grow faster, and the short course is a prominent manifestation of malignant gliomas. If the intracranial lesions can be detected as soon as possible, and the corresponding treatments can be implemented, the health hazards of brain tumors to humans can be reduced. CT or MRI imaging to analyze the pathological state of brain tissue is currently the mainstream method for examining brain tumors. Different imaging techniques have different advantages for tumor diagnosis. Compared with CT imaging, MRI uses a noninvasive imaging method, which can provide the observer with high-quality images without damage and skull artifacts, with clear anatomical structure, and with very good soft tissue resolution. At the same time, intracranial images in any direction can be obtained by adjusting the relevant parameters. In addition, using different imaging sequences, MRI of different angles or modalities of the same

tissue can be obtained. This type of image is generally referred to as a multimodal MRI image.

The effective diagnosis of brain tumors requires the successful segmentation of tumors in brain images. Based on the results of the segmentation, the doctor can determine the shape, size, and specific location of the tumor. According to the segmentation results of the tumor in the image, a corresponding treatment plan is given. Due to the increase in the number of patients with brain diseases and the development of intelligent diagnostic technology, the research work based on brain tumors continues to increase. The International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) began in 2012 and has organized competitions based on multimodal brain tumor segmentation for four consecutive years, greatly promoting the development of brain tumor segmentation technology. It is of great research value and practical significance to improve the diagnosis efficiency by mining potential pathological information of MRI brain tumor images through image processing technology and machine learning methods. However, the tumor is only characterized by gray-scale information on the MRI image, and the edge of the tumor structure and the normal tissue have significant gray-scale similarity. Simultaneously, the size, location, shape, and corresponding expansion of the tumor in the brain tissue will show different states with different patients. These characteristics pose challenges to the development of tumor segmentation technology.

The so-called brain tumor segmentation refers to the process of segmenting various tumor tissues from a variety of conventional brain tissues. In general, the segmentation methods of brain tumor images can be summarized into three categories [2]: purely artificial, semiautomatic, and fully automatic segmentation. Manual segmentation refers to manually drawing the outline of the target tissue. Figure 1 is a schematic diagram of manual segmentation. Manual segmentation is boring and time consuming, so it cannot meet the growing demand for segmentation. In addition, each segmenter has a different segmentation style, which leads to deviations in segmentation results. Although manual segmentation has many disadvantages, manual segmentation has the highest segmentation accuracy so far, and is often used as the ground truth for automatic segmentation. Semi-automatic segmentation is sensitive to initialization. Users need to input certain initialization data to get the final segmentation result. Fully automatic segmentation does not need to set any parameters manually and can automatically locate and segment the tumor area.

There is still a lack of a general method that can process all brain tumor images and obtain satisfactory results currently. Usually, the segmentation method is aimed at specific image data. Reviewing related literature, tumor image segmentation methods can be summarized as follows:

- (1) Threshold based method. The practicability and segmentation effect of this method are very good. The histogram in the global threshold can be expressed as a bimodal model, and a single threshold can be used to distinguish tumor from back-

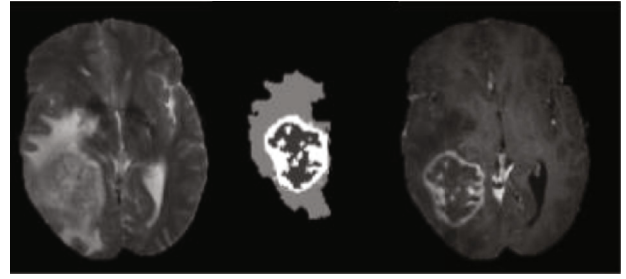


FIGURE 1: Tumor labels manually segmented on T1c and T2 modal images.

ground. Reference [3] proposes an unsupervised method to enhance pixel grayscale and utilize it to segment brain tumors in T1c images. If there are multiple types of regions in the image, a multi-threshold strategy needs to be added to the segmentation method, called local threshold. For the local mean, it can be obtained by estimating the local statistical characteristics, such as gray average [4] and data Gaussian distribution [5]. Generally, the threshold-based method cannot use all the information of the MRI image, and the segmentation result is relatively rough. Therefore, the threshold-based method in brain tumor segmentation is first applied

- (2) Area-based approach. Through predefined similarity criteria, in the way of merging neighboring pixels in the intersecting areas, the target MRI brain image is divided into the required subareas. Reference [6] applied region growth to MRI tumor segmentation image segmentation with good results. Reference [7] proposed an improved method of region growth. This method obtains a more exact boundary message by reducing the volume effect. The leak gap that may be generated after the division is also filled to a certain extent. As a morphological method, watershed segmentation represents the target contour edge as a partial watershed, which is widely used in brain tumor segmentation. References [8, 9] proposed a multiscale watershed transformation method. Reference [10] constructed an artificially assisted segmentation method by the hierarchical watershed method. From the principle of the watershed segmentation method, this kind of image edge and region watershed conversion easily produces oversegmentation. In order to solve this problem, some related processing methods have been excavated one after another
- (3) Pixel classification method. The collected MRI brain tumor data generally has two formats, namely, 2D slices and 3D volume. If it is a brain tumor segmentation based on slice format, its essence is the same as traditional image segmentation. The pixel-based method mainly uses the pixel characteristics of the image, and uses some related classifiers to classify



all the pixels in the brain tissue image, so as to achieve the effect of segmentation. Unsupervised classification is mainly represented by clustering [11, 12]. The core idea is to measure the relationship between tumor tissue and other tissues in the tumor image. The supervised classifier [13, 14] mainly uses those labeled training samples to train the relevant parameters in the model, which has reached the optimal tumor segmentation effect [15].

- (4) Model-based method. Model-based tumor segmentation methods are mainly 3D-oriented volume data, followed by 2D slice data. The most typical are the active contour model [16] and the level set method [17]. On the basis of these two models, tumor segmentation has formed two schools: the segmentation methods based on the generative algorithm and the discriminant algorithm. The generation algorithm uses the unique information of various organizations to predict the information of brain tissue that cannot be captured in the image [18–20]. In some generative models, in order to solve the problem of difficult coding of a priori knowledge of tumors, the diseased tissue of the tumor can be modeled as the desired shape [21–23], or it can be inferred using the given patient image and the tumor growth model's possible location of the tumor structure [24]. Discriminant methods generally require a certain size of training samples [25–27]. After many trainings, the processing effect of the discriminant model is more robust to the effects of MRI image artifacts and grayscale information. Figure 2 shows the basic flow of model-based tumor segmentation. For effective training, the first step of this type of method is generally to extract local grayscale differences [28] or gray-scale distribution and other voxel-wise features [29], and then send these features to the discriminant classifier of the model. In order to combine the advantages of discriminant models and generative models, a method called generative discriminant model [30, 31] was proposed

In this paper, the BRATS 2015 competition database is used as the experimental object, and the traditional segmentation method and the sparse subspace clustering method based on sparse representation are used to segment the brain tumor images. The main innovations of this article are as follows:

- (1) Introduce the sparse subspace clustering algorithm to achieve brain tumor image segmentation. The advantage of this algorithm is to use low-dimensional data to recover and approximate high-dimensional data, effectively reducing the dimension of high-dimensional data while retaining the correlation between the data. The introduction of this algorithm can solve the problem of excessive data dimension
- (2) This article focuses on the segmentation of MRI brain tumor images under multimodality. In the single-modality image fusion strategy, a simple and fast linear fusion strategy is selected. Before

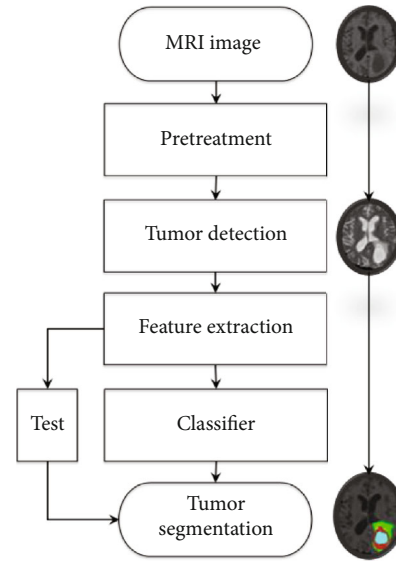


FIGURE 2: Flow of the brain tumor segmentation method based on the discriminant model.

segmenting multimodal images, the image is pre-processed by superpixel segmentation, feature vectors are extracted, and the data dimension is reduced. Experimental results show that for brain tumor segmentation, multimodal brain tumor information can be used as much as possible to obtain more accurate segmentation results

## 2. Related Information

**2.1. Multimodal MRI Brain Tumor Image Introduction.** Multimodal MRI images are images of the same tissue under different contrasts obtained through different MR development sequences. When tumors and other lesions occur in brain tissue, water molecules existing in free form in brain tumors begin to undergo lesion reactions, such as tissue edema. In Flair and T2 images, the water molecules in the bound state are displayed in the form of high signals. Therefore, it is theoretically feasible to use Flair modal MRI images as the main basis for segmenting the entire tumor. However, due to some special circumstances, the tumor will also show irregular changes in the Flair image. At this time, the image data of the T2 mode can provide additional reference. Figure 3 depicts three different sets of Flair and T2 images. Among them is (1) the Flair image, (2) the T2 image, and (3) the artificially labeled tumor structure image. The data used in this paper are all from the BRATS 2015 [32] database, and the database includes images in four modes: T1, T1c, T2, and Flair.

**2.2. Difficulties in MRI Brain Tumor Image Segmentation.** There are many difficulties in the segmentation of MRI brain tumor images. These difficulties can be summarized as follows:

- (1) The most typical problem of MRI comes from the different nonstandard intensity ranges obtained by



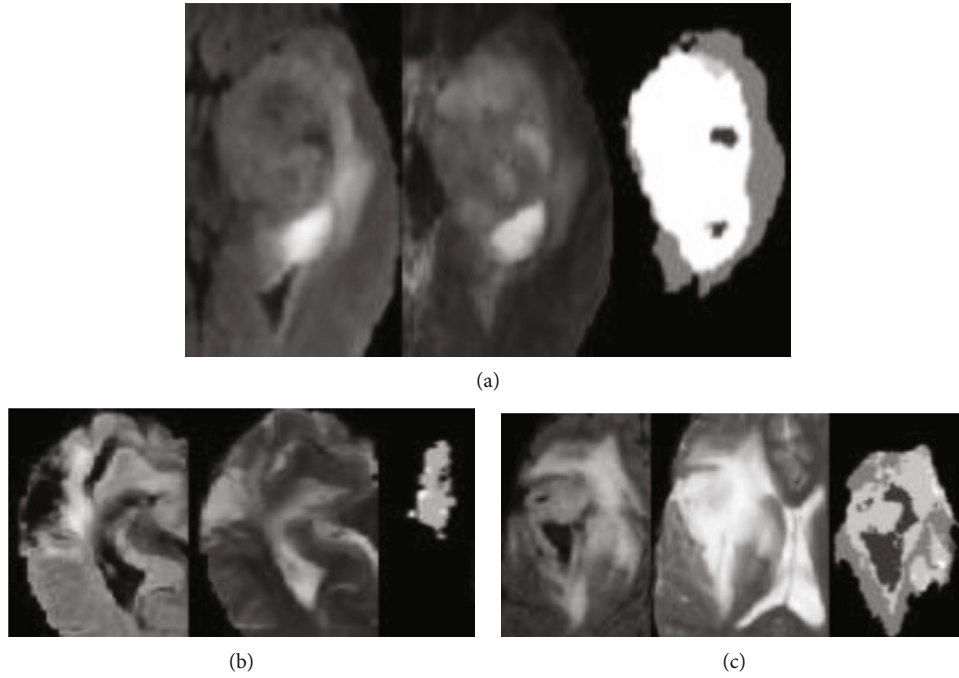


FIGURE 3: Flair and T2 images and corresponding tumor labels.

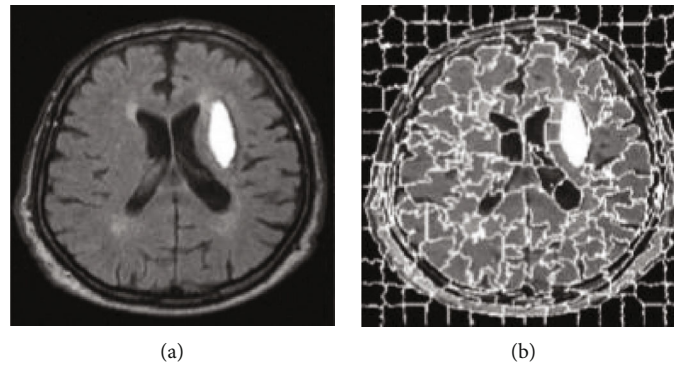


FIGURE 4: SLIC super pixel segmentation. (a) Original image. (b) SLIC superpixel segmentation image.

different scanners. Because of different magnetic field strengths and acquisition protocols, for the same patient, the brain MRI strength values are also different between hospitals

- (2) The brain tumor itself has no fixed shape or prior knowledge. Brain pathology can appear anywhere in the brain and can have any shape. In addition, the gray value range of this pathology may overlap with the gray value range of healthy tissues, making segmentation of brain tumors more complicated
- (3) MRI has nonnegligible white Rician noise during the acquisition process [33]
- (4) Uniform organization is often affected by changes in the spatial intensity of each dimension. This is caused by the bias field effect. The MRI bias affects the smoothed low-frequency signal of the image inten-

sity. This problem requires an offset field correction preprocessing step, which usually increases the intensity value around the brain

- (5) Large tumors or lesions in the brain may distort the overall structure of the brain, making some procedures impossible to perform. For example, a larger tumor may affect the overall symmetry of the brain, making it impossible to calculate the left-right symmetry feature. In addition, brains with large tumors are difficult to register with healthy brain templates

### 3. Brain Tumor Image Segmentation Based on Sparse Subspace Clustering Algorithm

Sparse representations are widely used in image segmentation algorithms. Sparse representations can effectively reduce the complexity of data operations and bring convenience

to the subsequent processing of data. SSC is a clustering algorithm based on sparse representation and subspace clustering [34]. Before segmentation, the target image needs to be preprocessed.

**3.1. Image Preprocessing.** Before SSC splits an image, the image needs to be split into superpixels. Superpixels are irregular image blocks composed of a series of adjacent pixels with similar characteristics, such as texture, color, and brightness. It replaces a large number of pixels with a few superpixels, which effectively reduces the amount of data that expresses the features of the picture, thereby reducing the complexity of image postprocessing. Superpixel segmentation algorithms are currently divided into two types, one is based on graph theory, and the other is based on gradient descent, such as Simple Linear Iterative Clustering (SLIC) [35]. The segmentation method based on gradient descent belongs to an iterative segmentation method. First, an initial clustering is given, and then the gradient clustering method is used to modify the result of the previous clustering, and iterate continuously until the convergence condition is satisfied. The superpixel rendering using SLIC segmentation is shown in Figure 4.

**3.2. Basic Model.** The algorithm is to assume that the data is composed of high-dimensional spatial data, and each data can be represented in a low-dimensional subspace. That is, by letting the data in the high-dimensional space be expressed linearly with the data in the low-dimensional subspace, the low-dimensional subspace to which the data belongs can be clearly known, which is beneficial to the clustering operation. The basic framework of sparse subspace clustering is shown in Figure 5.

The SSC model building process is as follows.

Given a set of datasets  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , the dimension is  $D$ , located in  $n$  linear subspaces  $\{S_i\}$ ,  $i = 1, 2, \dots, n$ . The dimensions of the linear subspace are  $\{d_i\}$ ,  $i = 1, 2, \dots, n$ . Then define the matrix

$$X = \{x_1, x_2, x_3, \dots, x_n\} = [X_1, X_2, \dots, X_n] \times Z, \quad (1)$$

where  $X_i \in R^{D \times N_i}$  is a matrix of rank  $d_i$  composed of the  $i$ th subspace data.  $Z$  is the permutation matrix. Subspace clustering is essentially to obtain the  $X_i \in R^{D \times N_i}$  matrix.

Subspace representation means that every data in matrix  $X$  can be linearly represented by data in the same subspace except for itself:

$$x_i = X a_i, \quad a_{ii} = 0, \quad (2)$$

where  $a_i = [a_{i1}, a_{i2}, \dots, a_{in}]^T$ . Formula (1) can be written in matrix form as follows:

$$X = XA, \quad A_{ii} = 0, \quad (3)$$

where  $A = [a_1, a_2, \dots, a_n] \in R^{N \times N}$  is a sparse matrix. In order to make the sparse matrix  $A$  the most sparse, that

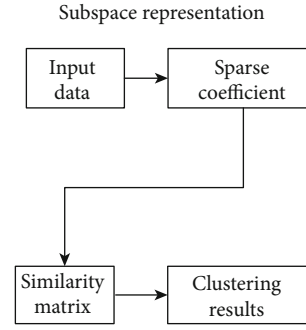


FIGURE 5: The basic framework of sparse subspace clustering.

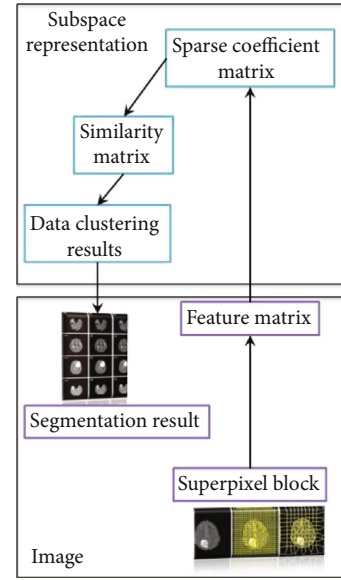


FIGURE 6: Image segmentation framework based on sparse subspace clustering.

is, the nonzero values in matrix  $A$  are minimized, by obtaining the  $l_0$  - norm to minimize, we use convex optimization to perform the following process:

$$\begin{aligned} \min \quad & \|A\|_0 \\ \text{s.t.} \quad & X = XA, \quad A_{ii} = 0. \end{aligned} \quad (4)$$

However, the solution of the  $l_0$  - norm is an NP-Hard problem in practical problems. Usually the  $l_1$  - norm is used to replace the  $l_0$  - norm to solve, so as to convert the subspace representation model to

$$\begin{aligned} \min \quad & \|A\|_1 \\ \text{s.t.} \quad & X = XA, \quad A_{ii} = 0. \end{aligned} \quad (5)$$

**3.3. Brain Tumor Image Segmentation Based on Sparse Subspace Clustering.** Image segmentation is the process of segmenting images into nonoverlapping regions and

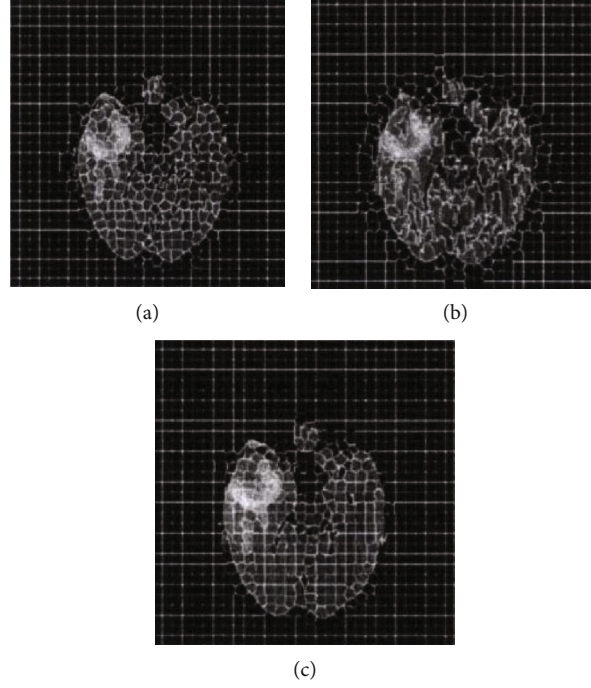


FIGURE 7: FLAIR image segmentation result when the  $m$  value changes. (a) Superpixel segmentation result when  $m = 10$ . (b) Superpixel segmentation result when  $m = 20$ . (c) Superpixel segmentation result when  $m = 30$ .

extracting ROI from them, while sparse subspace clustering is a process used to cluster data of the same class into the same subspace. An image contains multiple target images with a complex texture structure, but the features on the image are composed of multiple low-dimensional subspace data. Therefore, the sparse subspace clustering algorithm can be used to segment the image. First, divide the image to be divided into multiple superpixel blocks, and divide the superpixel blocks of the same target image into the same subspace, so as to achieve the purpose of extracting the target image. The process is shown in Figure 6.

A variety of modal image fusion strategies use linear fusion. Linear fusion is the simplest multimodal MRI brain tumor image fusion method. It is a pixel-level fusion method, and the processing object is pixels. It is mainly to operate the pixel unit in each modal brain image, so as to comprehensively process the pixel information in each modal brain tumor image. Through the linear fusion operation, multimodal brain images can be converted into single-modal brain images containing multimodal brain tumor tissue information. Thus, multimodal image segmentation is converted into single-modal image segmentation, and the operation of multimodal processing is simplified. The specific operation of linear fusion is as follows:

$$F_{ij} = \alpha T_1(i, j) + \beta T_2(i, j) + \varepsilon T_3(i, j), \quad (6)$$

where  $F_{ij}$  is the fused image;  $T_1(i, j)$ ,  $T_2(i, j)$ , and  $T_3(i, j)$  are the pixel values of  $T_1$ ,  $T_2$ , and  $T_3$  at position  $(i, j)$ ; and  $\alpha$ ,  $\beta$ , and  $\varepsilon$  are the weights of each modal image, and

meets  $\alpha + \beta + \varepsilon = 1$ . Figure 7 is a fusion image of multimodal images. Using the linear fusion operation, we use the following Flair ratio to obtain the fusion image in the figure:  $T_1 : T_1c : T_2 = 3 : 2 : 1 : 4$ . After preprocessing the fused image, SSC can be used to complete the multimodal image segmentation.

The steps of the SSC-based multimodal image segmentation algorithm are as follows:

- (1) Input image  $I$  and use the preprocessing algorithm described in Section 3.1 to divide the fused image into  $N$  superpixel blocks
- (2) Extract  $D$ -dimensional feature vectors from superpixel blocks to form a feature matrix  $\{X_i\}$  ( $i = 1, 2, \dots, n$ )
- (3) Use the basic model of sparse subspace clustering to obtain the sparse coefficient matrix  $C$
- (4) Calculate the similarity matrix  $W = |C| + |C^T|$ , where  $w_{ij} = w_{ji} = |c_{ij}| + |c_{ji}|$
- (5) The clustering result is obtained by using the spectral clustering algorithm

## 4. Simulation Experiment Analysis

**4.1. Experiment-Related Settings.** The comparison algorithms mainly include FCM, SVM, and the top 15 results of the Brats 2015 challenge. The experimental data of this paper is Brats 2015 [28]. The database contains data of two types of patients, those with benign tumors and those with malignant

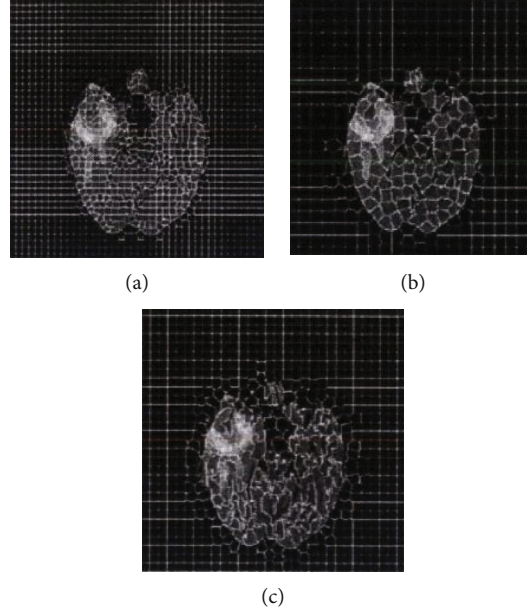


FIGURE 8: FLAIR image segmentation result when the  $n$  value changes. (a) Superpixel segmentation result when  $n = 300$ . (b) Superpixel segmentation result when  $n = 500$ . (c) Superpixel segmentation result when  $n = 1000$ .

TABLE 1: Evaluation indicator introduction.

Number	Index	Explanation
1	$Dice(P, T) = 2 P \cap T  /  P  +  T $ .	The Dice coefficient is a set similarity measurement method. In the image, it mainly refers to the degree to which the actual segmentation result and the golden segmentation result overlap each other, and the value is $[0, 1]$ . Among them, 0 represents that there is no overlap between the actual segmentation result and the golden segmentation result, which represents the worst segmentation accuracy at this time, and 1 represents that the actual segmentation result and the golden segmentation result completely overlap, which represents the optimal segmentation accuracy at this time.
2	$Jaccard(P, T) =  P \cap T  /  P  +  T  -  P \cap T $ .	The Jaccard coefficient is a method similar to the Dice coefficient that relies on similarity as a measure. It describes the degree of overlap between the actual segmentation result and the golden segmentation result from another perspective.
3	$Precision(P, T) =  P \cap T  /  P $ .	The false positive rate (Precision) reflects the accuracy of the actual segmentation result. The ratio of the overlap between the actual segmentation result and the golden segmentation result is used for description. The higher the ratio, the higher the proportion of the golden result included in the actual segmentation result.
4	$Recall(P, T) =  P \cap T  /  T $ .	The true positive rate (Recall) reflects the accuracy of the actual results in the actual segmentation results. It refers to the ratio of the overlap between the actual and golden section results. The higher the ratio, the higher the proportion of the true segmentation result in the golden section.

tumors, and contains brain image data of 274 patients. Each patient's brain image data contains Flair images, T1 images, T1c images, T2 images, and golden section results. The size of each modal image is  $240 * 240$ . We randomly selected data from 25 patients with brain tumors. Each patient's data includes five parts, namely, the Flair mode, the T1 mode, the T1c mode, the T2 mode, and the golden section results. The data size of each mode is  $240 * 240 * 163$ . Because the two-dimensional tumor pictures of the same patient are

similar, a set of two-dimensional multimodal brain tumor images is extracted from the data of each patient. There was a total of 25 sets of multimodal brain tumor image data. Among them, there are 15 groups of malignant tumor data and 10 groups of benign tumor data.

The performance of the algorithm in this paper mainly depends on the quality of the superpixels. The quality of the superpixels is controlled by the number  $K$  of the superpixels and the compact factor  $m$ . In this paper, the SLIC

TABLE 2: Comparison of multimodal image segmentation results.

Experimental sample		Index			Recall
		Dice	Jaccard	Precision	
Malignant tumor	1	0.8923	0.8042	0.9627	0.8287
	2	0.8596	0.7752	0.9748	0.7789
	3	0.8385	0.7431	0.9263	0.7821
	4	0.9507	0.8785	0.9874	0.9264
	5	0.9310	0.7886	0.9845	0.7954
	6	0.8746	0.7964	0.9678	0.8103
	7	0.9152	0.8522	0.9034	0.9371
	8	0.8386	0.7371	0.9976	0.7352
	9	0.8731	0.7796	0.9948	0.7911
	10	0.8694	0.7649	0.9563	0.8002
	11	0.8627	0.7628	0.9915	0.7832
	12	0.7016	0.5364	0.9997	0.5349
	13	0.8018	0.6742	0.9306	0.7132
	14	0.8220	0.6976	0.9736	0.7120
	15	0.8129	0.6842	0.9637	0.7058
Bright tumor	1	0.7961	0.6425	0.9264	0.6779
	2	0.8129	0.8413	0.9779	0.8646
	3	0.9298	0.6624	0.9836	0.6732
	4	0.9401	0.6830	0.7375	0.9118
	5	0.9228	0.8698	0.9990	0.8769
	6	0.9418	0.8891	0.9862	0.8996
	7	0.7147	0.5510	0.9996	0.5534
	8	0.7753	0.6256	0.9676	0.6394
	9	0.9027	0.8244	0.9834	0.8426
	10	0.8624	0.7632	0.9623	0.7824
Mean		0.8577	0.7451	0.9615	0.7743

superpixel segmentation method needs to consider the density factor  $m$  and the number of target superpixel blocks [36]. In order to study the influence of the density factor  $m$ , the number of predefined superpixel blocks is 1000 at first. Then, we explore the impact of the change of the compact factor size on the segmentation results. The compact factor  $m = 10$  leads to a more rigid boundary, while  $m = 20$  will produce a very flexible boundary, but it will increase the shape and irregularity of the superpixel. Figure 7 is the result of FLAIR image segmentation when the value of  $m$  is different. By visually checking the superpixel boundary and area, when  $m = 20$ , the boundary can obtain a better segmentation result.

The next step is to determine the number of target superpixel blocks. Figure 8 shows the result of the FLAIR image segmentation when the value of  $m$  is 20 and the number  $n$  of the target superpixel blocks is different. When the compaction factor is fixed at  $m = 20$ , by changing the number of the target superpixel blocks, the Dice measure is used to evaluate the formation performance of the superpixels.

Based on the above experimental results, the compact factor  $m = 20$  in this experiment and the number of superpix-

TABLE 3: The top 15 segmentation results of the Brats 2015 challenge.

Rank	Dice	Precision	Recall
1	0.8730	0.8715	0.8916
2	0.8710	0.8621	0.9140
3	0.8720	0.8531	0.8633
4	0.8511	0.8619	0.8633
5	0.8739	0.8532	0.9180
6	0.8650	0.8530	0.9011
7	0.8325	0.8344	0.8457
8	0.8670	0.8623	0.8820
9	0.7760	0.7475	0.8635
10	0.8513	0.8248	0.9150
11	0.8417	0.8345	0.8917
12	0.8580	0.8716	0.8635
13	0.8512	0.8343	0.8916
14	0.8327	0.8527	0.8363
15	0.8328	0.8055	0.9090

TABLE 4: Comparison of evaluation indexes of different segmentation methods.

Methods	Evaluation index			
	Dice	Jaccard	Precision	Recall
FCM	0.7110	0.5564	0.7205	0.6975
SVM	0.8012	0.7056	0.9013	0.7558
SSC	0.8577	0.7451	0.9615	0.7743

els  $n = 500$ . The fuzzy factor in FCM is 2, and the parameter in SVM  $\sigma \in [10^{-5}, 10^5]$ .

**4.2. Evaluation Index.** There are four evaluation indicators commonly used in objective evaluation criteria, namely, the Dice coefficient, the Jaccard coefficient, the false positive rate (Precision), and the true positive rate (Recall). The four evaluation indicators are shown in Table 1.

**4.3. Simulation Results and Analysis.** Table 2 shows the evaluation index results of the algorithm for different groups of multimodal image segmentation results, and Table 3 shows the top 15 segmentation results of the Brats 2015 challenge. It can be seen from the comparison of the data in the table that the average Dice index of this algorithm is 0.8577. Compared with the top 15 of the Brats 2015 competition, the accuracy is not much different, and it can even exceed the results of several of the rankings. The average Precision index is as high as 0.9615, which is a big advantage compared with the top 15 data. Compared with the top 15, the true positive rate is slightly inadequate. This is because the top 15 competition algorithms use a deep learning algorithm to segment the tumor in three dimensions and use the three-dimensional information of the brain tumor. Considering comprehensively, the algorithm in this paper can use the two-dimensional information of brain tumors to obtain a



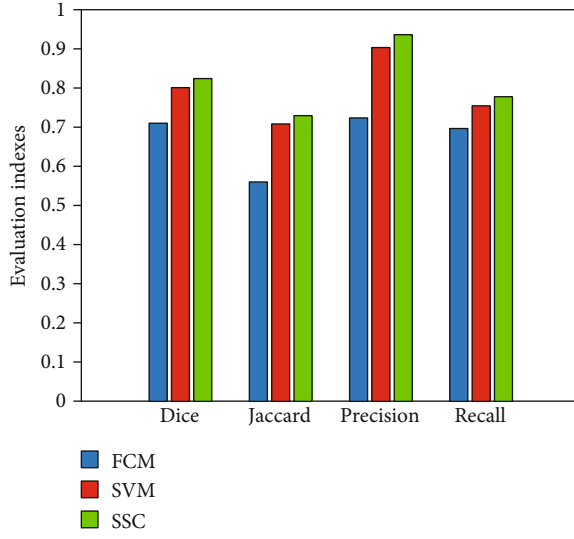


FIGURE 9: Comparison of evaluation indexes of various algorithms.

TABLE 5: Comparison of multimodal image segmentation results with 5% noise.

Experimental sample		Index			
		Dice	Jaccard	Precision	Recall
Malignant tumor	1	0.8123	0.7758	0.9036	0.8080
	2	0.8016	0.7469	0.9229	0.7568
	3	0.8001	0.7154	0.9086	0.7735
	4	0.8462	0.8369	0.9321	0.8858
	5	0.8528	0.7427	0.9650	0.7804
	6	0.8134	0.7528	0.9487	0.7940
	7	0.8347	0.8274	0.8936	0.9166
	8	0.8006	0.7144	0.9376	0.7130
	9	0.8104	0.7423	0.9721	0.7668
	10	0.8110	0.7417	0.9325	0.7878
	11	0.8234	0.7326	0.9639	0.7626
	12	0.6841	0.5146	0.9688	0.5229
	13	0.7695	0.6155	0.9129	0.7007
	14	0.7996	0.6639	0.9639	0.7013
	15	0.8005	0.6582	0.9470	0.6982
Bright tumor	1	0.7486	0.6301	0.9003	0.6663
	2	0.7952	0.8204	0.9575	0.8452
	3	0.8985	0.6471	0.9588	0.6620
	4	0.8625	0.6598	0.7176	0.9031
	5	0.8563	0.8446	0.9425	0.8206
	6	0.9012	0.8396	0.9393	0.8759
	7	0.6852	0.5329	0.9579	0.5414
	8	0.7410	0.6012	0.9493	0.6225
	9	0.8863	0.8071	0.9389	0.8223
	10	0.8401	0.7540	0.9522	0.7639
Mean		0.8110	0.7167	0.9315	0.7557

TABLE 6: Comparison of multimodal image segmentation results with 10% noise.

Experimental sample		Index			
		Dice	Jaccard	Precision	Recall
Malignant tumor	1	0.7585	0.7147	0.8452	0.7581
	2	0.7662	0.7020	0.8967	0.7052
	3	0.7596	0.6996	0.8746	0.7196
	4	0.8008	0.7989	0.9003	0.8320
	5	0.8020	0.7011	0.9114	0.7404
	6	0.7642	0.7102	0.9095	0.7462
	7	0.8001	0.7834	0.8482	0.8346
	8	0.7779	0.6996	0.9063	0.6730
	9	0.7823	0.7032	0.9101	0.7162
	10	0.7863	0.7142	0.9011	0.7285
	11	0.7903	0.7020	0.9039	0.7126
	12	0.6523	0.5011	0.9008	0.5028
	13	0.7124	0.6031	0.8557	0.6896
	14	0.7210	0.6313	0.8932	0.6745
	15	0.7695	0.6220	0.8712	0.6512
Bright tumor	1	0.7103	0.6102	0.8103	0.6326
	2	0.7533	0.7945	0.8410	0.8071
	3	0.8120	0.6103	0.8124	0.6426
	4	0.8236	0.6120	0.6731	0.8426
	5	0.8022	0.8008	0.8526	0.7945
	6	0.8471	0.8106	0.8989	0.8142
	7	0.6326	0.5030	0.9009	0.5231
	8	0.7002	0.5936	0.8855	0.6005
	9	0.8308	0.7852	0.8797	0.8030
	10	0.8001	0.7262	0.8722	0.7103
Mean		0.7662	0.6853	0.8702	0.7142

segmentation accuracy similar to the top 15 algorithms in the competition. It can be seen that the algorithm in this paper has certain value.

Table 4 shows the comparison of the average evaluation indexes of these three algorithms in 25 sets of data tests. From the comparison of the data in the table, we can see that the SSC algorithm used in each index is greatly improved compared to the other two algorithms.

Figure 9 is a comparison of the histograms of the various evaluation methods on the four evaluation indicators. From the figure, the greater advantages of the SSC algorithm can be clearly found.

In order to verify the noise resistance of the SSC algorithm, this paper adds 5%, 10%, 15%, and 20% Gaussian noise to the original image. The segmentation results after noise addition are shown in Tables 5–8. From the changing trends of the values of the four evaluation indicators in Tables 5–8, it can be analyzed that the tumor segmentation effect decreases with increasing noise. The greater the noise content, the worse the segmentation effect. This is completely consistent with theory.

TABLE 7: Comparison of multimodal image segmentation results with 15% noise.

Experimental sample		Index			
		Dice	Jaccard	Precision	Recall
Malignant tumor	1	0.7010	0.6235	0.7788	0.6963
	2	0.7121	0.6417	0.8293	0.6625
	3	0.7006	0.6582	0.8126	0.6741
	4	0.7259	0.7128	0.8253	0.6701
	5	0.7361	0.6733	0.8256	0.6693
	6	0.7140	0.6682	0.7896	0.6642
	7	0.7263	0.7117	0.7526	0.6723
	8	0.7030	0.6336	0.7864	0.6008
	9	0.7234	0.6402	0.8124	0.6037
	10	0.7026	0.6513	0.8102	0.6395
	11	0.7263	0.6412	0.8006	0.6279
	12	0.6136	0.4562	0.8152	0.4963
	13	0.6742	0.5846	0.7852	0.6230
	14	0.6892	0.6006	0.7984	0.6172
	15	0.7211	0.6001	0.8010	0.6003
Bright tumor	1	0.6982	0.5895	0.7142	0.6110
	2	0.7120	0.7312	0.7265	0.6753
	3	0.7361	0.5742	0.7416	0.5996
	4	0.7216	0.5863	0.6246	0.7582
	5	0.7121	0.7323	0.7693	0.7296
	6	0.7132	0.7125	0.8263	0.7369
	7	0.6030	0.4852	0.8082	0.5020
	8	0.6482	0.5611	0.8060	0.5801
	9	0.7413	0.7230	0.8132	0.6778
	10	0.7143	0.6736	0.8007	0.6256
Mean		0.7028	0.6346	0.7862	0.6405

Table 9 gives a comparison of the segmentation performance of the three algorithms under different noise ratios. Each data in the table is the average value of the above 25 sample data after division. Figure 10 shows the changing trend of the segmentation accuracy of the three algorithms with increasing noise. It can be concluded from Table 9 and Figure 10 that the SSC algorithm is relatively better in terms of the antinoise performance of the three algorithms. As the amount of noise increases, the performance of FCM declines the fastest, followed by SVM, and the relative decline of SSC is smaller. This further illustrates the feasibility and reference value of the SSC algorithm selected in this paper.

## 5. Conclusion

Different features have different effects on tumor segmentation results. In order to make better use of multimodal brain tumor image information, this paper proposes an SSC-based multimodal brain tumor image segmentation method. First, linear fusion is used to fuse multiple single-modality brain MRI images into one image to be processed; secondly, super-pixel features are extracted to construct a feature matrix; and

TABLE 8: Comparison of multimodal image segmentation results with 20% noise.

Experimental sample		Index			
		Dice	Jaccard	Precision	Recall
Malignant tumor	1	0.6013	0.5582	0.5786	0.6030
	2	0.6230	0.5631	0.6023	0.5436
	3	0.6058	0.5477	0.6113	0.5633
	4	0.6003	0.5693	0.6037	0.5721
	5	0.6146	0.5746	0.6012	0.5126
	6	0.6008	0.5832	0.5963	0.5362
	7	0.6001	0.5746	0.5748	0.5284
	8	0.6200	0.5365	0.5836	0.5369
	9	0.5963	0.5208	0.5862	0.5623
	10	0.5982	0.5300	0.5916	0.5123
	11	0.5996	0.5613	0.5746	0.5023
	12	0.5342	0.4203	0.5842	0.4523
	13	0.5846	0.5110	0.5532	0.5203
	14	0.5768	0.5030	0.5631	0.5417
	15	0.5636	0.5007	0.5711	0.5731
Bright tumor	1	0.5875	0.5114	0.5369	0.5324
	2	0.5939	0.5630	0.5284	0.6064
	3	0.5742	0.5023	0.5748	0.5412
	4	0.5936	0.5431	0.5303	0.6234
	5	0.5768	0.5623	0.5923	0.6127
	6	0.5693	0.5665	0.5830	0.6471
	7	0.5234	0.4528	0.5746	0.4864
	8	0.5236	0.5220	0.5822	0.5520
	9	0.5741	0.5360	0.5623	0.6113
	10	0.5698	0.5142	0.5722	0.5436
Mean	0.5842	0.5331	0.5765	0.5527	

TABLE 9: Comparison of segmentation performance of three algorithms under different noise ratios.

Noise ratio	Algorithm	Dice	Jaccard	Precision	Recall
5%	FCM	0.7002	0.5316	0.7010	0.6753
	SVM	0.7912	0.6902	0.8956	0.7412
	SSC	0.8410	0.7367	0.9515	0.7657
10%	FCM	0.6833	0.5241	0.6931	0.6595
	SVM	0.7800	0.6789	0.8763	0.7286
	SSC	0.8362	0.7353	0.9402	0.7542
15%	FCM	0.6658	0.5056	0.6767	0.6323
	SVM	0.7682	0.6574	0.8553	0.7001
	SSC	0.8268	0.7246	0.9362	0.7405
20%	FCM	0.6312	0.4712	0.6420	0.6125
	SVM	0.7404	0.6310	0.8211	0.6803
	SSC	0.8182	0.7131	0.9265	0.7327

finally, a sparse subspace clustering algorithm based on sparse representation is used to complete the segmentation. Using Brats 2015 competition data to experiment with the

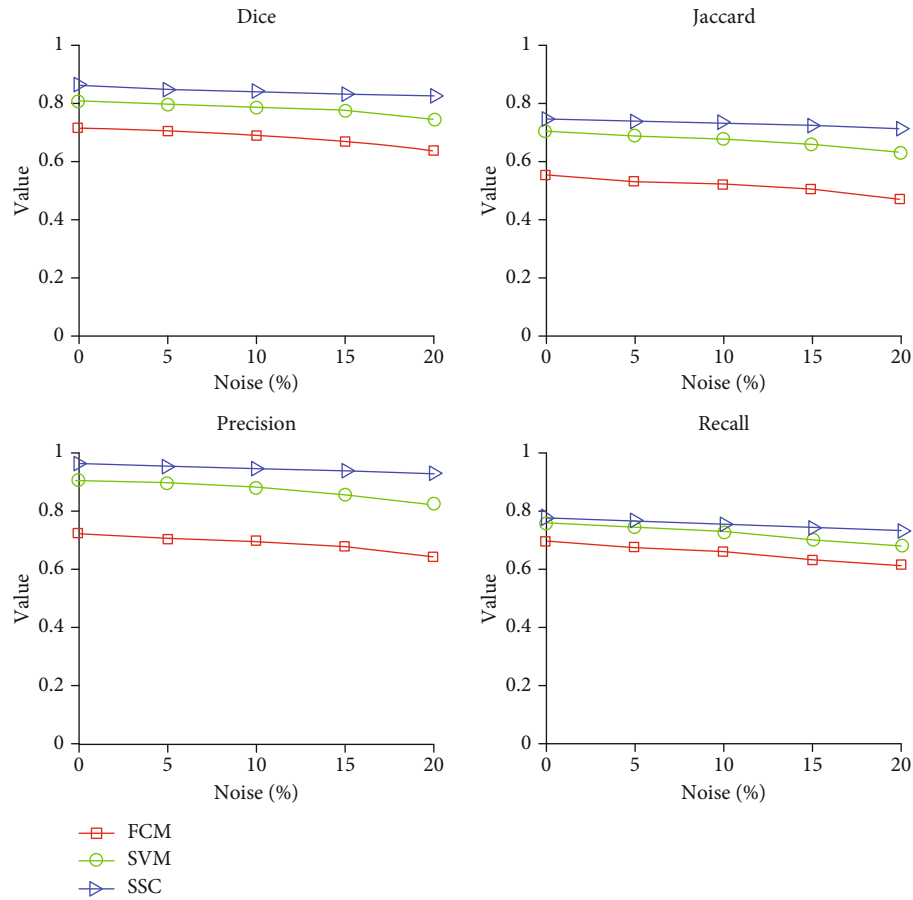


FIGURE 10: Segmentation performance of three algorithms under different noise contents.

proposed method, the results show that the method used can well integrate the tumor information of the multimodal images and obtain good segmentation results. After adding different proportions of noise, the segmentation performance of the proposed algorithm decreases significantly slower than that of the comparison algorithm, which also verifies that the proposed algorithm has good noise resistance. However, the method used in this paper has certain limitations. It needs to optimize the weights of various modal data fusions, which is very time consuming.

### Data Availability

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

### Conflicts of Interest

The authors declare no conflicts of interest.

### Acknowledgments

This work was supported by the Natural Science Foundation of Jiangsu Province (BK20131097), the Jiangsu Province Higher Vocational Education Industry-Education Integration Platform Construction Project, the Jiangsu Province

High-Level Backbone Professional Construction Project, the Jiangsu Province 333 High-Level Talent Cultivation Project ((2016) III-2887), and the Jiangsu Province Key Education Reform Project “Internet of Things Application Technology” Cross-Border Integration “Exploration and Practice of Project-Based Curriculum System Optimization” (2017JSJG077).

### References

- [1] R. Siegel, D. Naishadham, and A. Jemal, “Cancer statistics, 2013,” *CA: a Cancer Journal for Clinicians*, vol. 63, no. 1, pp. 11–30, 2013.
- [2] N. Gordillo, E. Montseny, and P. Sobrevilla, “State of the art survey on MRI brain tumor segmentation,” *Magnetic Resonance Imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.
- [3] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, “Tumour volume determination from MR images by morphological segmentation,” *Physics in Medicine and Biology*, vol. 41, no. 11, pp. 2437–2446, 1996.
- [4] A. Stadlbauer, E. Moser, S. Gruber et al., “Improved delineation of brain tumors: an automated method for segmentation based on pathologic changes of 1H-MRSI metabolites in gliomas,” *NeuroImage*, vol. 23, no. 2, pp. 454–461, 2004.
- [5] K. J. Shanthi and M. S. Kumar, “Skull stripping and automatic segmentation of brain MRI using seed growth and threshold

- techniques,” in *2007 International Conference on Intelligent and Advanced Systems*, pp. 422–426, Kuala Lumpur, Malaysia, 2007.
- [6] M. R. Kaus, S. K. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, and R. Kikinis, “Automated segmentation of MR images of brain tumors,” *Radiology*, vol. 218, no. 2, pp. 586–591, 2001.
  - [7] S. Lakare and A. Kaufman, “3D segmentation techniques for medical volumes,” in *Center for Visual Computing, Department of Computer Science*, State University of New York, 2000.
  - [8] M. Letteboer, W. Niessen, P. Willems, E. B. Dam, and M. Viergever, “Interactive multi-scale watershed segmentation of tumors in MR brain images,” in *Proceedings of the IMIVA workshop of MICCAI*, pp. 1–6, Utrecht, Netherlands, 2001.
  - [9] E. Dam, M. Loog, and M. Letteboer, “Integrating automatic and interactive brain tumor segmentation,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*, pp. 790–793, Cambridge, 2004.
  - [10] J. E. Cates, R. T. Whitaker, and G. M. Jones, “Case study: an evaluation of user-assisted hierarchical watershed segmentation,” *Medical Image Analysis*, vol. 9, no. 6, pp. 566–578, 2005.
  - [11] M. C. Clark, L. O. Hall, D. B. Goldgof, R. Velthuisen, F. R. Murtagh, and M. S. Silbiger, “Automatic tumor segmentation using knowledge-based techniques,” *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 187–201, 1998.
  - [12] A. S. Capelle, O. Alata, C. Fernandez, S. Lefevre, and J. C. Ferrie, “Unsupervised segmentation for automatic detection of brain tumors in MRI,” in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, pp. 613–616, Vancouver, BC, Canada, 2000.
  - [13] P. Qian, Y. Chen, J.-W. Kuo et al., “mDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 819–832, 2020.
  - [14] P. Qian, Y. Jiang, Z. Deng et al., “Cluster prototypes and fuzzy memberships jointly leveraged cross-domain maximum entropy clustering,” *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 181–193, 2016.
  - [15] B. Yegnanarayana, *Artificial Neural Networks*, PHI Learning Pvt. Ltd., 2009.
  - [16] J. Liu, H. W. Jung, A. Dubra, and J. Tam, “Cone photoreceptor cell segmentation and diameter measurement on adaptive optics images using circularly constrained active contour model,” *Investigative Ophthalmology & Visual Science*, vol. 59, no. 11, pp. 4639–4652, 2018.
  - [17] F. Gibou, R. Fedkiw, and S. Osher, “A review of level-set methods and some recent applications,” *Journal of Computational Physics*, vol. 353, pp. 82–109, 2018.
  - [18] Mahalakshmi, H. K. Krishnappa, and D. Jayadevappa, “Automated brain tumor segmentation and identification using MR images,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 2647–2651, 2019.
  - [19] B. Fischl, D. H. Salat, E. Busa et al., “Whole brain segmentation: automated labeling of neuroanatomical structure in the human brain,” *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.
  - [20] K. M. Pohl, J. Fisher, J. J. Levitt et al., “A unifying approach to registration, segmentation, and intensity correction,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, J. S. Duncan and G. Gerig, Eds., pp. 310–318, Springer, 2005.
  - [21] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, “A brain tumour segmentation framework based on outlier detection,” *Medical Image Analysis*, vol. 8, no. 3, pp. 275–283, 2004.
  - [22] D. T. Gering, W. E. L. Grimson, and R. Kikinis, “Recognizing deviations from normalcy for brain tumor segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2002*, T. Dohi and R. Kikinis, Eds., pp. 388–395, Springer, 2002.
  - [23] M. B. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. G. Villemeure, and J. P. Thiran, “Atlas-based segmentation of pathological MR brain images using a model of lesion growth,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1301–1314, 2004.
  - [24] A. Gooya, K. M. Pohl, M. Bilello et al., “GLISTR: glioma image segmentation and registration,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1941–1954, 2012.
  - [25] D. Cobzas, N. Birkbeck, M. Schmidt, M. Jagersand, and A. Murtha, “3D variational brain tumor segmentation using a high dimensional feature set,” in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, Rio de Janeiro, 2007.
  - [26] S. Ho, E. Bullitt, and G. Gerig, “Level-set evolution with region competition: automatic 3-D segmentation of brain tumors,” in *Object recognition supported by user interaction for service robots*, pp. 532–535, Quebec City, Quebec, Canada, 2002.
  - [27] C. H. Lee, S. Wang, A. Murtha, M. R. G. Brown, and R. Greiner, “Segmenting brain tumors using pseudo-conditional random fields,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, Eds., pp. 359–366, 2008.
  - [28] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi, and N. Ayache, “Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images,” *Neuro Image*, vol. 57, no. 2, pp. 378–390, 2011.
  - [29] D. Zikic, B. Glocker, E. Konukoglu et al., “Decision forests for tissue-specific segmentation of high grade gliomas in multi-channel MR,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., pp. 369–376, Springer, 2012.
  - [30] D. Zikic, B. Glocker, E. Konukoglu et al., “Context-sensitive classification forests for segmentation of brain tumor tissues,” in *Proc MICCAI-BRATS*, pp. 1–8, Nice, France, 2012.
  - [31] B. H. Menze, E. Geremia, N. Ayache, and G. Székely, “Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation,” in *Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge)*, pp. 22–30, 2012.
  - [32] B. H. Menze, A. Jakab, S. Bauer et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
  - [33] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, “A survey of MRI-based medical image analysis for brain tumor studies,” *Physics in Medicine & Biology*, vol. 58, no. 13, pp. R97–R129, 2013.
  - [34] E. Ehsan and V. René, “Sparse subspace clustering: algorithm, theory, and applications,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

- [35] M. Angulakshmi and G. G. Lakshmi Priya, “Walsh Hadamard transform for simple linear iterative clustering (SLIC) super-pixel based spectral clustering of multimodal MRI brain tumor segmentation,” *IRBM*, vol. 40, no. 5, pp. 253–262, 2019.
- [36] A. Radhakrishna, S. Appu, S. Kevin, L. Aurélien, F. Pascal, and S. Sabine, *SLIC superpixels*, Epfl, 2010.



## Research Article

# Factors Influencing Information Service Quality of China Hospital: The Case Study of since 2017 of a Hospital Information Platform in China

Lei Jiao <sup>1</sup>, HuaPing Xiao <sup>2</sup>, XiaoZhuo Zhu <sup>3</sup>, and Xu Zhao <sup>4</sup>

<sup>1</sup>Yale School of Medicine, Yale University, New Haven, CT 06510, USA

<sup>2</sup>Department of Anesthesiology, Jiangxi Cancer Hospital, Nanchang, Jiang Xi 330029, China

<sup>3</sup>College of Public Service and Management, Ningbo College of Health Sciences, Ningbo, Zhejiang 315100, China

<sup>4</sup>Department of Anesthesiology, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China

Correspondence should be addressed to Lei Jiao; [jiaoleiyale@163.com](mailto:jiaoleiyale@163.com)

Received 5 April 2020; Revised 19 May 2020; Accepted 3 June 2020; Published 1 July 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Lei Jiao 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:** As a country with the largest number of netizens around the world, China enjoys improving social information services based on the Internet. With such a large quantity of network users, it is inevitable for China's hospitals at various levels to provide patients and the public with information services by setting up their own official websites. But it is still elusive for the factors affecting the information service quality of China Hospital. **Objective:** Identifies the factors affecting the information service quality of the case of the Online website of a hospital in China. adding new content to the research fruits in this field. The research can effectively enhance the efficiency of hospital resource utilization, allocating limited resources to most efficient areas and leveling up the information service quality of hospitals to the largest extent. This ultimately improves patient satisfaction. **Method:** This research investigates the factors affecting the information service quality of the case of a Chinese hospital online website and by means of Delphi method, statistical analysis, and other research methods, formulates the Evaluation Indicator System for the Information Service Quality of the case of the Online website of a hospital in China. The research applies this system to the empirical research on the information service quality of the hospital's website and then makes a comparative analysis between the research results and traffic data of the websites of other hospitals over the same period. **Results:** By means of the Bivariate Correlation, the author carried out a correlation analysis of the comprehensive evaluations of the information service of the Online website of a hospital in China and the traffic data of the Online website of a hospital in China, including the total traffic, PV and UV. For details of the analysis results, indicates that the correlation coefficient among the three objects is 1, a significant correlation. It also suggests that the comprehensive evaluations of the information service of A Chinese hospital website and the traffic of A Chinese hospital website are positively correlated. The information service quality of China Hospital website is an important component of the hospital's overall service quality. This research on the information service quality of China Hospital website covers the website's service functions, service quality, resources and the front-end and back-end technology systems. **Discussion and conclusion:** In the case that the China Hospital information service function is still not perfect, perfecting the functions of China Hospital website plays a decisive role in improving the information service quality of the hospital. In addition, it can be inferred that after the information service function of China Hospital website is improved or the evaluation of the functional quality attribute of website information service scores higher, the supporting attribute of website information service will be the next key task for the hospital in enhancing its information service quality is the improvement ratio of Functional quality attribute of website information service and the supporting attribute of website information service tend to be the same., And even the improvement rate of the supporting attribute is sometimes higher than the improvement rate of Functional quality. So The construction of a model of the comprehensive evaluation system on the information service of has pointed out a new direction in China's research in this field, This model is both of high theoretical value and practical value.

## 1. Introduction

According to statistics of China Internet Network Information Center (CNNIC), the number of Chinese netizens, including Internet users on mobile phones, has reached 1.701 billion, which is the largest number in the world, and this number has been increasing year by years [1]. Therefore, the Internet has already been the main approach for the public to acquire information in China. In the meanwhile, medical institutions that provide services for the public are also carrying through the informatization reform. The hospital information platforms thus become the main representative in this reform. Hospitals usually provide more than just medical treatments to patients. They also offer other services including information service. Among them, using hospital information platform is one of the important approaches for a hospital to provide information service to patients. After ten years of informatization, as of 2019, the average rate of third-level hospitals launching information platforms in each province of China has reached 82% [2]. China Hospitals usually provide more than just medical treatments to patients. They also offer other services including information service, consultation service, and psychological service. Among them, using hospital information platform is one of the important approaches for a hospital to provide information service to patients.

In China's mainland, when patients have medical needs and choose hospitals online, they tend to choose a hospital that has high quality information service on the hospital website. They gain information service on the website and proceed with subsequent medical service accordingly. They usually recommend the website providing good experience of online information service to their families or friends in need of medical service. It can be seen that an advanced hospital information platform with high quality information service can not only bring profits to the hospital but also establish the hospital reputation, and more importantly patients can enjoy high quality information service. Therefore, it is of importance to study factors influencing information service quality of hospital information platforms so as to improve their information service quality.

Research on information service quality of hospital information platforms is firstly carried out on information service quality of hospital information system, such as the measurement of effectiveness of the information system and the study on SERVQUAL scale that is used to measure the service quality of information system [3, 4]. Then, with the popularity of computers and networks as well as the increase in web browsing, the academic community pays more attention to the research on the service quality of the information system of webpages since 2000. Tan in 2003 studied the service quality frame of the information system [5]. Pinho in 2011 studied the influences of network service quality on network usage [6]. Baldwin in 2014 studied the relationship between the rating of patients' satisfaction and the quality of hospital information platforms and measures [7]. However, China's research on information service quality of hospital information platforms has just started. The major contributions have been made mainly by some scholars in Taiwan and coastal

universities in China's mainland. For example, Liang in 2009 conducted a research on the influences of the quality of information platforms on the performance of an organization's relationship with customers [8]. Pei and Wang in 2009 researched on the evaluation system of information service quality of user-oriented information platforms [9]. Zhu and Xiong in 2012 studied the evaluation criteria on information service quality of medical information platforms [10]. This article shows investigating and studying the information platform of a Hospital Information Platform from the perspective of management, the thesis intends to identify factors influencing the information service quality of hospital information platforms, so as to figure out factors affecting the information service quality of the information platform of a Hospital Information Platform in China and the interaction among these factors.

## 2. Methods

This study aimed to understand Factors Influencing Information Service Quality of the Information Platform of China Hospital. To achieve this goal a case study and Delphi and Questionnaire methodology were followed. This study will discuss information service quality of hospital information platforms in two aspects—the essence of the information service of hospital information platforms and perception of information service. Then an evaluation system will be established and verified.

To begin with, following the case study method, this study's object is set within a certain range in order to perform in-depth and comprehensive analysis through dynamic follow-up investigation into them. This study selects the a Hospital Information Platform in China as the object of the case study. Through study design, data collection, data analysis, and study report draft, factors influencing information are sought out as reference for the improvement of service quality of hospital information platforms in China.

Second, Use Delphi and Questionnaire method. In this study Use Delphi, experts in this field are consulted to discuss how to determine or quantify indicators of various dimensions, indicator weights of medical information service quality of information platforms in Chinese hospitals so as to help finish the study [11, 12]. In this study Use Questionnaire, In this Questionnaire method, the subjects are given brief forms and asked to offer suggestions or file complaints on certain issues so as to collect information. Based on literature research and experts' views, a questionnaire is designed for the front-line hospital staff and patients. The questionnaire consists of hospital staff investigation, patient investigation and expert investigation. Every investigation will be carried out two times, and there are two rounds for all the investigations. The first investigation is on the dimension structure of evaluation system for medical information service function and quality of a Hospital Information Platform in China, and the second one is on evaluation system index coefficient of medical information service function and quality of a Hospital Information Platform in China. At last, Statistical Analysis. In this study, Excel is used for data-entry of information in the expert consultation, and SPSS is used for

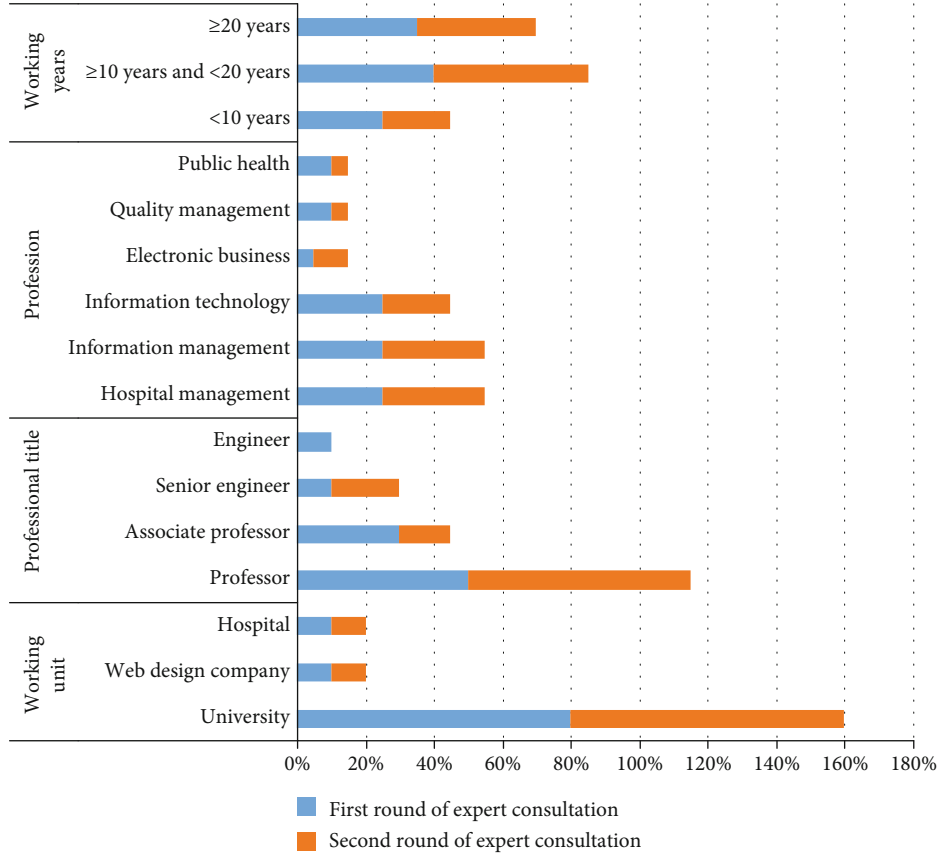


FIGURE 1: Statistics of Status of Consultation Experts.

statistical analysis of data in the questionnaire so as to examine the consistency and logic of the data and to ensure the correctness of the data.

### 3. Comprehensive Evaluation System for Medical Information Service of a Hospital Information Platform in China

**3.1. The Structure of System Evaluating the Information Service.** This study has randomly chosen 101 experts across China to put into the expert bank of consultation object, ranging from professors in management or other disciplines with deputy senior or above titles to information technology experts at hospitals and professional technicians at information technology web design companies. Next, send the initial dimension indicator set to about 20 experts via mobile phones, letters and emails for at least two rounds of consultation to settle the names, categories, quantity and structure of the evaluation system's dimension indicators. Each round of consultation randomly chooses experts among the 101 consultation objects to send consultation forms, See Figure 1 for details.

The expert difference coefficient  $Q$  indicates the difference of expert opinion. This variable coefficient demonstrates different degree of expert recognition of the evaluated indicators. The smaller the variable coefficient is, the less the differ-

ence of expert opinion is. Calculation equation of variable coefficient:

$$Q_i = \frac{s_i}{v_i} \quad (1)$$

In the equation,  $Q_i$  refers to the variable coefficient of indicator  $i$ ,  $s_i$  standard deviation and  $v_i$  weighted arithmetic mean. In the two rounds of expert consultations, 45 consultation forms were sent out, and 40 valid feedbacks were collected. The effectiveness of the two rounds of consultations was 86%, both of which were higher than 60%. 18 dimensional indicators' difference coefficients in the first round are over 0.25, and all these 18 dimensions have been adjusted accordingly; while the difference coefficients of 73 dimensional indicators are all between 0.08-0.24 (less than 0.25). In the second round, only one dimension's difference coefficient is bigger than 0.25, which has been modified, and the difference coefficients of 72 dimensional indicators are all between 0-0.24 (less than 0.25).

In summary, The expert authority degree is 0.846. In the Delphi Method, when the expert authority degree  $> 0.7$  it can be seen as relatively high. Hence, the expert authority degree in this round of expert consultation is quite good. Experts have given positive responses to the dimension structure of the indicator system in this thesis quite unanimously, which demonstrates that both the first and second round of

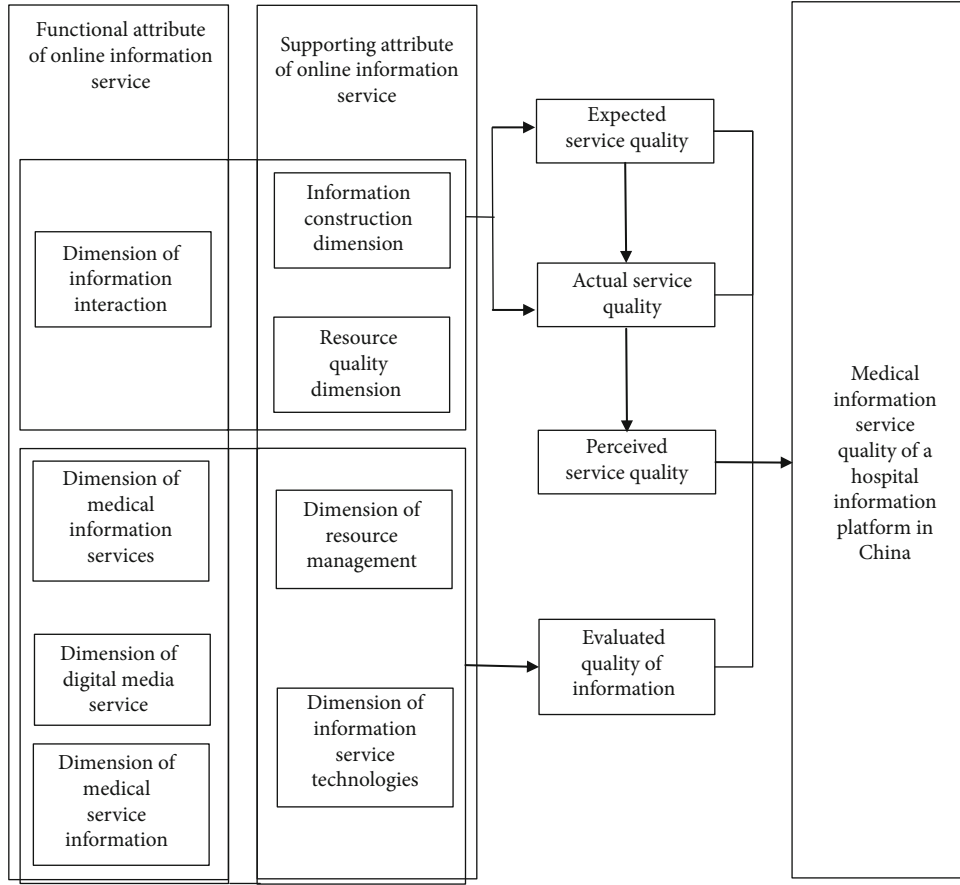


FIGURE 2: Theoretical Research Platform of Information Service Quality of a Hospital Information Platform in China (Formal Version).

research have been recognized by consultation experts as scientific and rational. The attributes, dimensions and indicators of the information service evaluation system of a Hospital Information Platform in China have been identified. According to Their respective arithmetic averages are less than 3, which does not reach the predetermined indicator of consistency. After two rounds of questionnaires, the old indicators are eliminated to construct new indicators. Based on the above indicators, this thesis has developed it into Figures 2 and 3 the formal version.

**3.2. System Construction of Information Service Quality Indicators and Weight Determination.** Through the Delphi method, this thesis conducts two rounds of expert consultation and discussion according to the importance of the dimensional indicators of information service evaluation system of a Hospital Information Platform in China. The consultation form is basically the same as the one constructed in the information service evaluation system in the hospital information platform. The difference lies in that the consulting content is changed into the 4 initial weight schemes and the experts can either choose a scheme from the above four schemes or propose their own ones. If the experts do not have a unified opinion after the first round of consultation, the author can merge and sort out the four schemes and the schemes put forward by experts in the first round.

Through the first and second rounds of expert consultation, the final weight distribution scheme is determined. The combined weight values of secondary and tertiary indicators are obtained through multiplication. The combined weight values of the secondary indicators are obtained by Equation (2), while those of tertiary indicators are got by Equation (3).

$$S_{wi} = S_{1i} \times S_{2i} \quad (2)$$

$S_{wi}$  is the combined weight value of the secondary indicators,  $S_{1i}$  refers to the weight of primary indicator corresponding to indicator  $i$ , and  $S_{2i}$  the weight of secondary indicator corresponding to indicator  $i$ . See Table 1.

$$S_{wi} = S_{1i} \times S_{2i} \times S_{3i} \quad (3)$$

$S_{wi}$  is the combined weight value of the tertiary indicators,  $S_{1i}$  refers to the weight of primary indicator corresponding to indicator  $i$ ,  $S_{2i}$  the weight of secondary indicator corresponding to indicator  $i$ , and  $S_{3i}$  the weight of tertiary indicator corresponding to indicator  $i$ . The weighting allocation interval is  $[0,1]$ , See Table 1.

**3.3. Explanation of Evaluation Indicator Assignment.** After determining the structure and weight of the comprehensive evaluation system, to improve it and strengthen its

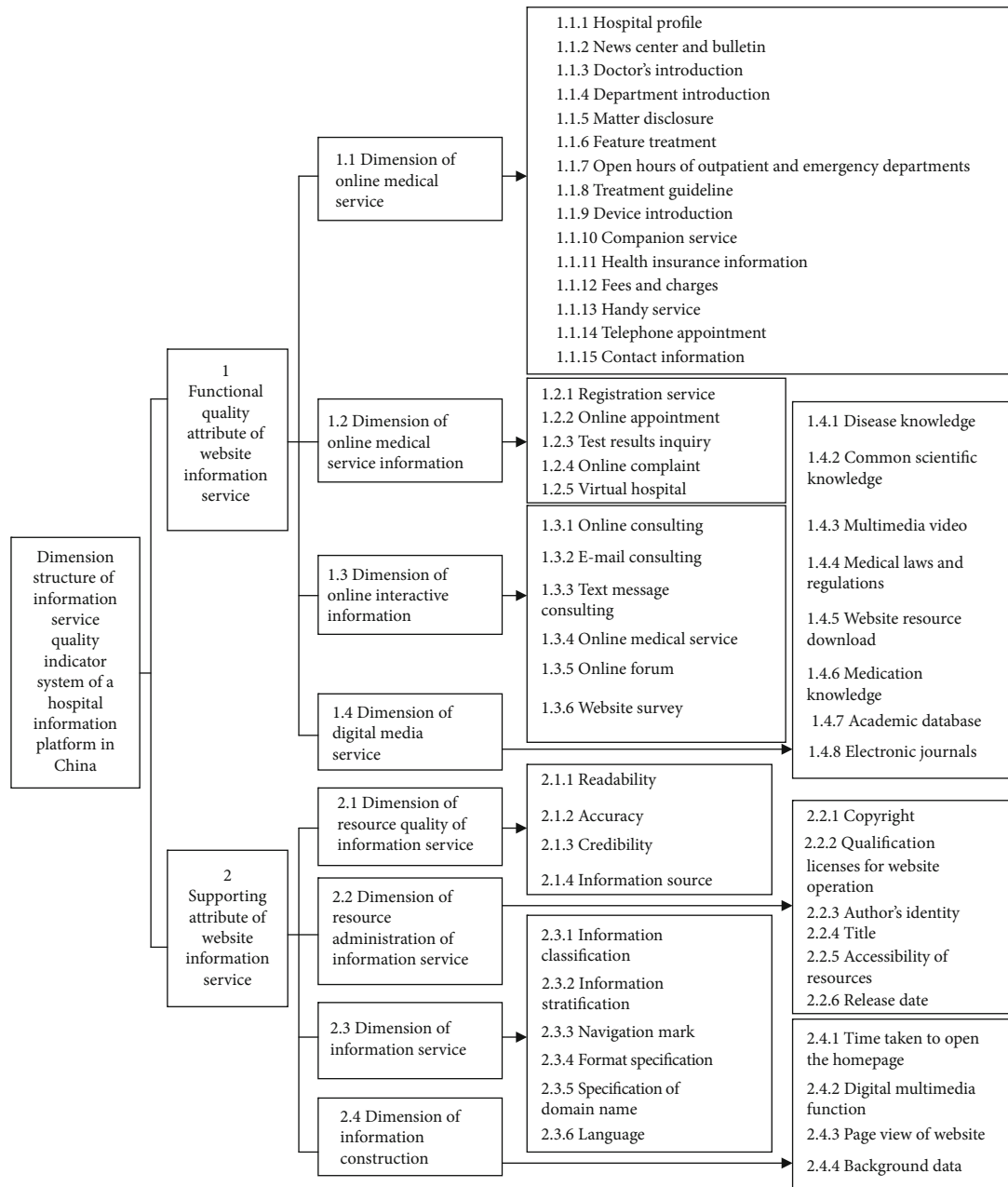


FIGURE 3: Dimension Structure of Information Service Quality Indicator System of a Hospital Information Platform in China (Final Version).

operability, there is still need to take different quantitative assignment methods in accordance with the characteristics of each indicator. Since the assignment interval of all the indicators in the comprehensive evaluation system is set within  $[0, 1]$ . According to the characteristics of each indicator, different grades and quantified assignments are divided, and the quantitative scores of all the indicators are calculated. For details about the quantitative assignment of all standards in the comprehensive evaluation system, see Table 2.

**3.4. Indicator Quantification of Evaluation System and Assignment Standard and Construction of Comprehensive Evaluation Model.** By weighting the dimensions of the system and assigning values in a quantitative way, the author

establishes a comprehensive evaluation model for the information service of the a Hospital Information Platform in China according to the weight of its information service and the quantitative value assignment standards. For the total weighted score of the comprehensive indicators, see Equation (4).

$$R = \sum_{i=1}^m \sum_{j=1}^{n_i} W_{ij} R_{ij} \times 100 \quad (4)$$

$R$  representing the total score is in the range between 0 and 100,  $R_{ij}$  is the value of the  $j$  indicator in the  $i$  dimension, and  $W_{ij}$  is the combined weight of the  $j$  indicator in the  $i$



TABLE 1: Indicator Combined Weight Values of Information Service Evaluation System of a Hospital Information Platform in China.

(1) distribution scheme of primary indicators			
No.	Evaluation Indicator	Scheme 4	Weight value
1	Functional quality attribute of website information service	70	0.7
2	Supporting attribute of website information service	30	0.3
Total			1
(2) distribution scheme of secondary indicators			
No.	Evaluation Indicator	Scheme 3 Original first-round my scheme 2)	Combined weight value
1.1	Dimension of online medical information services	35	0.245
1.2	Dimension of online medical service information	30	0.21
1.3	Dimension of online interactive information	20	0.14
1.4	Dimension of digital media service information	15	0.105
No.	Evaluation Indicator	Scheme 2	Combined weight value
2.1	Dimension of resource quality of information service	40	0.12
2.2	Dimension of resource administration of information service	30	0.09
2.3	Dimension of information service technology	15	0.045
2.4	Dimension of information construction	15	0.045
Total			1
(3) distribution scheme of tertiary indicators			
No.	Evaluation Indicator	Scheme 1	Combined weight value
1.1.1	Hospital profile	10	0.0245
1.1.2	News and bulletin	7	0.01715
1.1.3	Introduction of hospital experts	7	0.01715
1.1.4	Introduction of departments	5	0.01225
1.1.5	Publicity of medical events	5	0.01225
1.1.6	Special medical treatment	5	0.01225
1.1.7	Emergency service and outpatient service hour	10	0.0245
1.1.8	Treatment guidelines	10	0.0245
1.1.9	Device introduction	10	0.0245
1.1.10	Accompanying services	4	0.0098
1.1.11	Medical insurance information	5	0.01225
1.1.12	Charging standard notes	5	0.01225
1.1.13	Handy Services	5	0.01225
1.1.14	Appointments and registration by phone	6	0.0147
1.1.15	Contact information	6	0.0147
No.	Evaluation Indicator	Scheme 4	Combined weight value
1.2.1	Signing in service	25	0.0525
1.2.2	Online appointment and registering	35	0.0735
1.2.3	Examination results query	25	0.0525
1.2.4	Online complaint	10	0.021
1.2.5	Virtual hospital	5	0.0105
No.	Evaluation Indicator	Scheme 4	Combined weight value
1.3.1	Online advisory	30	0.042
1.3.2	E-mail advisory	10	0.014
1.3.3	Text message advisory	10	0.014
1.3.4	Online medical service	30	0.042
1.3.5	BBS	15	0.021
1.3.6	Online message board	5	0.007
No.	Evaluation Indicator	Scheme 2	Combined weight value

TABLE 1: Continued.

1.4.1	Disease knowledge	15	0.01575
1.4.2	Popular science	15	0.01575
1.4.3	Multimedia video	7.5	0.007875
1.4.4	Medical policy	15	0.01575
1.4.5	Medical education	7.5	0.007875
1.4.6	Drug usage	15	0.01575
1.4.7	Academic database	12.5	0.013125
1.4.8	Hospital E-journal	12.5	0.013125
No.	Evaluation Indicator	Scheme 2	Combined weight value
2.1.1	Readability	20	0.024
2.1.2	Accuracy	30	0.036
2.1.3	Completeness	20	0.024
2.1.4	Information source	30	0.036
No.	Evaluation Indicator	Scheme 4	Combined weight value
2.2.1	Copyright	35	0.0315
2.2.2	Qualification licenses for website operation	25	0.0225
2.2.3	Author's identity	10	0.009
2.2.4	Title marking	10	0.009
2.2.5	Accessibility of resource	10	0.009
2.2.6	Release date	10	0.009
No.	Evaluation Indicator	Scheme 3	Combined weight value
2.3.1	Information classification	25	0.01125
2.3.2	Information stratification	10	0.0045
2.3.3	Navigation mark	20	0.009
2.3.4	Format specification	20	0.009
2.3.5	Specification of domain name	10	0.0045
2.3.6	Language	15	0.00675
No.	Evaluation Indicator	Scheme 4	Combined weight value
2.4.1	Time taken to open the homepage	30	0.0135
2.4.2	Digital multimedia function	20	0.009
2.4.3	Website page view	20	0.009
2.4.4	Background data	30	0.0135
Total			1

dimension.  $i$  is 1 to  $m$ ,  $j$  is 1 to  $n_i$ ,  $n_i$  is the number of indicators covered by the  $i$  dimension,  $m$  is the number of the dimensions and equals 8, and  $S$  is the number of indicators in the tertiary level and equals 54.

The author first collects data by using the Hospital Website Information Service Evaluation Form, then processes the data, assigns values in a quantitative way according to the evaluation of indicators at each level, and in the end, uses the equation of the comprehensive evaluation model to calculate the comprehensive score of the information service of the a Hospital Information Platform in China.

#### 4. Empirical Research Results

Since the website of a Hospital Information Platform in China is the only channel for the hospital's information platform to deliver information services, the author sets the website of a Hospital Information Platform in China as the

object of the empirical research. Established on June 19, 2013, the website of a Hospital Information Platform is redesigning its overall pages on June 19, 2017, as the empirical research goes on. Therefore, the author sets both the pre-revised website and the post-revised website as the object of the empirical research. Formulated on the basis of the Comprehensive Evaluation Model for the Information Service of a Hospital Information Platform Information Platform, the a Hospital Information Platform Website Information Service Evaluation Form is an online data collection form for the evaluation of hospitals' website information service.

In the empirical research, December, 2016 to January, 2020, 80 participants falling of Internet Explorer 11 as the uniform browser and logged in a Hospital Information Platform's website in ten periods. Stratified random sampling is adopted to select participants, and in accordance with the sampling indicators, these participants are classified into four groups: out-patients, in -patients, hospital employees,

TABLE 2: The Quantification and Assignment Standards of the Dimensions of Online Medical Information Services.

No.	Evaluation standards	Description of quantitative assignment
Quantitative assignment of the dimensions of online medical information services		
1.1.1	Hospital profile	<p>The hospital profile item includes ten sub-items: (introduction, logo, history, leadership, culture, qualification, honor, environment, words by the president and video of the hospital.) each of these items has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>“n” is the number of the sub-items and equals 10.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
1.1.2	News and bulletin	<p>The news and announcement include two sub-items: News and announcement, respectively, valued 0.7 and 0.3. According to the updating frequency, the sub-item news falls into four categories, no updating, updating every one month and over, updating every week and updating every two days, which are valued, respectively, 0, 0.34, 0.67 and 1. The sub-item announcement has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $0.7 \times W_1 + 0.3 \times W_2$ <p><math>W_1</math> is the value of news and <math>W_2</math> is that of announcement.</p>
1.1.3	Introduction of hospital experts	<p>The introduction of hospital experts includes five sub-items: (personal introduction, expertise, honor and academic achievements, positions and contacts.) this item has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>“n” is the number of the sub-items and equals 5.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
1.1.4	Introduction of departments	<p>The item introduction of departments includes five sub-items: (departments introduction, expertise, medical teams, honor and contacts.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>“n” is the number of the sub-items and equals 5.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
1.1.5	Publicity of medical events	<p>The publicity of medical events includes five sub-items: (clinical service management, nursing management, hospital regulations, academic research and medical education.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>“n” is the number of the sub-items and equals 5.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
1.1.6	Special medical treatment	<p>The special medical treatment includes three sub-items: Programs, introduction and effects. It has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>“n” is the number of the sub-items and equals 3.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
1.1.7	Emergency service and outpatient service hour	<p>The emergency service and outpatient service hour include two sub-items: Time and doctors. It has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>“n” is the number of the sub-items and equals 2.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
1.1.8	Treatment guidelines	<p>The treatment guidelines include eight sub-items: (transportation guides, hospital indoor navigation, outpatient information, registration information, emergency information, medical</p>

TABLE 2: Continued.

No.	Evaluation standards	Description of quantitative assignment
		test information, inpatient treatment and discharge guidelines.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is: $1/n \sum_{i=1}^n W_i$ “n” is the number of the sub-items and equals 8. $W_i$ is the specific value of the $i$ sub-item.
1.1.9	Device introduction	The device introduction includes five sub-items: (types, names, functions, distribution and numbers.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is: $1/n \sum_{i=1}^n W_i$ “n” is the number of the sub-items and equals 5. $W_i$ is the specific value of the $i$ sub-item.
1.1.10	Accompanying service	The accompanying service includes five sub-items: (home care, inpatient care, postpartum care, pregnancy care and others.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is: $1/n \sum_{i=1}^n W_i$ “n” is the number of the sub-items and equals 5. $W_i$ is the specific value of the $i$ sub-item.
1.1.11	Medical insurance information	The medical insurance information includes four sub-items: (news, policies, guidelines, and organizations.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is: $1/n \sum_{i=1}^n W_i$ “n” is the number of the sub-items and equals 4. $W_i$ is the specific value of the $i$ sub-item.
1.1.12	Charging standard notes	The charging standard notes include six sub-items: (outpatient, emergency and comprehensive medical services, traditional Chinese and folk medicine, medical technologies, clinical treatment, special medical treatment services and pharmaceutical drugs.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is: $1/n \sum_{i=1}^n W_i$ “n” is the number of the sub-items and equals 6. $W_i$ is the specific value of the $i$ sub-item.
1.1.13	Handy Service	The item Handy Services includes ten sub-items: (guiding services, drinking water, body temperature measurement, Presbyopic glasses, wheelchairs and stretchers, medical record copying, free magazines, parking guides, testing report post service and others.) it has two levels: “Established” valued 1 and “to be established” valued 0. The calculation of the value of the quantitative standard is: $1/n \sum_{i=1}^n W_i$ “n” is the number of the sub-items and equals 10. $W_i$ is the specific value of the $i$ sub-item.
1.1.14	Appointments and registration by phone	According to its functions, the item appointments and registration by phone has four levels: (1) no telephone appointment service is provided on the website, which is valued 0 (2) the website offers telephone number for appointment but it is not accessible. The value is 0.34 (3) the website has accessible telephone number but users cannot make appointment through it. The value is 0.67 (4) the website provides telephone appointment service and users can make an appointment successfully, which is valued 1 The state that the website provides telephone appointment service and users can make an appointment successfully is: When researchers acting as patients make an appointment over the telephone, operators answer the phone promptly, make a registration in appropriate departments for the patients according to the symptoms described by the researchers, record the patients’ information and thus finish all the registration procedures.

TABLE 2: Continued.

No.	Evaluation standards	Description of quantitative assignment
1.1.15	Contact information	<p>The contact information includes ten sub-items for contacting the hospitals: (hospital switchboard, phone numbers of departments, discipline inspection phone number, E-mail address, hospital address, location, traffic, digital map, President's mail-box and others.) it has two levels: "Established" valued 1 and "to be established" valued 0. The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>"n" is the number of the sub-items and equals 10.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
Online medical service information		
1.2.1	Signing in service	<p>According to its functions, the signing in service has four levels:            (1) the hospital website does not provide registration service. The value is 0            (2) the website provides registration service but users cannot finish the registration. The value is 0.34            (3) the website has registration service and users can register successfully. The value is 0.67            (4) the website has registration service and users can register successfully and use the registered accounts to use website's functions. The value is 1</p>
1.2.2	Online appointments and registering	<p>The online appointments and registering has two levels: "Established" valued 1 and "to be established" valued 0.</p>
1.2.3	Examination results query	<p>According to patients' names and medical service numbers, the examination results query has two levels: "Established" valued 1 and "to be established" valued 0.</p>
1.2.4	Online complaint	<p>The online complaint item has two levels: "Established" valued 1 and "to be established" valued 0.</p>
1.2.5	Virtual hospital	<p>The virtual hospital includes six sub-items:            (medical service appointment, guidelines, specialist clinic, hospital indoor navigation, online test and 3D virtual hospital.) it has two levels: "Established" valued 1 and "to be established" valued 0.            The calculation of the value of the quantitative standard is:</p> $1/n \sum_{i=1}^n W_i$ <p>"n" is the number of the sub-items and equals 6.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.</p>
Online interactive information		
1.3.1	Online advisory	Standards of this kind include two parts:
1.3.2	E-mail advisory	<p>Service functions valued 0.6 and service quality valued 0.4. The service functions has two levels: "Established" valued 1 and "to be established" valued 0. In terms of the value of service functions, researchers acting as users send consultation to designated e-mail address or phone number.</p>
1.3.3	Text message advisory	<p>According to the time period before the consultation is answered and the accuracy of the answers, the service function has four levels: Half a month and over, valued 0, in half a month, valued 0.34, in a week, valued 0.67, and in three days, valued 1. The calculation of the value of the quantitative standard is:</p> $0.6 \times W_1 + 0.4 \times W_2$ <p><math>W_1</math> is the value of the service functions and <math>W_2</math> is that of the service quality.</p>
1.3.4	Online medical service	<p>According to its functions, the online medical service has three levels:            (1) researchers acting as users try the medical services provided on the hospital website to assign the value of the item. If the website does not provide online medical services, the value is 0            (2) if the website provides online medical services but users cannot register and further access these services, the value is 0.5            (3) if the website provides online medical services and these services are accessible for users, the value is 1</p>
1.3.5	BBS	<p>The BBS item has four levels:            (1) if the website does not provide the function of online forum, the value is 0            (2) if the website provides this function and users can register and log in to use it, the value is 0.34            (3) if the website has an online forum and users can register, log in and search for information by categories, but the number of posts is relatively small (less than 100), the value is 0.67            (4) if the website has an online forum with various functions, users can register, log in and search for information by categories, and the number of posts is relatively large (more than 100), the value is 1</p>



TABLE 2: Continued.

No.	Evaluation standards	Description of quantitative assignment
1.3.6	Online message board	This item has two levels: “Established” valued 1 and “to be established” valued 0.
Digital media service information		
1.4.1	Disease knowledge	<p>According to the amount of its contents, the disease knowledge has four levels:</p> <p>(1) if the website does not provide this function, the value is 0</p> <p>(2) if the website has this function and the number of knowledge items is no more than 50, the value is 0.34</p> <p>(3) if the website provides knowledge about diseases and categorizes these knowledge, and the number of knowledge items is between 50 and 100 (100 included), the value is 0.67</p> <p>(4) if the website provides knowledge about diseases and categorizes these knowledge, and the number of knowledge items is over 100, the value is 1</p>
1.4.2	Popular science	<p>According to the amount of its contents, the popular science has four levels:</p> <p>(1) if the website does not provide this function, the value is 0</p> <p>(2) if the website has this function and the number of knowledge items is no more than 50, the value is 0.34</p> <p>(3) if the website provides knowledge about science and categorizes these knowledge, and the number of knowledge items is between 50 (50 included) and 100, the value is 0.67</p> <p>(4) if the website provides knowledge about science and categorizes these knowledge, and the number of knowledge items is over 100 (100 included), the value is 1</p>
1.4.3	Multimedia video	<p>According to the amount of its contents, the multimedia video has four levels:</p> <p>(1) if the website does not provide this function, the value is 0</p> <p>(2) if the website has this function and the number of videos is less than 10, the value is 0.34</p> <p>(3) if the website provides multimedia videos and the number of videos is between 10 (10 included) and 50, the value is 0.67</p> <p>(4) if the website provides multimedia videos and categorizes videos, and the number of videos is 50 and over, the value is 1</p>
1.4.4	Medical policy	The medical policy has two levels: “Established” valued 1 and “to be established” valued 0.
1.4.5	Medical education	<p>According to the amount of its contents, the medical education has four levels:</p> <p>(1) if the website does not provide this function, the value is 0</p> <p>(2) if the website has this function and the number of resources for downloading is no more than 10 (10 included), the value is 0.34</p> <p>(3) if the website provides resources for downloading and the number is over 10, the value is 0.67</p> <p>(4) if the website provides resources for downloading and categorized these resources, and the number is over 50, the value is 1</p>
1.4.6	Drug usage	The drug usage has two levels: “Established” valued 1 and “to be established” valued 0.
1.4.7	Academic database	The academic database has three levels: An academic database with owned academic resources, an academic database and to be established, which valued, respectively, 1, 0.5 and 0.
1.4.8	Hospital E-journal	<p>According to its function and contents, the hospital E-journal has four levels:</p> <p>(1) if the website does not provide this function, the value is 0</p> <p>(2) if the website provides this function, and the number of e-journals is no more than 5 (5 included), the value is 0.34</p> <p>(3) if the website has this function, and the number of e-journals is between 5 and 10 (10 included), the value is 0.67</p> <p>(4) if the website has e-journals and provides the function of categorization and searching, and the number of the e-journals is over 10, the value is 1</p>
The resource quality of information service		
2.1.1	Readability	According to whether the contents on the hospital website are easy to be understood by users, readability has four levels valued, respectively, 0, 0.34, 0.67 and 1.
2.1.2	Accuracy	The accuracy concerns with the accuracy of information published on the website, pronunciation and grammar. According to this accuracy, the item has four levels valued, respectively, 0, 0.34, 0.67 and 1.
2.1.3	Completeness	The completeness is whether the information released by the website covers a sufficiently wide range of topics. According to this completeness, the item has four levels valued, respectively, 0, 0.34, 0.67 and 1.
2.1.4	Information source	<p>The information source has five sub-items:</p> <p>(source of the article, reprint mark, reprint review, online reference links and reminding,) it has two levels: “Established” valued 1 and “to be established” valued 0. The total score is 3. The</p>

TABLE 2: Continued.

No.	Evaluation standards	Description of quantitative assignment
		<p>calculation of the value of the quantitative standard is:  <math>1/3(\min(3, \sum_{i=1}^n W_i))</math>  “n” is the number of the sub-items and equals 5.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.  The “min” function is used to find the smallest value in the given parameter list.</p>
Information service resources management		
2.2.1	Copyright	<p>The copyright has two levels: “Established” valued 1 and “to be established” valued 0.</p> <p>The qualification licenses for website operation includes ten sub-items: (internet drug information service license, MOH health information service network management review, internet medical and health information service license, internet education information service license, internet domain name accreditation, internet publishing license, internet broadcasting program license, internet payment license, value added telecommunication business license and others.) it has two levels: “Established” valued 1 and “to be established” valued 0. The total score is 5. The calculation of the value of the quantitative standard is:  <math>1/5(\min(5, \sum_{i=1}^n W_i))</math>  “n” is the number of the sub-items and equals 10.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.  The “min” function is used to find the smallest value in the given parameter list.</p>
2.2.2	Qualification licenses for website operation	
2.2.3	Author’s identity	<p>The Author’s identity includes six sub-items: (name, education, title, organization, contact and others.) it has two levels: “Established” valued 1 and “to be established” valued 0. The total score is 4. The calculation of the value of the quantitative standard is:  <math>1/4(\min(4, \sum_{i=1}^n W_i))</math>  “n” is the number of the sub-items and equals 6.  <math>W_i</math> is the specific value of the <math>i</math> sub-item.  The “min” function is used to find the smallest value in the given parameter list.</p>
2.2.4	Title marking	<p>The title marking item has four levels:  (1) if the website does not mark any titles, the value is 0  (2) if the homepage or a certain web page has a marked title, the value is 0.34  (3) if the homepage and other pages all have marked titles, the value is 0.67  (4) if the homepage and other pages have different title marks, the value is 1  The level of the sub-item is figured out through the random examination of 20 page titles of the hospital’s website.</p>
2.2.5	Accessibility of resources	<p>According to the number of the website’s invalid links, the accessibility of resources has four levels:  (1) if the number of invalid links is no less than 10 (10 included), the value is 0  (2) if the number of invalid links is between 5 and 10, the value is 0.34  (3) if the number of invalid links is less than 5 (5 included), the value is 0.67  (4) if the website has no invalid links, the value is 1  The number of invalid links is figured out through the random examination of 20 internal links of the website.</p>
2.2.6	Release date	<p>The release date has four levels:  (1) if the publishing time or date is not marked, the value is 0  (2) if the time or date is marked, the value is 0.5  (3) if the time and date are both marked, the value is 1</p>
Information service technology		
2.3.1	Information classification	<p>The information classification includes five sub-items:  (1) the layout of pages on the website  (2) the homepage has information classification  (3) the information classification is reasonable  (4) the information classification meets users’ demands  (5) the names of classified information are easy to understand  The above five sub-items have four levels: Excellent, good, general and bad, valued, respectively, 0, 0.34, 0.67 and 1. The calculation of the value of this item is:</p>

TABLE 2: Continued.

No.	Evaluation standards	Description of quantitative assignment
2.3.2	Information stratification	$0.2 \times (W_1 + W_2 + W_3 + W_4 + W_5)$ $W_1, W_2, W_3, W_4$ And $W_5$ are, respectively, the values of the five sub-items. According to the number of the grades of information classification, the item has four levels: (1) if the number of the grades is no less than 10, the value is 0 (2) if the number of the grades is between 5 and 10, the value is 0.34 (3) if the number of the grades is between 4 (4 included) and 5 (5 included), the value is 0.67 (4) if the number of the grades is less than 4, the value is 1  The navigation mark includes five items: (A back link to the last page, a back link to the homepage, a back link to the information classification page, a link to homepage prompt message and a navigation map for the website.) it has two levels: "Established" valued 1 and "to be established" valued 0. The total score is 3. The calculation of the value of the quantitative standard is:
2.3.3	Navigation mark	$1/3(\min(3, \sum_{i=1}^n W_i))$ "n" is the number of the sub-items and equals 5. $W_i$ is the specific value of the $i$ sub-item. The level of the sub-item is figured out through the random examination of 20 page titles of the hospital's website. The "min" function is used to find the smallest value in the given parameter list. The format specification includes eight sub-items: (well-designed pages, user-friendly font size, uniform page layout, appropriate page margin, pagination on long pages, a horizontal scroll bar on large pages, zoom-in and zoom-out function, all browsers supported.) it has two levels: "Established" valued 1 and "to be established" valued 0. The total score is 7. The calculation of the value of the quantitative standard is:
2.3.4	Format specification	$\frac{1}{4} \left( \min \left( 4, \sum_{i=1}^n w_i \right) \right)$ "n" is the number of the sub-items and equals 8. $W_i$ is the specific value of the $i$ sub-item. The level of the sub-item is figured out through the random examination of 20 pages of the hospital's website. The "min" function is used to find the smallest value in the given parameter list. The specification of domain name is whether the URL of the hospital website is composed of the hospital's English name or Chinese pinyin initials, whether it has multiple corresponding domain names and whether users find it easy to remember and use. This item has three levels: (1) if the domain name is not composed of the hospital's English name or Chinese pinyin initials, and it does not have multiple corresponding domain names, the value is 0 (2) if the domain name is composed of the hospital's English name or Chinese pinyin initials, while it does not have multiple corresponding domain names, the value is 0.5 (3) if the domain name is composed of the hospital's English name or Chinese pinyin initials, and it has multiple corresponding domain names formed by the hospital's English name and Chinese pinyin initials, the value is 1
2.3.5	Specification of domain name	If the website only has simple Chinese version, the value is 0.5; if it has two or more language versions including Chinese, the value is 1.
2.3.6	Language	
	Information architecture	
2.4.1	Time taken to open the homepage	According to the time users need to open the hospital website's homepage, the time taken to open the homepage (before flashes and videos are played) has four levels: (1) if the homepage open time exceeds 10 seconds, the value is 0 (2) if the homepage open time is between 8 and 10 (10 included) seconds, the value is 0.34 (3) if the homepage open time is between 4 and 8 (8 included) seconds, the value is 0.67 (4) if the homepage open time is no more than 4 seconds, the value is 1  The digital multimedia function refers to dynamic forms of information services provided by the hospital website, including pictures, sounds, videos, flashes, dynamic pages and 3D interaction programs. Each of the above six sub-items has two levels: "Established" valued 1 and "to be established" valued 0. The total score is 4. In case that advertisement appears in the forms of dialog boxes and float bars on the website, 1 point is deducted. The calculation of the value of the quantitative standard is:
2.4.2	Digital multimedia function	

TABLE 2: Continued.

No.	Evaluation standards	Description of quantitative assignment
		$\frac{1}{4} \left( \min \left( 4, \sum_{i=1}^n w_i \right) - R \right)$ <p>“<math>n</math>” is the number of the sub-items and equals 6.</p> <p><math>W_i</math> is the specific value of the <math>i</math> sub-item. <math>R</math> is the case that incurs any point deduction.</p> <p>The level of the sub-items is figured out through the random examination of 20 pages of the hospital’s website.</p> <p>The “min” function is used to find the smallest value in the given parameter list.</p> <p>The website page view is the function to calculate the number of visitors to the hospital website.</p> <p>According to its function, the item has two levels: “Established” valued 1 and “to be established” valued 0.</p>
2.4.3	Website Pageview	
2.4.4	Background data	<p>The background data refers to that the information service resources on the website are posted in the format and form of background data. It has four sub-items:</p> <p>(1) if the resources are not in the format and form of background data, the value is 0</p> <p>(2) if part of the contents on the website is in the format and form of background data, the value is 0.34</p> <p>(3) if half or more of the contents of the website are in the format and form of background data, the value is 0.67</p> <p>(4) if all the contents of the website are in the format and form of background data, the value is 1</p>

Source: the author.

and the public and professionals. Through the empirical research, we get the comprehensive evaluations of the information service of a Hospital Information Platform’s information platform. If participants give different scores for evaluations of these two items, a mean value is adopted as the final score. The total averages of Pre-revision are: 19.4 and 20.3875, and the total averages of Post-revision are: 73.5, 74.4, 77.7, 77.3, 77.5, 79.4, 81.7, 84.3. Data in show that after the overall revision, the website of a Hospital Information Platform gets clearly higher comprehensive evaluations for its information service. (for details, see Table 3).

By means of the Bivariate Correlation in SPSS, the author carried out a correlation analysis of the comprehensive evaluations of the information service of a Hospital Information Platform’s information platform and the traffic data of a Hospital Information Platform website, including the total traffic, PV (page view) and UV (unique visitor). The total traffic is equal to PV plus UV. For details of the analysis results, the correlation coefficient among the three objects is 1, a significant correlation. It also suggests that the comprehensive evaluations of the information service of a Hospital Information Platform’s information platform and the traffic of a Hospital Information Platform’s website are positively correlated. That is to say, the traffic of a Hospital Information Platform’s website will vary to the changes in the comprehensive evaluations of the information service of a Hospital Information Platform’s information platform.

## 5. Conclusions

Based on the actual situation of Chinese hospitals, creates a theoretical research platform of medical information service quality based on the reality of a Hospital Information Platform, and sorts out the core of website information service of Chinese hospitals as well as the influencing factors and their correlations. Focusing on the function and quality of

hospital information service, it has established a complete and scientific evaluation system of hospital information service that conforms to the actual situation of Chinese hospitals. Still, through the empirical research on a Hospital Information Platform website, it also validates the scientificity, rationality and practicability of the model proposed in this thesis. The model’s scores in various dimensions in empirical research See Figure 4 for details.

For the first time in China, the author conducts comparison and correlation analysis of survey results on comprehensive evaluation of hospital website information service and traffic data through empirical research, in order to discuss the relations between the evaluation results and the traffic data. And through the above empirical research to prove the positive correlation between the two, see Figure 5 for details.

## 6. Suggestions

The results of the empirical research show Enriching the information service functions plays an important role for hospitals to improve the information service quality of their websites and their website usage. The weight of the functional quality attribute of website information service is 0.7, significantly higher than that of the supporting attribute. In addition, the second-level indicators of the functional quality attribute are the dimensional indicators for measuring the functions of hospital websites, while in the evaluation on January 9, 2020, the evaluation score of the functional quality attribute is 55.41, suggesting a improving.

As the evaluation of the functional quality attribute of website information service shows a large room for improvement, perfecting the information service function of a Hospital Information Platform’s website is the key task that the hospital should do to enhance its information service. In addition, that after the information service function of a

TABLE 3: Comprehensive Evaluations of the Information Service of a Hospital Information Platform's Information Platform.

No.	Period	Date	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	Total	Total score
1	Pre-revision	2016/12/20	0.088	0.000	0.000	0.003	0.033	0.025	0.021	0.025	0.194	19.4
2	Pre-revision	2017/01/05	0.088	0.000	0.000	0.003	0.042	0.025	0.021	0.025	0.204	20.4
3	Post-revision	2017/01/20	0.151	0.205	0.086	0.051	0.096	0.071	0.033	0.043	0.735	73.5
4	Post-revision	2017/02/06	0.151	0.205	0.089	0.051	0.101	0.071	0.033	0.043	0.744	74.4
5	Post-revision	2017/08/18	0.157	0.205	0.089	0.055	0.113	0.078	0.037	0.043	0.777	77.7
6	Post-revision	2018/02/06	0.157	0.205	0.089	0.055	0.113	0.078	0.037	0.038	0.773	77.3
7	Post-revision	2018/08/03	0.157	0.205	0.089	0.055	0.113	0.078	0.037	0.041	0.775	77.5
8	Post-revision	2019/02/21	0.157	0.205	0.090	0.058	0.116	0.090	0.037	0.041	0.794	79.4
9	Post-revision	2019/08/21	0.161	0.205	0.101	0.067	0.116	0.090	0.037	0.041	0.817	81.7
10	Post-revision	2020/01/09	0.166	0.205	0.117	0.067	0.116	0.090	0.038	0.045	0.843	84.3

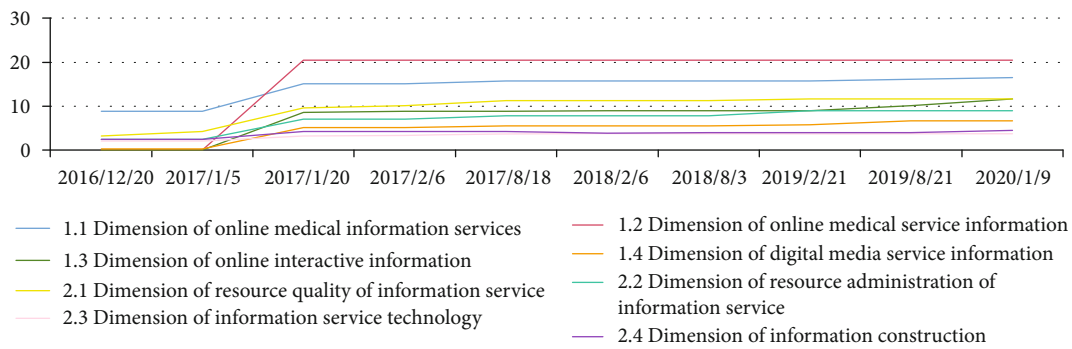
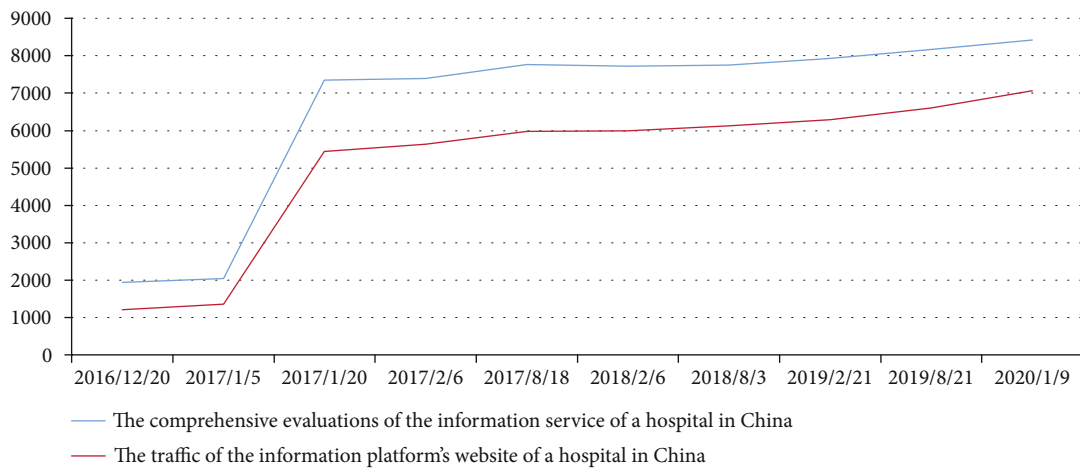


FIGURE 4: Line Chart of the Correlation Analysis of the Comprehensive Evaluations of The Secondary Indicators of The Information Service of a Hospital in China.

FIGURE 5: Line Chart of the Correlation Analysis of the Comprehensive Evaluations of the Information Service of a Hospital in China and the Traffic of Information Platform's Website of a Hospital in China (Comprehensive Evaluations of the Information Service of a Hospital in China of value = value $\times$ 100).

Hospital Information Platform's website is improved or the evaluation of the functional quality attribute of website information service scores higher, the supporting attribute of website information service will be the next key task for the hospital in enhancing its information service quality.

From the data in Figure 6 and Table 3, we can see that the improvement ratio of Functional quality attribute of website information service and the supporting attribute of website information service between 2017/8/18 and 2020/1/9 tend to be the same. And even the improvement rate of the



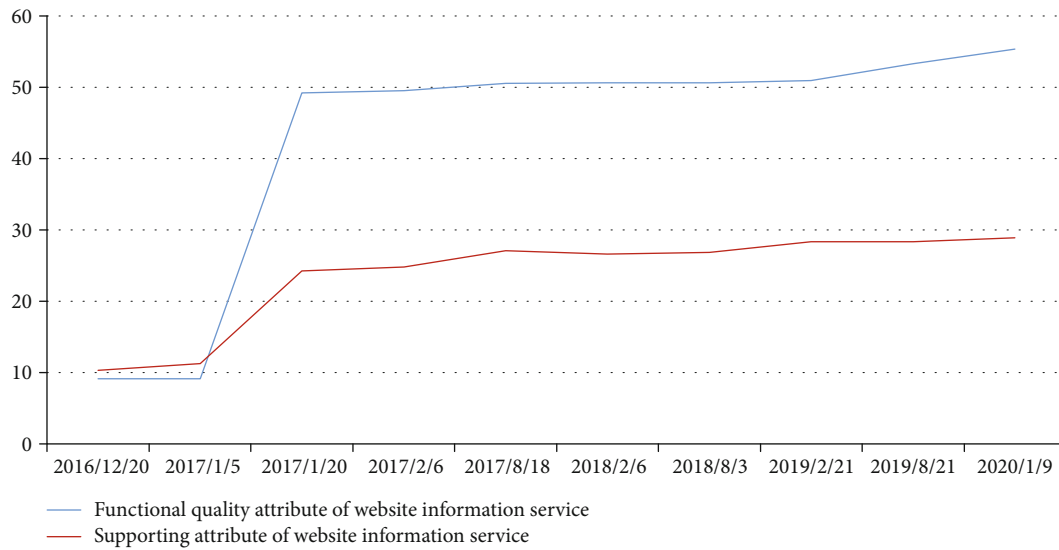


FIGURE 6: Line Chart of the Functional quality attribute of website information service and Supporting attribute of website information service.

TABLE 4: Evaluation score of the information services of a Hospital Information Platform's website in China of Growth of rate in empirical research.

Date	1	2	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4
2016/12/20	0	0	0	0	0	0	0	0	0	0
2017/1/5	0.5000	1.0777	1	0	0	1	1.2923	1	1.0183	1
2017/1/20	4.6790	2.1007	1.7159	0	0	17.0000	2.2857	2.8400	1.5569	1.7200
2017/2/6	1.0088	1.0169	1	1	1.0351	1	1.0521	1	1.0154	1
2017/8/18	1.0313	1.0843	1.0389	0.9988	1.0055	1.0819	1.1213	1.0990	1.1229	0.9942
2018/2/6	1.0013	0.9733	1	1	1.0052	1	1	1	0.9975	0.8958
2018/8/3	1	1.0147	1	1	1	1	1	1	1	1.0588
2019/2/21	1.0144	1.0449	1	1	1.0103	1.0471	1.0262	1.1534	1	1
2019/8/21	1.0744	1	1.0250	1	1.1140	1.1586	1	1	1	1
2020/1/9	1.0473	1.0337	1.0305	1	1.1588	1	1	1	1.0251	1.1099

supporting attribute is sometimes higher than the improvement rate of Functional quality.

All in all, between 2017/8/18 and 2020/1/9, the final score of the functional quality attribute of the website information service in 2020/1/9 is about 6 times that of 2017/8/18, and the support attribute of the website information service The final score in 2020/1/9 is about 3 times that of 2017/8/18, see Figure 6 and Table 4 for details.

This study is the first one in China that applies both the Delphi method and composite grade method to construct the model of comprehensive evaluation system of hospital information service. For the first time in China, the author conducts comparison and correlation analysis of survey results on comprehensive evaluation of hospital website information service and traffic data through empirical research, in order to discuss the relations between the evaluation results and the traffic data. Therefore, it bears considerable reference value. This model is both of high theoretical value and practical value. But the development of new network information technology often goes beyond

hospital managers imagination. Therefore, the information service evaluation system in this thesis will be enriched with the development of the times and information technology.

### Data Availability

The Survey results and network traffic data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Authors' Contributions

All authors have made a substantial, direct, intellectual contribution to this study. Lei, Jiao are Corresponding author and First author. XiaoZhao Zhu and HuaPing Xiao are Co-First author. Xu Zhao are Second author.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

## References

- [1] China Internet Network Information Center (CNNIC), "Cnnic.Net.Cn, 2020," January 2020, [https://www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwtjbg/201908/t20190830\\_70800.htm](https://www.cnnic.net.cn/hlwfzyj/hlwzxbg/hlwtjbg/201908/t20190830_70800.htm).
- [2] Y. Fei and Y. C. Sheng, "A Comparative Study of the Operation of Third-Class Class A Hospital Websites in China," *Chinese Hospital Management(Chinese)*, vol. 9, pp. 25–27, 2019.
- [3] S. J. Philip George, "Assessment of the service quality of a cancer information service using a new P-C-P attributes model," *Managing Service Quality*, vol. 9, no. 3, pp. 167–179, 1999.
- [4] M. M. Jeon and M. Jeong, "Customers' perceived website service quality and its effects on e-loyalty," *International Journal of Contemporary Hospitality Management*, vol. 29, no. 1, pp. 438–457, 2017.
- [5] K. C. Tan, M. Xie, and Y. N. Li, "A service quality framework for web-based information systems," *The TQM Magazine*, vol. 15, no. 3, pp. 164–172, 2003.
- [6] J. C. Pinho, M. D. L. Martins, and I. Macedo, "The effect of online service quality factors on internet usage," *International Journal of Quality & Reliability Management*, vol. 28, no. 7, pp. 706–722, 2011.
- [7] A. E. Baldwin, "Service quality in an Australian private dental network," *TQM Journal*, vol. 26, no. 4, pp. 360–367, 2014.
- [8] C.-J. Liang, "a study of the impacts of website quality on customer relationship performance," *Total Quality Management & Business Excellence*, vol. 20, no. 9, pp. 971–988, 2009.
- [9] L. W. Pei and Jie Tang, "Research on customer-faced Information Platform information service quality evaluation system," *Information Magazine*, vol. 5, pp. 60–66, 2009.
- [10] L. Zhu and Xin Jiang, "Evaluation criteria of medical information platform information service quality," *Chinese Journal of Medical Library and Information Science*, vol. 6, pp. 1–4, 2012.
- [11] J. Lei and R. M. C. P. de Menezes, "Fairness evaluation method of resource allocation based on BPSO multidimensional perspective," *Cluster Computing*, vol. 22, no. S2, pp. 4283–4290, 2019.
- [12] J. S. Armstrong, "Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared," *SSRN Electronic Journal*, vol. 8, pp. 17–20, 2008.

## Research Article

# An Indirect Multimodal Image Registration and Completion Method Guided by Image Synthesis

Huan Yang<sup>1,2</sup>, Pengjiang Qian<sup>1,2</sup> and Chao Fan<sup>1,2</sup>

<sup>1</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

<sup>2</sup>Jiangsu Key Lab of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China

Correspondence should be addressed to Pengjiang Qian; [qianpjiang@jiangnan.edu.cn](mailto:qianpjiang@jiangnan.edu.cn)

Received 1 May 2020; Accepted 8 June 2020; Published 30 June 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Huan Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multimodal registration is a challenging task due to the significant variations exhibited from images of different modalities. CT and MRI are two of the most commonly used medical images in clinical diagnosis, since MRI with multicontrast images, together with CT, can provide complementary auxiliary information. The deformable image registration between MRI and CT is essential to analyze the relationships among different modality images. Here, we proposed an indirect multimodal image registration method, i.e., sCT-guided multimodal image registration and problematic image completion method. In addition, we also designed a deep learning-based generative network, Conditional Auto-Encoder Generative Adversarial Network, called CAE-GAN, combining the idea of VAE and GAN under a conditional process to tackle the problem of synthetic CT (sCT) synthesis. Our main contributions in this work can be summarized into three aspects: (1) We designed a new generative network called CAE-GAN, which incorporates the advantages of two popular image synthesis methods, i.e., VAE and GAN, and produced high-quality synthetic images with limited training data. (2) We utilized the sCT generated from multicontrast MRI as an intermediary to transform multimodal MRI-CT registration into monomodal sCT-CT registration, which greatly reduces the registration difficulty. (3) Using normal CT as guidance and reference, we repaired the abnormal MRI while registering the MRI to the normal CT.

## 1. Introduction

Deformable image registration (DIR) is to find the spatial relationship between two or more images and is abundantly used in medical image analysis, such as image fusion, lesion detection, disease diagnosis, surgical planning, and navigation. It is necessary to analyze the relationships among images that were acquired from different viewpoints, at different times, or using different sensors/modalities [1]. Computer Tomography (CT) and Magnetic Resonance Imaging (MRI) are two of the most commonly used medical images in the clinical diagnosis due to the complementary information and multicontrast images they provided. Among them, CT shows precise skeletal location information and electron density information, which is often used for the dose planning of cancer patients. On the other hand, MRI has clear anatomical structures and multiple imaging modalities that enable the detection and segmentation of diseased organs

and tissues. The DIR of MRI and CT is essential and a challenging task, due to the inherent structural differences among different modalities and the missing dense ground truth.

Typically, image registration is an iterative optimization process. It requires both a metric that quantifies the similarity between a moving image and a fixed image and an optimization algorithm that iteratively updates the transformation parameters such that the similarity between the images is maximized. The main difficulty of MRI-CT DIR is the definition of the image similarity measure, which is an inherent problem in multimodal image registration. What is more, the registration of MRI and CT is more difficult to perform due to MR images that may be contaminated or damaged by abnormal magnetic fields during acquisition, resulting in unknown deformation. This is a challenging task, not only for image registration but also for image completion.

The significant variations exhibited from images of different modalities make multimodal registration much more

troublesome than monomodal registration. Various methods have been proposed to solve multimodal registration tasks, such as mutual information-based [2, 3], elastodynamics-based [4], and learning-based [5, 6]. However, most of these methods are task-sensitive, high-dimensional iterative optimization methods, which are disadvantageous in terms of computational complexity and are difficult to apply in the clinic [7]. Inspired by image fusion, we explore to utilize synthetic images to address the challenge of multimodal image registration and ultimately repair the problematic images. That is, we try to adopt image synthesis algorithms to generate synthetic CT (sCT) from corresponding multicontrast MRI and then use sCT and corresponding CT scans from the same subject to perform image registration. The main idea of our method is to replace the traditional input for multimodal registration with synthetic and real CT images. In this way, multimodal registration is approximately converted into monomodal registration, so MRI-CT registration is indirectly transformed into sCT-CT registration. Guided by the real CT, sCT is generated by the fusion of multiple modality MRI from the same subject, so it has the ability to incorporate the deep features of MRI and the surface features of CT to the greatest extent.

Concerning image synthesis, there are already several existing methods, such as atlas-based [8] and learning-based [9–11]. Atlas-based methods are aimed at generating an atlas among a set of images and then applying the atlas to a new subject. Since this method only relies on geometric transformations, effectiveness and stability are difficult to guarantee. Learning-based methods are achieved by learning a potential nonlinear mapping between source and target domain images. Among them, three themes are mainly used: Variational Auto-Encoder (VAE) [12], Generative Adversarial Network (GAN) [13], and Autoregression [14]. VAE pairs a differentiable encoder network with a decoder network to reconstruct images. However, due to the lack of imperfect similarity measurements, its outputs are often blurry. GAN automatically learns the measurement via a discriminative network and thereby promoting the generative network to generate near-real synthetic images. Inspired by conditional GAN (CGAN) [15] and VAE/GAN [16], we proposed a new deep generative network that combines the idea of VAE and GAN under a conditional process to tackle the problem of sCT synthesis.

The main target of our method is to convert multimodal registration into monomodal registration through a robust image synthesis algorithm and ultimately repair the problematic images. Therefore, our main efforts can be summarized into three aspects:

- (1) We combined the current popular image synthesis algorithms, i.e., VAE and GAN, to propose a new Conditional Auto-Encoder Generative Adversarial Network, called CAE-GAN. Compared with existing methods, the synthetic images generated by our network are of higher quality and more generalized
- (2) We utilized the generated sCT to transform multimodal MRI-CT registration into monomodal sCT-CT

registration, which greatly reduces the registration difficulty caused by inherent structural differences among different modalities

- (3) Using normal CT as reference and guidance, through the deformation fields obtained by sCT and CT, we succeed not only in registering multiple modality MRI but also in repairing abnormal MRI

The following parts of the manuscript are arranged as such. Related work, e.g., image synthesis and image registration algorithms, is reviewed briefly in Section 2. The proposed method, CAE-GAN and sCT-guided image registration and completion, is introduced specifically in Section 3. The experimental studies, results, and discussion are presented in Sections 4. Conclusions are given in Section 5.

## 2. Related Work

**2.1. Image Synthesis.** Most traditional image synthesis algorithms include feature extraction, modeling, and target reconstruction, with an assumption that data have a simple formation. They have difficulty in modeling complex patterns of high dimensional, irregular distributions. There have been many recent developments of deep learning-based generative models [17–20], since deep hierarchical architectures are capable of capturing underlying complex features in data. VAE [12] and GAN [13] are two of the most commonly used methods for image synthesis.

VAE is composed of a recognition model and a generative model. Recognition model is also referred to as a probabilistic encoder (E), since given a datapoint  $x$ , it can produce a distribution (e.g., a Gaussian). Through this distribution, a latent code  $z$  can be sampled, so that datapoint  $x$  can be reconstructed from  $z$  by the generative model (G, also called probabilistic decoder). However, a disadvantage of VAE is that, due to the injected noise and imperfect elementwise measurements such as the squared error, the generated samples are often blurry.

GAN is a generative model that generally includes two subnetworks, a generator  $G$  and a discriminator  $D$ .  $G$  learns a mapping to generate fake samples from random noise, and  $D$  tries to discriminate whether the samples are from real or fake. The two networks are like the two sides of a game, and the performance of the two networks gradually improves in the confrontation until  $D$  cannot discriminate whether the sample is real or fake. No prior knowledge is needed, and the fake samples are fitted by random noise. This is both an advantage and a disadvantage of original GAN, because the mapping without premodel is too free and broad, it is difficult to get good results. With the proposal of CGAN [15], it adds constraints (condition) to the original GAN to guide image synthesis, which alleviates this problem to some extent. However, the problem still exists, and GAN-based models are hard to converge in the training stage.

There have been some methods which tried to combine GAN and VAE, such as VAE/GAN [16], adversarial autoencoders [21], CVAE-GAN [22]. Inspired by these works, we proposed a new condition-driven deep generative network

combining the idea of VAE and GAN, called CAE-GAN. In addition, inspired by the classifier used in Auxiliary Classifier GAN (ACGAN) [23], we reuse the discriminative network as a classifier to further enhance the performance of the network. The detailed network framework is described in Section 3.1.

**2.2. Image Registration.** Generally speaking, image registration is an iterative optimization process, which is achieved by maximizing a predefined image similarity metric calculated from the moving image and the fixed image through an optimization algorithm. Several manually crafted metrics are frequently used, such as the sum of squared differences (SSD), cross-correlation (CC) [24], mutual information (MI) [25], normalized cross-correlation (NCC), and normalized mutual information (NMI). The optimization algorithms are mostly intensity-based [26, 27] and feature-based [28–30]. Actually, image registration generally includes linear (rigid) registration and deformable (nonrigid) registration, where linear registration intends to globally align the two images, and deformable registration is used to correct local deformations. Deformable registration is used to compute a displacement vector for each voxel of an image according to a metric, enabling the estimation of the spatial variations of the anatomy. The displacement vectors are computed to point to the best corresponding location of the voxels in another image according to the metric which is a measure of the image matching.

Among the existing methods, nonparametric deformable registration algorithms are widely used to estimate tissue deformations in highly deformable anatomies due to the advantage of being fast and easy-to-use, such as Demons [31] and Morphons [32]. Demons is an intensity-based registration algorithm, which requires no particular preprocessing nor patient-specific modeling. It is aimed at calculating a regular displacement field which produces a good matching of the intensities in fixed and moving images, along with a measure of the field regularity. However, intensity-based methods are not suitable for registering images with different contrast enhancements. Morphons is a method similar to Demons but phase-based, whose principle is to match transitions rather than intensities, by looking locally at the spatial oscillations in intensities. This method uses Gaussian smoothing as a regularization of the displacement field and additive accumulation during the iterative process. What is more, diffeomorphic transforms [33], which preserve topology and invertibility on the transformation, have shown remarkable superiority in various computational anatomy studies [34–38]. Actually, there have been many learning-based registration methods, such as Quicksilver [39], VoxelMorph [40], and BIRNet [41]. They utilize Convolutional Neural Network (CNN) and spatial transformation function [42] to estimate the similarity measure of the two images or directly predict the transformation parameters of the deformation field. But they are beyond the scope of this article, so we do not introduce much here. In this work, we adopted a registration method based on local phase differences and diffeomorphic accumulation like Diffeomorphic

Morphons [38]. The detailed algorithm is presented in Section 3.2.

### 3. Our Proposed Method

Our proposed method consists of two main phases. The first phase is sCT synthesis, two advanced deep learning-based methods, i.e., VAE and GAN, are applied. The second phase is sCT-guided multimodal image registration and image completion. Further details will be given below.

**3.1. Phase 1: sCT Synthesis Based on Deep Generative Network.** In order to obtain high-quality sCT, we combined the most popular image synthesis methods, i.e., VAE and GAN, to propose a new deep generative network, called CAE-GAN. As shown in Figure 1, it consists of four subnetworks: (1) the encoder network (E), (2) the generator network (G), (3) the discriminator network (D), and (4) the classifier network (C).

The function of networks E and G is the same as that in VAE [12]. The network E maps the input data  $x$  (e.g., brain fat, R2, and water MR images) to a latent distribution  $P_z$  with mean  $\mu$  and standard deviation  $\delta$ . The latent variable sample  $\tilde{z}$  is sampled from  $P_z$  through sample  $\tilde{z} = \mu + (\delta * z)$  as the input data of G, where  $z$  is random noise. The function of network G and D is the same as that in GAN [13]. The network G tries to generate synthetic images from the latent variable sample  $\tilde{z}$ , and D tries to discriminate whether the input images are synthetic or real and outputs a probability distribution  $d$  over input images. They automatically update gradients through adversarial training to improve network performance. The network C is realized by reusing the network structure of D, which only need to change the last activation function of D from sigmoid to softmax. Because the anatomical structure of the brain is relatively simple and clear, its internal tissues can be roughly divided into three types: soft tissue, air, and bone. Therefore, C performs 3-class classification and outputs a probability distribution  $c$  over the class labels.

However, it is not sufficient to reconstruct images only through the data distribution mapped by E, which results in blurred images. We redesigned the network E and G in the form of concatenation and residual blocks, to form a symmetric encoder-decoder structure. Using this structure, the features of the same resolution at different stages can be fused on the channel via concatenation operation. What is more, we designed encoder and decoder residual blocks to retain more latent features and accelerate network convergence. The use of residual blocks can also avoid possible problems of gradient disappearance or gradient explosion to a certain extent. The detailed structures can be seen in Figure 1.

As for the specific network structure, all convolutional layers (Conv) are followed by normalization (Norm) and the Rectified Linear Unit (Relu) to form a module of Conv-Norm-Relu and all deconvolutional layers (Deconv) are followed by Norm, Dropout (Dp), and Relu to form a module of Deconv-Norm-Dp-Relu. The normalization methods commonly used in deep learning are Batch Normalization (BN) and Instance Normalization (IN). We use IN in our



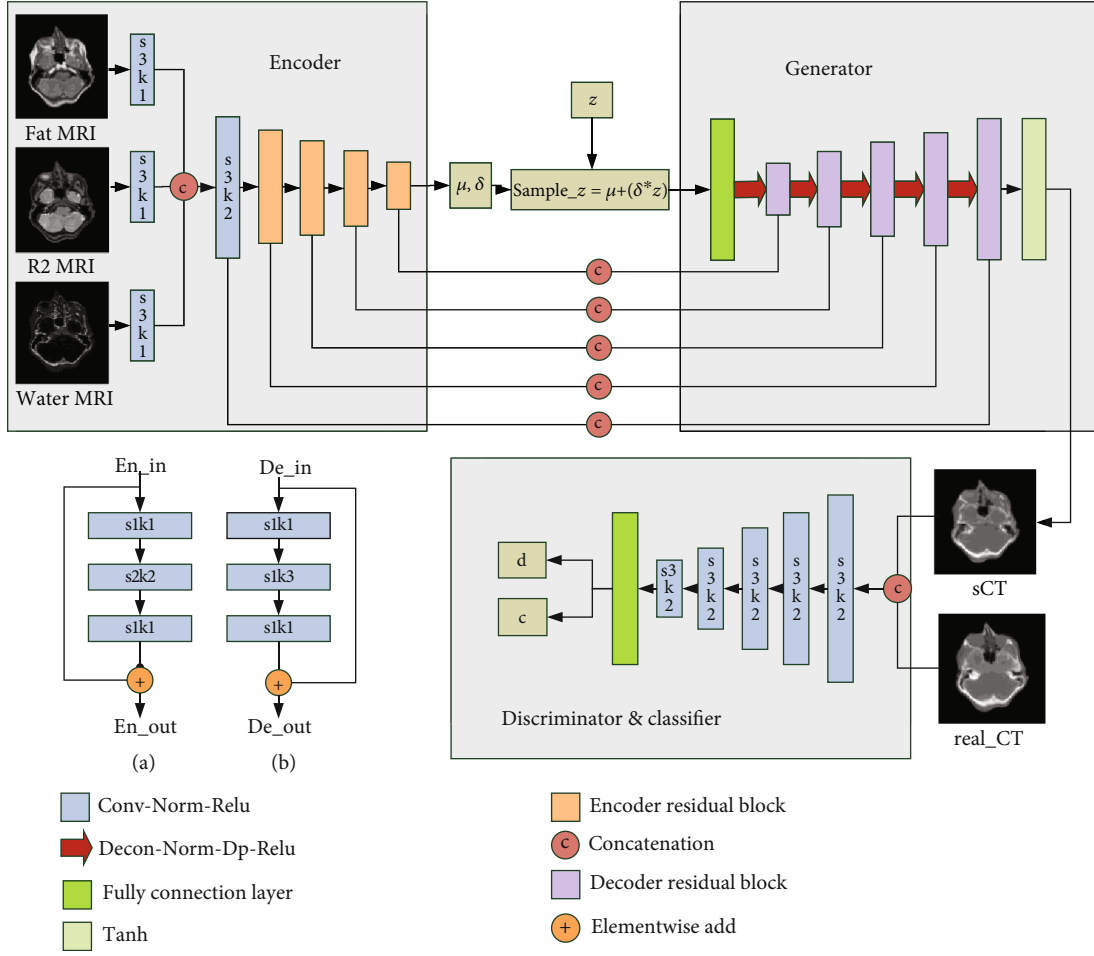


FIGURE 1: The overall framework of our proposed ACE-GAN. It consists of four subnetworks, encoder (E), generator (G), discriminator (D), and classifier (C). The detailed structure of the encoder and decoder residual block is shown in (a) and (b), respectively. Specifically, s1k1 represents that the convolution stride is 1 and the kernel size is  $1 \times 1 \times 1$ .

network, because it is normalized from image pixels that are more suitable for image translation tasks. There are three contrast (modality) brain MRI (fat, water, and R2) as input data; we start with three convolutional layers for each input data separately (late-fusion) instead of stacking them as channels (early-fusion). We found late-fuse is better than early-fuse during the experiment.

Using an appropriate and effective loss function is essential for network parameter optimization and performance improvement. As illustrated in Equation (1), we utilize the adversarial loss  $L_{adv}$  to ensure the adversarial training of G and D. Specifically, G tries to minimize the adversarial loss and D tries to maximize it. In addition, we also introduce an image reconstruction loss  $L_{l1}$  to minimize the difference between synthetic images and reference images, as presented in Equation (2). Therefore, the ultimate objective function is summarized as Equation (3). In these equations below,  $x$  represents the source domain images (MR),  $y$  represents the target domain images (CT), and  $c$  represents class labels. We use a parameter  $\lambda$  to control the relative weight of the image reconstruction loss function, set to 100 by default.

$$L_{adv} = E_y[\log D(y)] + E_x[1 - \log D(G(E(x)))] + E_y[\log D(c|y)] + E_x[1 - \log (D(G(E(c|x))))] \quad (1)$$

$$L_{l1} = E_{x,y}[\|y - G(E(x))\|_1], \quad (2)$$

$$L_{total} = L_{adv} + \lambda L_{l1}. \quad (3)$$

**3.2. Phase 2: sCT-Guided Multimodal Image Registration and Image Completion.** In our method, we adopted a registration algorithm similar to Diffeomorphic Morphons [38]. This is a nonparametric deformable registration method based on local phase differences at multiple scales. The deformation field estimation is stabilized by using a simple smoothing and downsampling procedure to decompose the fixed and the moving images on several scales. This registration algorithm contains an iterative loop that mainly consists of four interconnected steps, i.e., deformation field computation, deformation field accumulation, deformation field regularization, and image deformation. The idea is to progressively build a proper displacement field by iteratively improving

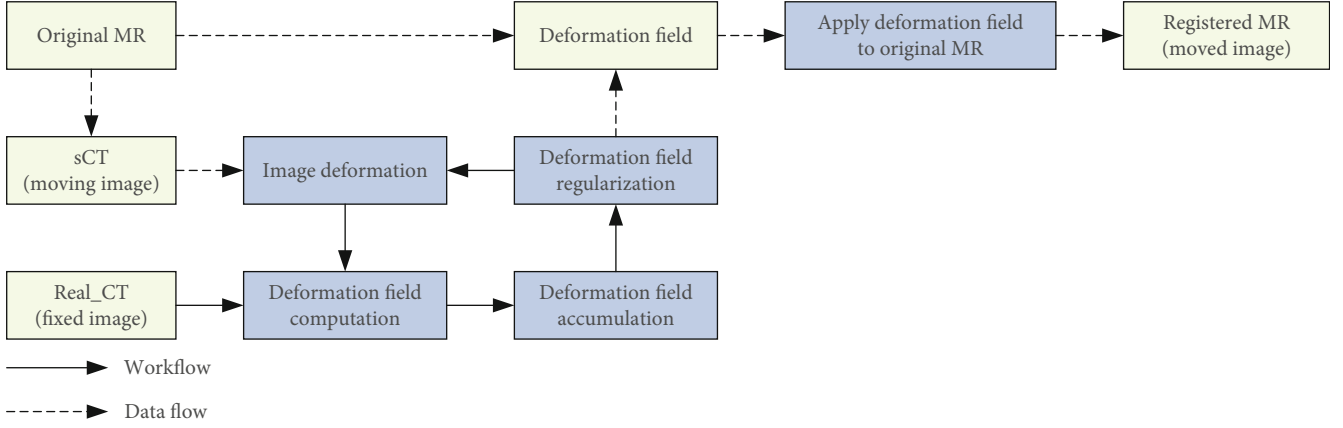


FIGURE 2: The pipeline of sCT-guided multimodal image registration used in our method.

the matching between the fixed and moving images warped by the displacement field, according to a certain metric. The overall framework of this phase is shown in Figure 2.

- (1) Deformation field computation: the deformation field is estimated by the dephasing between the local phase of the fixed and moving images. This local phase can be probed at a certain frequency and in a particular direction using quadrature filters [43]. Given the fixed image  $f$  and moving image  $m$ , the deformation field  $D_u$  can be calculated by solving a weighted least square optimization problem. Accordingly, the global certainty mapping of deformation field can be expressed mathematically as

$$\Theta_c(f, m) = \sum_k c_k(x), \quad (4)$$

where  $c_k(x) = A_f(x; k)A_m(x; k)$  is the certainty mapping of the filter mentioned above, with  $A_f(x; k)$  and  $A_m(x; k)$  denoting the amplitudes of the fixed and moving images, respectively. Then, this update of certainty map will be combined with an accumulated certainty computed from previous iterations

- (2) Deformation field accumulation: the complete accumulated field is computed as a weighted sum of the update field and the previous accumulated field. The weights are given by the certainty on the update field and the accumulated certainty map. Besides, some new tricks are adopted in this method. Let the deformation field be represented by a vector field  $D$ , and let  $\Delta = \Delta Id + D$  denote the deformation operation, where  $Id$  denotes the identity deformation:  $Id(x) = \Delta x$ . Then, the accumulation process can be expressed as

$$D_1 \oplus D_2 \triangleq \Delta_1 \circ \Delta_2 - Id, \quad (5)$$

where  $\circ$  represents the common function composition operation, denoting the warping of two objects. The composition will remain diffeomorphic in the case of two diffeomorphic deformation fields

- (3) Deformation field regularization: field regularization is in order to get a smoother transformation and reduce the impact of image noise on the registration output. In our method, regularization is achieved by using a normalized convolution [44] of the deformation field by the Gaussian kernel. In addition, the certainty map obtained from local phase computation is also used to adjust the importance of the target locations, i.e., the higher significance is attached to locations with larger values in certainty map
- (4) Image deformation: the three steps mentioned above are performed iteratively a certain number of times at each scale from coarse to refine, until it reaches a certain stopping criterion. Then, the resulting regularized deformation field is used to warp the moving image to obtain the registered image

In our method, we use sCT (moving image) and real CT (fixed image) as input and then apply the optimal deformation field output to original MRI, so as to obtain the final registered MRI (moved image). Because sCT is completely generated on the basis of MRI, which is representative of the characteristics of the original MRI, so the deformation field obtained by registering sCT and real CT can be used to register original MRI to reference normal CT. In this process, we transformed multimodal MRI-CT registration into monomodal sCT-CT registration and repaired the problematic MRI at the same time. Thus, our overall approach consists of two interrelated phases, i.e., sCT synthesis based on deep generative network and sCT-guided multimodal image registration and image completion. To some extent, the image registration and completion results in the second phase largely depend on the quality of synthetic CT generated in the first phase.

TABLE 1: Performance comparison of three image synthesis algorithms. The best results of each subject are presented in bold.

Sub	MAE			CC		
	FCM_based	FCN_based	CAE-GAN	FCM_based	FCN_based	CAE-GAN
1	157.81	77.11	<b>58.88</b>	0.75	0.94	<b>0.96</b>
2	176.91	<b>95.24</b>	98.44	0.61	<b>0.91</b>	0.89
3	157.94	99.77	<b>69.99</b>	0.71	0.90	<b>0.92</b>
4	178.49	87.34	<b>67.49</b>	0.74	0.92	<b>0.93</b>
5	165.03	103.34	<b>99.32</b>	0.70	0.89	<b>0.89</b>
6	152.67	105.57	<b>97.54</b>	0.71	0.89	<b>0.89</b>
7	169.38	82.08	<b>78.96</b>	0.67	0.92	<b>0.93</b>
8	146.57	94.91	<b>71.76</b>	0.67	0.89	<b>0.92</b>
9	171.82	110.18	<b>88.94</b>	0.79	0.88	<b>0.90</b>

## 4. Experiment and Discussion

**4.1. Experiment Setup.** The dataset used in our experiment includes nine sets of brain MR-CT images from nine healthy subjects after providing informed consent. For each subject, there are three types of MR sequences, e.g., fat, water, and R2, and corresponding CT scans from the same anatomy. To simulate the deformation and other unknown corruption that MR images may encounter during acquisition, we randomly warped the original MR images with salt-and-pepper noise. To make full use of these brain MR-CT images, we adopted a leave-one-out (LOO) strategy to get unbiased cross-validation results of each subject. LOO trains the network on all but one case and tests on that case and then repeats this process on each case. In other words, for the sCT of each subject, we obtained it by testing on the network which was trained with the images from the other nine subjects. To fit the memory and computing resources of the computer, we divided the original MR and CT images with the resolution of  $256 \times 256 \times 256$  into many  $64 \times 64 \times 64$  patches. The training and testing patches for the network are extracted by sliding over the original images. After the network converges, the generated patches are then recombined to form a complete sCT.

In order to demonstrate the superiority of our proposed framework, two image synthesis methods, i.e., FCN-based deep generative network and FCM-based clustering algorithm, are chosen as opponents for the comparison of sCT synthesis. On the other hand, two classical traditional registration methods, i.e., Diffeomorphic Morphons (D\_Morphons, a method based on local phase differences) [38] and Diffeomorphic Demons (D\_Demons, a method based on the sum of squared differences) [37], are selected for the comparison of image registration. In all non-learning-based approaches, the methods using diffeomorphic transforms can preserve topology and the invertibility of the transformation, so they are still effective and representative methods nowadays. We utilize the OpenREGGUI, an open-source image registration package to implement these registration methods.

Our experimental studies were carried out on a computer with an Intel Xeon® CPU E5-2640 V4, GeForce RTX 2080 Ti GPU using the Ubuntu 16.04 (64 bit) operating system. We use TensorFlow (Google, California, USA) and MATLAB

2016a (MATHWorks, Natick, MA, USA) to implement our proposed method.

**4.2. Metrics for Evaluation.** To demonstrate the effectiveness of our proposed deep generative network, we use the Mean Absolute Error (MAE) and the Pearson Correlation Coefficient (CC) to quantitatively measure the similarity between generated and reference images. That is, the values (see Table 1) are calculated by measured CT and sCT generated by different algorithms, i.e., FCN-based, FCM-based, and our proposed CAE-GAN. A lower value of MAE and higher value of CC indicate a smaller error and higher quality of generated images.

As for the evaluation of image registration, two metrics are used in our research: the mutual information (MI) and the sum of local phase differences (SLPD). Giving the fixed image  $f$  and moved image  $m$ , the MI measures the mutual dependence between two images and can be defined as

$$MI(f; m) = \sum_{f \in F} \sum_{m \in M} p(f, m) \log \left( \frac{p(f, m)}{p(f)p(m)} \right), \quad (6)$$

where  $p(f, m)$  is the joint probability function of  $f$  and  $m$ , and  $p(f)$  and  $p(m)$  are the marginal probability distribution functions of  $f$  and  $m$ , respectively. The SLPD is calculated as the sum of the local phase differences in all directions between two images. Mathematically, it can be defined as

$$SLPD = \sum (\sin(\Delta\varphi)), \quad (7)$$

where  $\Delta\varphi$  denotes the local phase differences between two images. A lower value of SLPD and higher value of MI indicate higher registration accuracy and better results.

**4.3. Results and Discussion.** The sCT synthesis phase is an important component of our proposed sCT-guided multi-modal image registration and completion method. We separately ran CAE-GAN, FCN\_based, and FCM\_based methods on the brain images obtained from 9 subjects. To illustrate the effectiveness of our proposed deep generative network, we calculated the evaluation metrics between measured CT and sCT that are generated by FCM\_based, FCN\_based, and our CAE-GAN methods, respectively. The values of

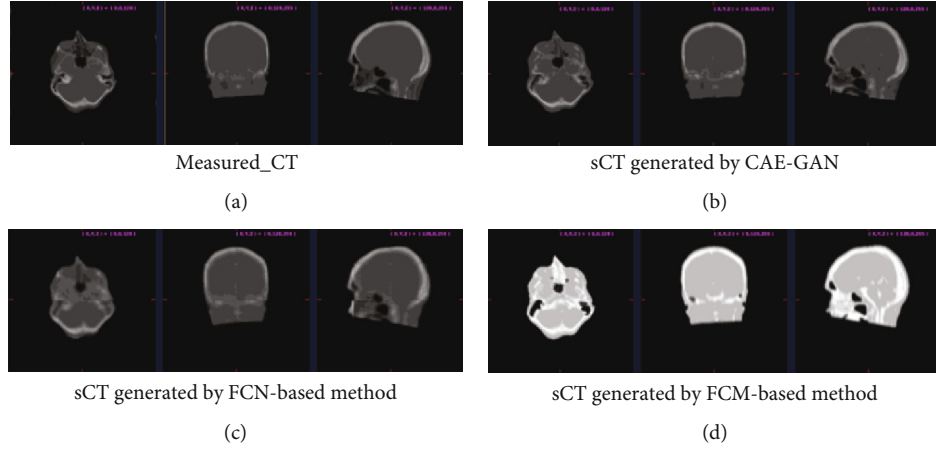


FIGURE 3: Results of different image synthesis algorithms on subject 1. (a) is the measured CT, and (b)–(d) are the synthetic CT generated by different methods. Each slice image is viewed from three perspectives, transverse, coronal, and sagittal, from left to right. It can be seen that (b) is the closest to the measured CT with high synthesis accuracy and quality.

TABLE 2: Performance comparison between our proposed method and other registration methods. The best results are presented in bold.

Sub	MI			SLPD		
	Ours	D_Morphons	D_Demons	Ours	D_Morphons	D_Demons
1	<b>0.9482</b>	0.9465	0.9466	<b>1.6159E + 05</b>	1.6803E + 05	1.7738E + 05
2	<b>0.9220</b>	0.9177	0.9161	<b>2.6173E + 04</b>	4.1678E + 04	2.9328E + 04
3	<b>0.9338</b>	0.9307	0.9328	<b>1.9668E + 04</b>	1.1733E + 05	5.8928E + 04
4	<b>0.9397</b>	0.9340	0.9386	<b>6.0200E + 04</b>	7.6741E + 04	9.0167E + 04
5	<b>0.9251</b>	0.9184	0.9238	<b>1.0450E + 05</b>	1.3337E + 05	1.5216E + 05
6	0.9316	<b>0.9456</b>	0.9311	9.4646E + 04	<b>8.2707E + 04</b>	9.2987E + 04
7	<b>0.9437</b>	0.9378	0.9435	<b>1.1808E + 05</b>	1.3797E + 05	1.2557E + 05
8	<b>0.9414</b>	0.9348	0.9411	<b>1.4943E + 05</b>	1.6041E + 05	1.8494E + 05
9	<b>0.9286</b>	0.9271	0.9278	<b>4.3636E + 04</b>	1.6514E + 05	7.6691E + 04

MAE and CC of different methods for each subject are listed in Table 1. As can be seen from the bold fonts (best) marked in the table, except for sub2, our proposed method has lower MAE and higher CC than other methods, which demonstrates that our method produces more precise and higher quality synthetic images. Figure 3 shows the results of sCT synthesis by different methods on subject 1. From these synthetic slices, we can clearly see that the sCT generated by our method is very similar to the measured CT with clearer internal tissue textures and higher-quality synthesis effects. In addition, we can conclude that deep learning-based methods (FCN\_based and our method) are superior to traditional machine learning-based methods (FCM\_based), which can be seen from Table 1 and Figure 3.

For the purpose of the performance comparison of different registration methods, we measured the MI and SLPD introduced above for each subject and listed in Table 2. They are calculated by measured CT and registered MRI obtained from different registration methods. It can be seen from evaluation metrics that our proposed sCT-guided image registration method has similar and even more superior performance to other classical traditional registration methods. From Figure 4

that shows the results of registration and completion on subject 1, we can clearly see that the registered MRI obtained using our method has a high matching of tissue deformation and spatial location with the corresponding reference CT and it repairs the initial abnormal MRI. Besides, it can be illustrated that the matching of structures in images based on intensity is unsuitable for registering variable contrast images, as can be seen from Figure 4(f). It may lead to uncontrollable abnormal tissue deformation, making the registration results not credible. What is more, considering that in clinical applications, time is very important. We recorded the total time consumption used in our experiments, which is listed in Table 3. Our proposed method in this work consists of two interrelated phases, i.e., sCT synthesis and image registration. Once the image synthesis network is trained well, the total time cost of our method is less than two minutes, which is much more efficient than the other two registration methods. Thus, our method obviously consumes less time and obtains better results than either of these methods.

Overall, in our method, we utilized a deep generative network to obtain sCT and then used sCT as an intermediary to convert the problem of multimodal MRI-CT registration into

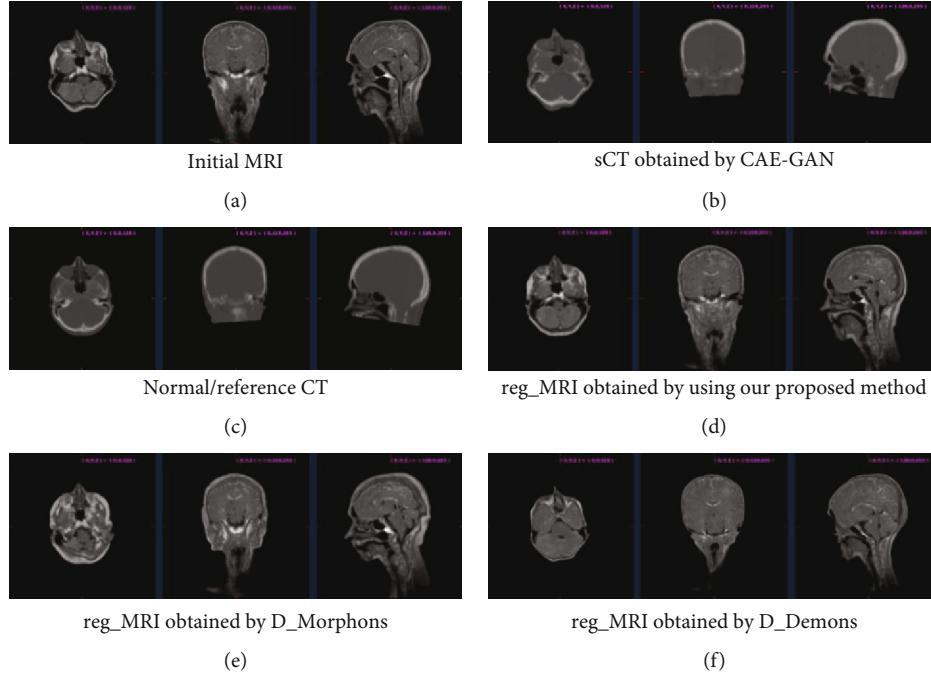


FIGURE 4: Results of image registration and completion on subject 1. (a) is the initial contaminated MRI, (b) is the synthetic CT as an intermediary, (c) is the normal CT as guidance and reference, and (d)–(f) are registered MRI obtained by using different registration methods. Each slice image is viewed from three perspectives, transverse, coronal, and sagittal, from left to right. The results demonstrate that the reg\_MRI obtained by our method has a high matching of tissue deformation and spatial location with the reference CT and repairs the abnormal MRI.

TABLE 3: Time consumption of three image registration methods.

Sub	sCT synthesis (s)	Ours Registration (s)	Total time (s)	D_Morphons Total time (s)	D_Demons Total time (s)
1	3.5	84	87.5	450	376
2	3.5	84	87.5	582	380
3	3.5	89	92.5	583	353
4	3.7	90	93.7	586	354
5	3.5	85	88.5	585	355
6	3.6	84	87.6	585	356
7	3.5	82	85.5	582	353
8	3.5	89	92.5	582	352
9	3.6	92	95.6	583	378

the problem of monomodal sCT-CT registration. In this way, we successfully addressed the challenges of multimodal registration caused by the significant differences between different contrast images and reduced the difficulty of registration. We utilized a traditional deformable registration method based on local phase differences and diffeomorphic accumulation to perform the image registration process and enable the completion of problematic MR images. We did not use the latest popular deep learning-based registration algorithms due to insufficient data and the lack of optimal optimization goals. The deep generative models are relatively mature now, but the deep registration methods still face many challenges, especially for multimodal registration. It is very difficult to design a

common optimal similarity index to measure the accuracy of registration. Due to the inherent defects of deep learning, strong data dependence, and poor interpretability, model stability and generalization are limited in practice. Besides, one disadvantage of our method is that we require normal CT scans from the same subject to serve as a guide and reference to enable indirect multimodal registration and completion, which may have trouble in practical application scenarios. Nevertheless, with limited data, we have proved the effectiveness of our proposed deep generative network and indirect multimodal image registration and completion method. Compared with the existing methods of multimodal registration, our method exhibits clear superiority.



## 5. Conclusion

In this paper, inspired by image synthesis, we utilized synthetic CT as an intermediary to solve the challenging multimodal image registration problem. We proposed a sCT-guided multimodal MRI registration and completion method and designed a new deep generative network called Conditional Auto-Encoder Generative Adversarial Network (CAE-GAN), which combined the idea of VAE and GAN, to obtain high-quality synthetic CT. We conducted experiments on brain MR-CT images provided by nine subjects. The experimental results illustrated that our designed deep generative network can yield high-quality synthetic images and our proposed image registration method can exhibit clear superiority in registration accuracy and time consumption compared with other methods.

## Data Availability

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

## Conflicts of Interest

We declare that there are no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grants 61772241 and 61702225, by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under grant 18KJB520048, by the Science and Technology Demonstration Project of Social Development of Wuxi under grant WX18IVJN002, and by JiangSu 333 expert engineering.

## References

- [1] D. Ruecker and J. A. Schnabel, "Medical Image Registration," in *Biomedical Image Processing*, Biological and Medical Physics, Biomedical Engineering, T. Deserno, Ed., pp. 131–154, Springer, Berlin, Heidelberg, 2010.
- [2] W. M. Wells, P. Viola, H. Atsumi, and S. Nakajima, "Multimodal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.
- [3] J. Woo, M. Stone, and J. L. Prince, "Multimodal registration via mutual information incorporating geometric and spatial context," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 757–769, 2015.
- [4] S. Ahmad and M. F. Khan, "Multimodal non-rigid image registration based on elastodynamics," *The Visual Computer*, vol. 34, no. 1, pp. 21–27, 2018.
- [5] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018.
- [6] Y. Hu, M. Modat, E. Gibson et al., "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical Image Analysis*, vol. 49, pp. 1–13, 2018.
- [7] C. Xue and F. H. Tang, "An adaptive patient specific deformable registration for breast images of positron emission tomography and magnetic resonance imaging using finite element approach," in *Medical Imaging 2014: Image Processing*, San Diego, CA, USA, 2014.
- [8] R. Farjam, N. Tyagi, J. O. Deasy, and M. A. Hunt, "Dosimetric evaluation of an atlas-based synthetic CT generation approach for MR-only radiotherapy of pelvis anatomy," *Journal of Applied Clinical Medical Physics*, vol. 20, no. 1, pp. 101–109, 2019.
- [9] Y. Wu, W. Yang, L. Lu et al., "Prediction of CT substitutes from MR images based on local sparse correspondence combination," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9349 of Lecture Notes in Computer Science, pp. 93–100, Springer, Cham, 2015.
- [10] T. Huynh, Y. Gao, J. Kang et al., "Estimating CT image from MRI data using structured random forest and auto-context model," *Medical Imaging, IEEE Transactions on*, vol. 35, no. 1, pp. 174–183, 2016.
- [11] H. V. Nguyen, K. Zhou, and R. Vemulapalli, "Cross-domain synthesis of medical images using efficient location-sensitive deep network," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds., vol. 9349 of Lecture Notes in Computer Science, pp. 677–684, Springer, Cham, 2015.
- [12] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, <https://arxiv.org/abs/1312.6114>.
- [13] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems* 27, vol. 2, pp. 2672–2680, 2014.
- [14] H. Larochelle and I. Murray, "The neural autoregressive distribution estimator," *Journal of Machine Learning Research*, vol. 15, pp. 29–37, 2011.
- [15] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, <https://arxiv.org/abs/1411.1784>.
- [16] A. B. L. Larsen, S. K. Sønderby, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," 2015, <https://arxiv.org/abs/1512.09300>.
- [17] D. Nie, R. Trullo, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. Collins, and S. Duchesne, Eds., vol. 10435 of Lecture Notes in Computer Science, Springer, Cham, 2017.
- [18] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017.
- [19] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.
- [20] Z. Yi, H. Zhang, and P. T. M. Gong, "DualGAN: unsupervised dual learning for image-to-image translation," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.
- [21] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, <https://arxiv.org/abs/1511.05644>.

- [22] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua, "CVAE-GAN: fine-grained image generation through asymmetric training," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017.
- [23] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," 2016, <https://arxiv.org/abs/1610.09585>.
- [24] B. B. Avants, C. L. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.
- [25] P. Viola and W. M. Wells, "Alignment of maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [26] Z. Tang, S. Ahmad, P. T. Yap, and D. Shen, "Multi-atlas segmentation of MR tumor brain images using low-rank based image recovery," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2224–2235, 2018.
- [27] Z. Tang, P. T. Yap, and D. Shen, "A new multi-atlas registration framework for multimodal pathological images using conventional monomodal normal atlases," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2293–2304, 2019.
- [28] D. Shen and C. Davatzikos, "HAMMER: hierarchical attribute matching mechanism for elastic registration," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.
- [29] G. Wu, M. Kim, Q. Wang, and D. Shen, "S-HAMMER: hierarchical attribute-guided, symmetric diffeomorphic registration for MR brain images," *Human Brain Mapping*, vol. 35, no. 3, pp. 1044–1060, 2014.
- [30] S. Reaungamornrat, T. D. Silva, A. Uneri et al., "MIND demons: symmetric diffeomorphic deformable registration of MR and CT for image-guided spine surgery," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2413–2424, 2016.
- [31] J. P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.
- [32] H. Knutsson and M. Andersson, "Morphons: segmentation using elastic canvas and paint on priors," in *IEEE International Conference on Image Processing 2005*, Genova, Italy, 2009.
- [33] M. F. Beg, M. I. Miller, A. Troune, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139–157, 2005.
- [34] C. Ceritoglu, K. Oishi, X. Li et al., "Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging," *NeuroImage*, vol. 47, no. 2, pp. 618–627, 2009.
- [35] M. Hernandez, M. N. Bossa, and S. Olmos, "Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows," *International Journal of Computer Vision*, vol. 85, no. 3, pp. 291–306, 2009.
- [36] M. Zhang, R. Liao, A. V. Dalca et al., "Frequency diffeomorphisms for efficient image registration," in *Information Processing in Medical Imaging. IPMI 2017*, M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P. T. Yap, and D. Shen, Eds., vol. 10265 of *Lecture Notes in Computer Science*, pp. 559–570, Springer, Cham, 2017.
- [37] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
- [38] G. Janssens, L. Jacques, J. O. de Xivry, X. Geets, and B. Macq, "Diffeomorphic registration of images with variable contrast enhancement," *International Journal of Biomedical Imaging*, vol. 2011, Article ID 891585, 16 pages, 2011.
- [39] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration - A deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.
- [40] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.
- [41] J. Fan, X. Cao, P. T. Yap, and D. Shen, "BIRNet: brain image registration using dual-supervised fully convolutional networks," *Medical Image Analysis*, vol. 54, pp. 193–206, 2019.
- [42] M. Jaderberg, K. Simonyan, and A. Zisserman, "Spatial transformer networks," 2015, <https://arxiv.org/abs/1506.02025>.
- [43] H. Knutsson and M. Andersson, "Implications of invariance and uncertainty for local structure analysis filter sets," *Signal Processing: Image Communications*, vol. 20, pp. 569–581, 2005.
- [44] J. Pettersson, H. Knutsson, and M. Borga, "Normalized convolution and morphons for non-rigid registration," in *SSBA 2006 Symposium on Image Analysis*, Umeå, Sweden, 2006.

## Research Article

# Attention Optimization Method for EEG via the TGAM

Yu Wu<sup>1</sup> and Ning Xie<sup>2</sup>

<sup>1</sup>Glasgow College, University of Electronic Science and Technology of China, 611731, China

<sup>2</sup>Center of Future Media, School of Computer Science and Engineering, University of Electronic Science and Technology of China, 611731, China

Correspondence should be addressed to Ning Xie; seanxiening@gmail.com

Received 27 March 2020; Accepted 7 May 2020; Published 18 June 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Yu Wu and Ning Xie. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the 21st century, noninvasive brain-computer interface (BCI) has developed rapidly, and brain-computer devices have gradually moved from the laboratory to the mass market. Among them, the TGAM (ThinkGear Asic Module) and its encapsulate algorithm have been adopted by many research teams and faculty members around the world. However, due to the limited development cost, the effectiveness of the algorithm to calculate data is not satisfactory. This paper proposes an attention optimization algorithm based on the TGAM for EEG data feedback. Considering that the data output of the TGAM encapsulate algorithm fluctuates greatly, the delay is high and the accuracy is low. The experimental results demonstrated that our algorithm can optimize EEG data, so that with the same or even lower delay and without changing the encapsulate algorithm of the module itself, it can significantly improve the performance of attention data, greatly improve the stability and accuracy of data, and achieve better results in practical applications.

## 1. Introduction

In recent years, a control system based on related concepts such as human mind, thought, and consciousness has developed rapidly. The difference between this system and traditional computerized control systems is that it is based directly on the human brain, using electroencephalogram (EEG) as the basis for data analysis. This class is also known as brain-computer interface (BCI), in which the human brain controls the corresponding system [1].

EEG was first discovered by Richard in 1875 while studying the brains of exposed rabbits. In 1924, Hans published the first paper on scalp signal which can trace to EEG. Although electroencephalography has been found for more than a hundred years, in earlier studies, because the brain's thoughts, minds, and other activities typically produced signals with much lower amplitude than the EEG, most of the signals were submerged in spontaneous potentials [2]. Based on the original EEG alone, the researchers were unlikely to get "specific sensory and complex cognitive information" about the subject [3]. This is a major difficulty in EEG analysis.

## 2. Technical Background

Recently, the breakthrough of biotechnology and the rapid development of computer technology make the EEG and brain-computer interface, which is in the interdisciplinary field, develop rapidly. With the discovery of key characteristics of EEG such as event-related potential (ERP) [4, 5], P300 potential [6, 7] and contingent negative variation (CNV) [8], the technology gradually moved from ideal to reality. In 1999, *Nature* published the first paper on brain-computer interface. As a landmark achievement at that time, this paper showed that slow cortical potential (CSP) could realize the construction of brain-computer interface for spelling function [9]. In the 21st century, EEG technology has achieved the evolution from multinode invasive to multinode noninvasive and then to single-node noninvasive. EEG devices are moving from the laboratory to the mass market.

Single-node portable EEG equipment is a prevailing research field and has great potential in the future. At the core of many of these kinds of devices is the TGAM (ThinkGear Asic Module) and its encapsulate algorithm, with the

difference being the Bluetooth version, the type of API, and how data is received by terminals. However, even though the single-node EEG device of the mind has been adopted by many research teams and faculty members around the world, due to the limited cost of technology development, the algorithm effect of calculating data is not satisfactory. There has been a lot of research using the TGAM, such as EEG control of wheelchairs to develop disability services [10], robotic hand controlled [11, 12], and brain training [13]. This device has a wide range of application fields. In addition to the traditional service equipment for the disabled, it has profound application potential and use value in the fields of game entertainment, VR environment, education [14], and human-computer interaction [15]. As early as 2014, there was a precedent for using the TGAM to test users' brain waves in games [16] which only detect the change of beta wave. However, the beta wave value given by the paper is quite different from the output value of the TGAM, so it can be seen that the modification module did not play a role in this experiment.

In particular, the application of EEG in VR is a hot topic in recent years. As early as 2012, the idea of building a hybrid system based on the advantages of EEG and VR appeared [17], and it can be used to optimize the sense of immersion in VR environment [18]. EEG must be an important part of the development of VR technology in the future.

The original TGAM was available in 2012, and despite some algorithm updates over the next eight years, the output data was still disappointing. The stability and accuracy of this module are far less than those of research-level equipment, and the existing open-source optimization method is quite feeble; whether official or private, there is no optimization algorithm for this widely used module, which is a pity for all the research projects using it.

### 3. The Design of Data Transmission via the TGAM

**3.1. Data Transmission Method.** The TGAM locks the wired transmission protocol from the inside and can only use Bluetooth for data transmission. The transmitted data is divided into large packets and small packets, which are all represented in a hexadecimal system in this paper. A packet, whether large or small, consists of three components: the packet header, the payload, and the checksum.

The small packets contain 8 bytes, one for every 2 ms. The big packets contain 36 bytes, with one big packet transferred after every 512 small packets. The data in the big packet (except for the data used for synchronization and checksum detection) is the result of calculating the raw wave value (raw data in short) in the previous 512 small packets and exporting it.

In the small packets, the value of the first two bytes are 0xAA for data synchronization, and the value of the third byte is 0x04, which means that the package is a small package with only 4 bytes of payload. The first three bytes are called packet headers, while data starts with the fourth byte (after which all data except the checksum is called Payload).

```

1  if (High <<8) + Low < 32768
2    raw_data <- ((High <<8) + Low)
3  else
4    raw_data <- (((High <<8) + Low)) -65536)

```

ALGORITHM 1: Raw data processing logic.

The fourth (0x80) and fifth (0x02) bytes represent the code and the length (2 bytes) of raw data, respectively. Sometimes, data like 0x80 and 0x02 are easily considered as packet header. However, it is clearly pointed out in the official manual that the checksum verifies all payloads, so 0x80 and 0x02 which are included in the checksum calculation formula are not packet header data, and the same is true for large packet data.

The two-byte length of the raw data is in the sixth (high eight bits) and seventh (low eight bits) bytes and is artificially set to transmit eight bits high and then eight bits low. Therefore, two bytes need to be spliced when the data is extracted, but the data needs to be further processed for the convenience of analysis. To obtain the new raw data with the absolute value within 32678, the data processing logic can be expressed as follows.

The last byte is the checksum, the module's own data detection mechanism. Add up all the payloads, invert, and then take the lower eight bits, so that the value and the eight-bit checksum are equal that means the packet is extracted correctly. That is Equation (1);

$$\text{Checksum} = \text{Low\_eight\_bits}(\text{Sum\_of\_payload XOR } 0xFFFFFFFF). \quad (1)$$

The ratio of the number of packets transferred to the number of packets with checksum errors is the packet loss rate. The official manual states that the packet loss rate below 10% will not affect the final result [19].

The big packet also starts with two 0xAA for data synchronization. The third byte is 0x20 for 32 payloads. The first three bytes are packet headers. The payload starts from the fourth byte. The specific meaning of the data is not described here. It is important to note that the eight types of brainwave data transmitted in the packet are all divided into three bytes (the lowest bit is transmitted first) and need to be manually spliced. The transmission characteristics of the wave value can be found in Yin et al.'s research [20]; the corresponding description will not be elaborated in this paper.

**3.2. Limitations of TGAM Encapsulate Algorithm.** The encapsulate algorithm mainly has three limitations.

First of all, it may be the problem of development cost that TGAM, as a single-node EEG device, does not have good enough antinoise and data analysis step. No matter, in the experiment of this paper or the experimental data in some articles [21], the attention data calculated by the module's own algorithm is very unsatisfactory in practical application, with large fluctuations, high delay, and low accuracy (Figure 1). Some research projects have used large



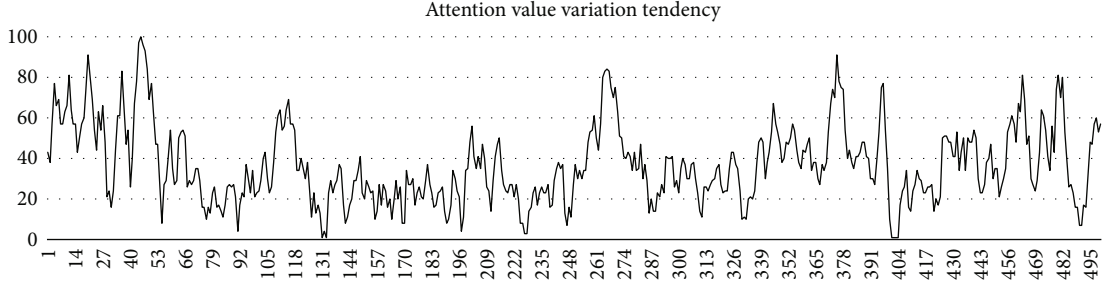


FIGURE 1: 500 consecutive concentration values were randomly selected from the test data.

amounts of data to get baselines for attention and meditation to improve accuracy [22], but the results have been extremely limited.

Tests have shown that the aperiodical dips of attention value are caused by skin movements in the prefrontal, which is usually caused by the subjects blinking their eyes. In addition, subtle changes in the subjects' facial expressions can lead to irregular changes in EEG signals and sometimes even long periods of constant values, as well as spikes or drops in the difference of values above 50 (which has a range from 0 to 100).

Additionally, although there are 32 bytes of payloads in the big package (24 bytes of eight kinds of wave value data), it is difficult to make regular analysis because the hardware is a single-node sampling. In that case, the focus of the data analysis is mainly attention and meditation [23]. Both of these values are integers from 0 to 100, but the criterion of meditation is not as practical as attention in most cases, which is the reason why most research work using the TGAM will only analyze attention.

When we tried to find the relationship between attention and meditation, we found that these two values which are an output by the encapsulate algorithm of the TGAM were actually divided into many parts artificially (Figure 2), and this conclusion derives from a data set which contains over ten thousand samples from the experiment. Lots of meaningless empty value areas will increase the analysis cost. Specifically, the empty value areas are null region data points that do not appear in Figure 2. In other words, theoretically, there are  $100 \times 100 = 10,000$  kinds of data, because attention and meditation are all integers from 0 to 100. However, a large part of the data will not appear in experiments, and these areas that will not appear are called null value areas. The lack of stability in the data is partly due to the fact that 100 numbers divide the level of attention too much. Finding the attention-meditation relationship is not the first of its kind; a research had analyzed the relationship and used the wave value to make data correction [24]. But perhaps because of the small amount of data, the study did not find that the data had been artificially divided.

Finally, the encapsulate algorithm does not have error detection steps, even if there are a checksum and poor signal, which is used to verify the accuracy of the payload. In particular, the checksum is a standard for developers to check whether the data they receive from a module is correct. In other words, as long as the data is extracted correctly, the checksum will never report an error (our experiment can

support this point). Poor signal is an integer between 0 and 200; 200 means that the developers cannot receive any signals by the wrong manner of wearing. This parameter is used to determine the quality of the device. Although the poor signal is 0, wrong data will still be output.

And there are three obvious problems with these two data (poor signal and checksum) that can be used for detection:

- (1) The output of a module does not consider checksum and poor signal. In other words, the data, such as attention, output of the encapsulate algorithm at each moment does not consider whether the checksum or poor signal at that moment is normal or not
- (2) After many tests, it can be seen that as long as the correct collection method is used, the checksum is almost always correct, which is of no essential help to data optimization. A total of 9698 groups of big packets were collected in one test, and none of them had incorrect checksum
- (3) Poor signal does not change continuously. In fact, there are only a few values such as 26, 51, 55, and 200, with a large number of null value areas. In addition, although the signal noise is zero in the case of data receiving, there are still unstable situations in Figure 1. It is obvious that this parameter is extremely limited

## 4. Attention Optimization Method

*4.1. The Pipeline of the Proposed Optimization Method.* Our algorithm optimizes two kinds of packages (big package and small package) to improve the performance of the data. The optimization process is summarized as follows (refer to Figure 3 for the overall optimization process):

- (1) Add checksum verification; that is, check whether the data value of the packet is consistent with the checksum and improves the accuracy of data collection
- (2) Cross-check the data with the previous big package. A big package with 36 bytes of data detects, but only attention and meditation will be detected
- (3) Because the eight kinds of wave value are a big integer with no practical significance which are defined personally, the result will be bad if dealt with directly. Consequently, use the Isolation Forest method to detect an outlier (abnormal data points) from a large



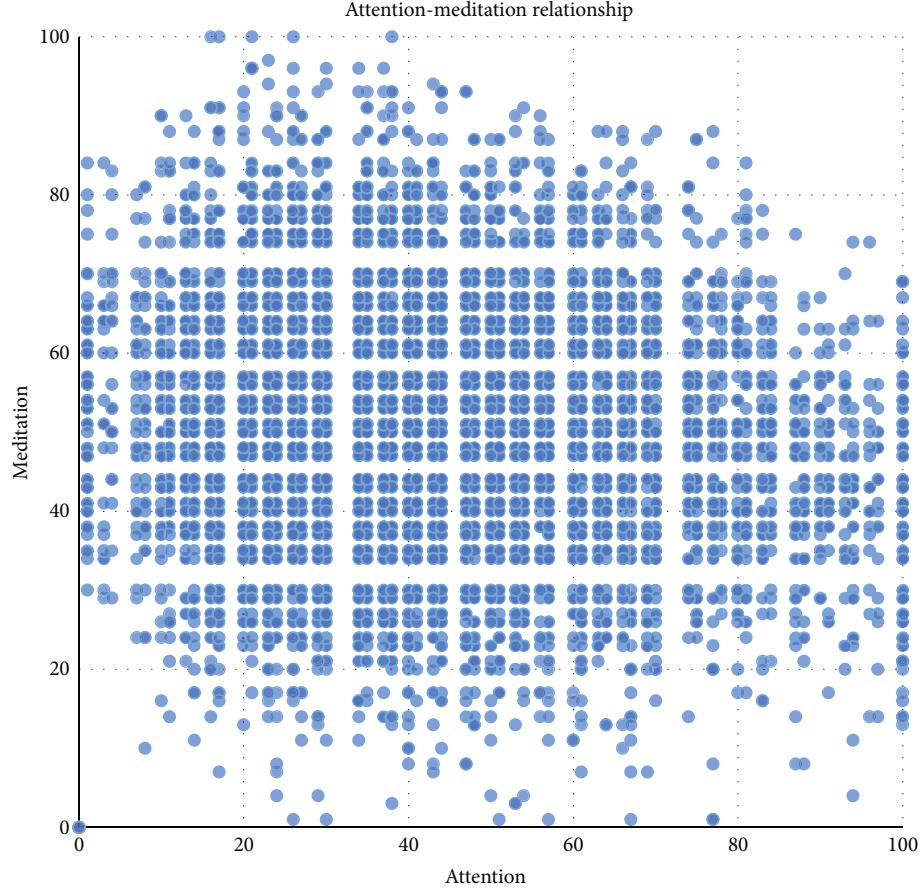


FIGURE 2: Encapsulate algorithm's output of attention-meditation relationship.

number of wave value which is collected from experiment, and then find out the upper bound of eight wave values after the regularization. If at the same time there are three or more values beyond the upper bound, the packet at this moment will be judged as a wrong package. In addition, Isolation Forest is also used to find the relationship between the attention value and the meditation value and draw the decision boundary of the value of concentration on the meditation value

- (4) Because the blink signal will cause the instability of the raw data, which is manifested as a sudden drop in the attention value, the blink can be detected by the shake change of the small packets' data. When the user blinks, the algorithm will stabilize the value of attention from abnormal range to normal range. Additionally, there is a link between how often people blink and how focused they are [25, 26]. Therefore, a step was set to detect the blink frequency and to further optimize the data to improve the return value of attention
- (5) The output of attention is step by step, which means two constantly updated attention values are used to make the data with strong shake much smoother

There are three most important steps of this optimization algorithm, and the reasons and principle will be introduced in the next sections:

- (1) Time interval setting in big packet data extraction
- (2) Blink feedback mechanism
- (3) Output classification (removal of meaningless null)

**4.2. Big Package Optimization via Isolation Forest.** Before optimization, big packets' data extraction is the basis of the whole optimization algorithm system. Through experiments and investigations, it is found that the use of various versions of the official development kit or open-source code often encountered difficulties in extracting big packets' data. This is because when extracting, a code needs to wait a certain amount of time between each byte to be extracted. It is really important. Big packets lose data in a continuous read because of the defects of the original encapsulate algorithm. A large number of experiments have proved that if the pause time is not set, a large packet of data is successfully extracted every 3-4 seconds in general, but in theory, the large packet data is one for every 1026 ms. In order not to affect the accuracy of reading data, the pause time is smaller than the byte transmission time of large packets, and our algorithm takes  $10\ \mu\text{s}$ .

After the checksum check is completed, cross-check with the previous packet is the second step of optimization. In a data set with more than 10,000 samples (under normal wearing conditions), no group of data has the same attention and meditation values as the previous group, and the two data are the same only when there is a problem in data collection.

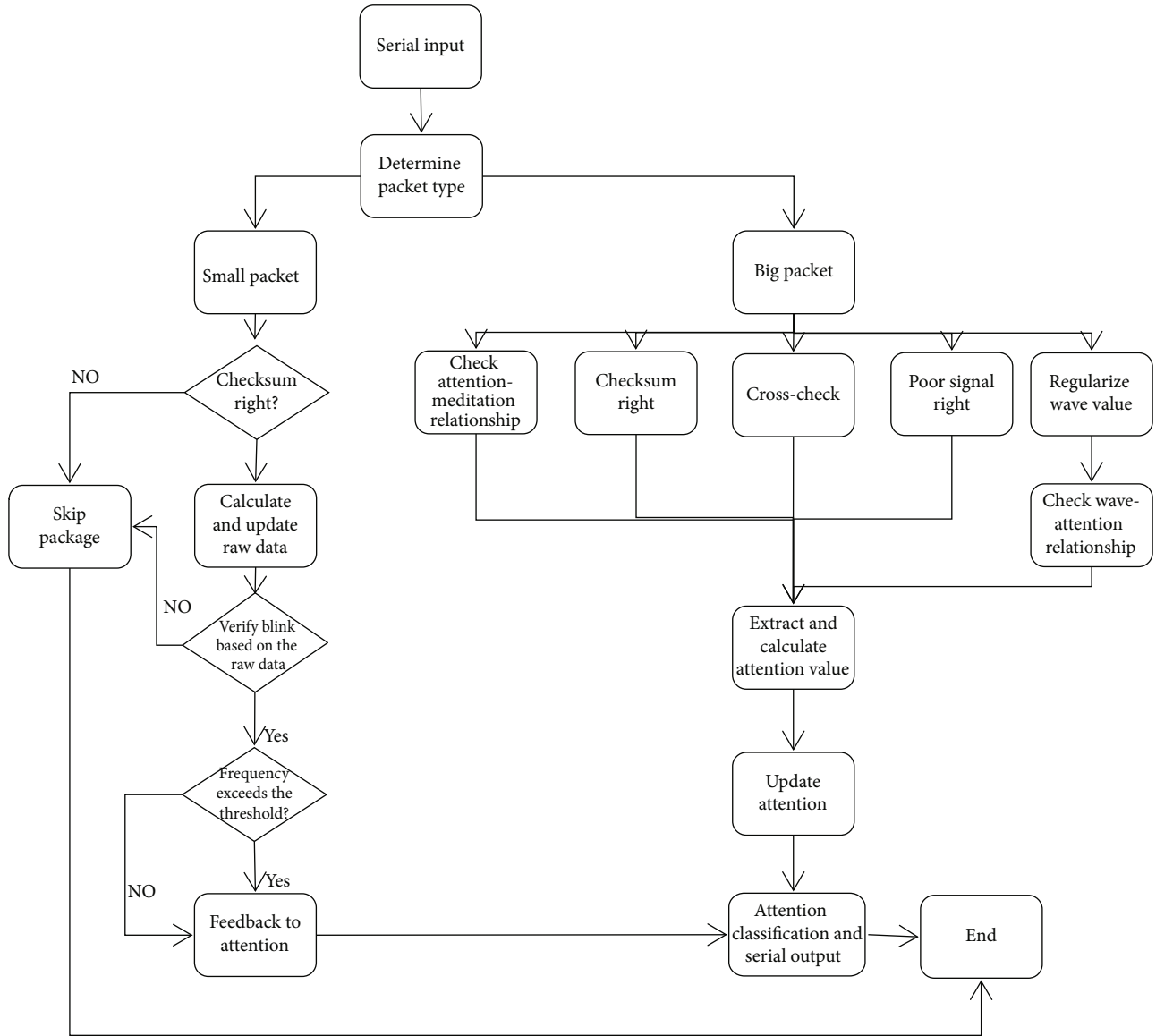


FIGURE 3: Flow diagram of optimization process.

Therefore, cross-check can effectively eliminate incorrect data due to hardware contact or noise problems. After that, the poor signal at that moment will be extracted, which is located in the fifth byte of big packets. Poor signal can directly reflect the wearing quality.

After that, there are two boundary detection mechanisms. Our algorithm uses Isolation Forest to detect abnormal points and outline the decision boundary. Isolation Forest is an unsupervised anomaly detection method, which is suitable for continuous data. Isolation Forest uses a binary search tree structure (also known as “isolated tree”) to isolate samples and detect outliers through the isolation of sample points, which is different from the anomaly detection algorithm that expresses the degree of alienation between samples by quantitative indexes such as distance and density [27]. Because most of the attention values in the TGAM are in the correct range, the number of outliers is small and there

is a significant difference from most samples. In that case, the isolation of sample points will cause the outliers to be isolated earlier; that is, the outliers will be closer to the root node, while the normal values will be further away from the root node. This is an important reason why this optimization algorithm uses Isolation Forest for upper- and lower-bound analysis.

Finding the relationship between meditation and attention using the Isolation Forest method (Figure 4) is the first boundary detection mechanism. In the figure, there are four colored lines representing the boundary drawn by data from four different situations, and the final average of the four was taken. The boundary in the figure is the same as the trend of boundary which is drawn artificially, which is enough to prove the rigor and accuracy of this machine learning method.

Another boundary detection, specifically, is the detection of eight wave value data (respectively, Delta, Theta,

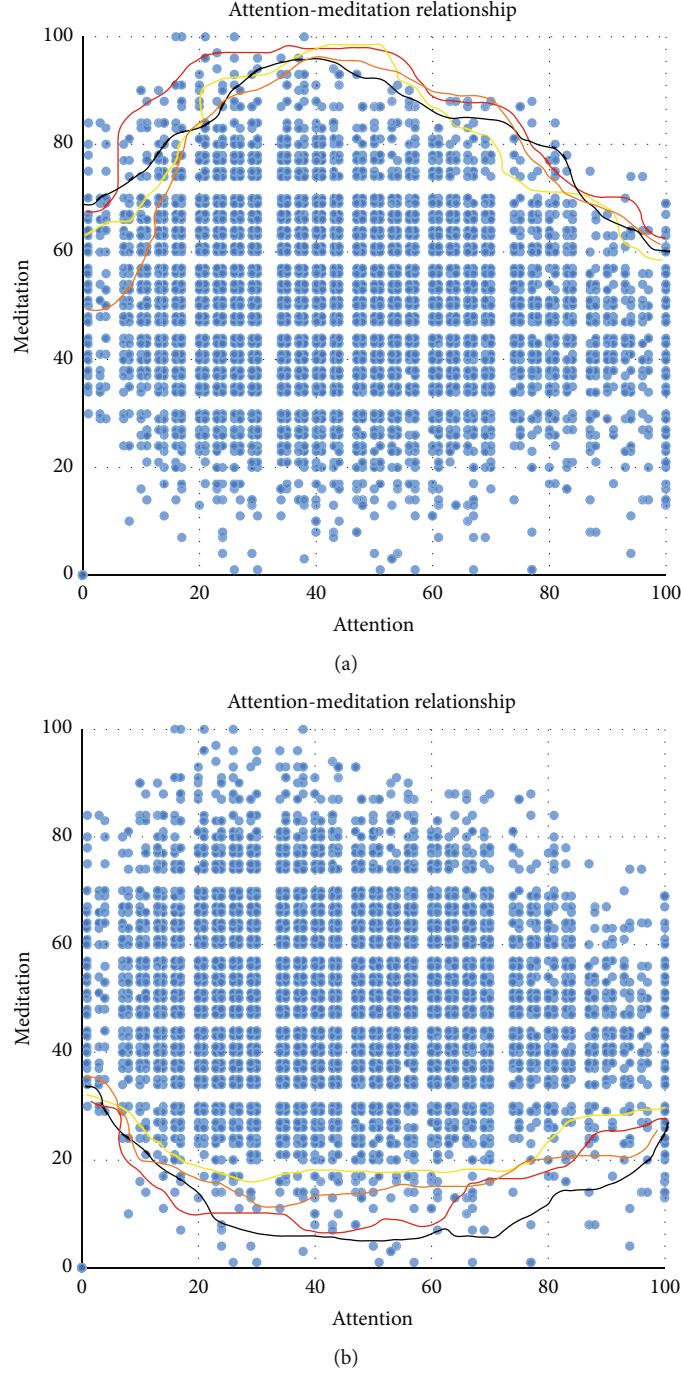


FIGURE 4: (a) Upper bound and (b) lower bound of attention-meditation relationship.

TABLE 1: Threshold of wave value after regularizing.

Wave name	Delta	Theta	LowAlpha	HighAlpha	LowBeta	HighBeta	LowGamma	MidGamma
Threshold value (regularization)	0.635	0.610	0.640	0.600	0.615	0.605	0.620	0.630

LowAlpha, HighAlpha, LowBeta, HighBeta, LowGamma, mid-gamma); if more than three wave values fall outside the normal upper bound, it is judged as wrong data.

Before the decision, the wave value will be regularized, because it is a meaningless huge integer, which is too difficult

to analyze. After trying standardization, normalization, regularization, and other methods and comparing with the untreated values, it is found that regularization is a method with obvious differentiation. Moreover, regularization will make the data fall between 0 and 1, and the upper and lower

```

1  if data_is_big_packet
2    while count < 36
3      data[count] <- extract_Byte
4      time_delay(0.01ms)
5      Verify Wave_value_Availability\
6        and Checksum and Last_Packet\
7        and Mediation-Attention_Relationship
8    if Blink
9      Attention <- Blink_Attention + data[32]
10   else
11     Attention <- data[32]
12   Output
13   Round_down((Attention + Last_Attention)/2)

```

ALGORITHM 2: Big packet optimization logic.

bounds of analysis are more convenient and effective. The results are as follows (Table 1).

The big packet optimization logic is shown below.

**4.3. Small Package Optimization with Region Segmentation of Blink Frequency.** The calculation method of the raw data in the small packets is mentioned in Section 3, and the range after parsing is between -32767 and 32768. Through the data set analysis, when the EEG signal is transmitted regularly, the raw data will be limited to  $\pm 550$ . The highest value is no more than 528, the lowest value is no less than -427, and most (over 95%) are within  $\pm 350$ .

The raw data output from the device in eight seconds were randomly extracted (Figure 5) in a total of 4143 groups, and the obvious wave peaks were caused by the blink moment (7 times) of the tester. It can be seen that the raw data is quite stable when a blink does not happen, and the blink activity can be monitored only by detecting the fluctuations of the raw data.

In this algorithm, 1000 is set as the difference of the raw data to detect the blink signal, and the huge fluctuation of the raw data represents the blink movement of the user. The blink frequency can be calculated by calculating the time difference between each two blinks, and the attention value can be improved by calculating the specific frequency value, so as to further optimize the attention value. The method of judging the blink of the raw data can be expressed as follows.

The Blink\_type above refers to the region where the blink frequency falls, which will be explained in more detail below.

Regarding the relationship between blinking and attention, Chen demonstrated that blinking not only affects the voltage of the cortex (that is, affects the EEG signal), but also negatively correlates the concentration with the blink frequency by comparing the blink frequency between the two groups of subjects during meditation and under normal conditions [26]. In the experiment, Chen concluded that relaxation and concentration resulted in a 3.8-6.0 times/min blink rate difference between the same subjects in different environments and at different times. In addition, Maffei and Angrilli have designed experiments that measure how quickly people blink when they complete tasks of different difficulty [25]. The harder and stranger the task, the less likely

participants were to blink. This study also supports the interesting idea that when people focus on one thing, their concentration increases, but their reaction time to other things decreases [28]. The same conclusion emerged in a study of computer users who blinked significantly less when they looked at a screen [29].

To sum up, it is feasible to optimize the attention data by blinking.

The blink frequency was divided into five parts for analysis. The frequency selection was derived from the experimental results of Chen. The specific distinction is in Table 2.

The reason for setting too small and too large frequency as an invalid zone is to effectively avoid collection errors caused by hardware or wearing problems. The fluctuations in raw data caused by the blink itself can induce attention to plumb. In most cases, the blink will bring an 8-20 drop in the next attention value and after that, the next value will become regular. Therefore, even if the frequency is not in the focus zone, "blink compensation" needs to be set to keep the data close to the previous packet.

According to the data analysis and experimental feedback of blink moment, a relatively lower feedback value of 10 is used as the cardinal number of blink compensation. Additionally, frequency and maximum values are used to constitute the entire blink optimization system.

It should be noted that after the blink signal is detected, the system does not immediately modify the attention value but waits for the next big packet to arrive before making corrections. Doing this, relative to the direct update, fundamentally avoids the occurrence of the next abnormal attention value.

**4.4. Tuning on the Attention Level.** As mentioned in Section 4.1, there is no specific difference between each individual attention-meditation interval, and meaningless null value area will increase the analysis cost. The instability of the data is also partly due to the fact that 100 numbers divide the attention level too much. In the end, attention classification is a necessary step to optimize data, enhance accuracy, and reduce analysis costs.

It is different from the numerical classification in similar articles [30]. The algorithm classification system is essentially reflecting the TGAM itself way of classification. As shown in Figure 3, if not done classification in advance, cannot explain why the null value area so coincidence slices the image into many areas. Therefore, we remove the null value areas and divide the rectangular regions which are composed of blue data points in Figure 3. In this way, it can be divided into seven levels (Table 3) according to the abscissa (attention). The reliability of this attention level is proved in Section 5; it is crucial.

Among them, 0-4 and 98-100 actually have the attention value, but the algorithm still abandons them. For the former, this is because only 147 (1.52%) of the 9698 data sets are between 0 and 4, and more than half of them have nonzero poor signal. In addition, there are only two forms of these data: one is the continuous occurrence of multiple times with a huge difference with the data before and after. The other is the occurrence of only once with no connection with the data

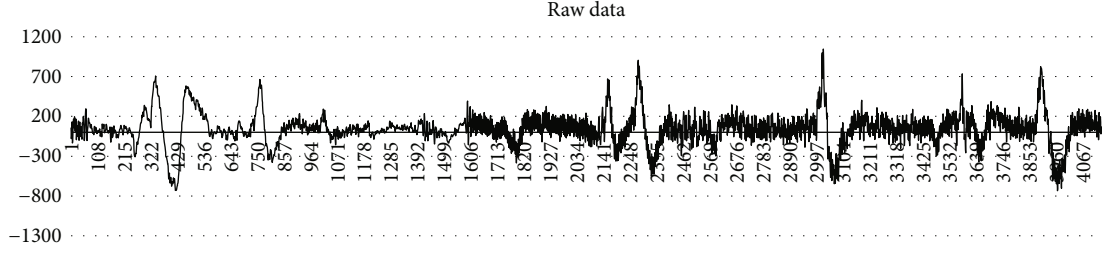


FIGURE 5: The raw data fluctuates with blink.

```

1  if data_is_small_packet
2    extract: sixth_and_seventh_Bytes
3    transform two_Bytes to Raw_data
4    Raw_data[0] <- data[1]
5    Raw_data[1] <- Raw_data
6    if Raw_data[ ] stop_increase and >528
7      save Raw_data[1] as Raw_data_High
8    if Raw_data[ ] stop_decrease and < -427
9      save Raw_data[1] as Raw_data_Low
10   if Raw_data_High - Raw_data_Low >1000
11     Cal_Frequency
12     Record_Blink_type

```

ALGORITHM 3: Small packet optimization logic.

TABLE 2: Region segmentation of blink frequency.

Frequency (second/once)	Region	Attention management
<2	Invalid area	None
2-4.81	Normal area	Plus 10
4.81-6.51	Focus area	Plus (10 + frequency $\times$ 0.8)
6.51-8.33	Highly focus area	Plus (10 + frequency $\times$ 1.5)
>8.33	Invalid area	None

before and after as well. For the latter, only 100 will appear in the interval of 98-100, accounting for only 1.15% of the total number. The data will appear for several times in a row, and the difference with the previous data is more than 20 in most of times.

Combined with the fact that attention data rarely appear continuously and do not plunge or surge under regular circumstances, two areas can be abandoned.

## 5. Experimental Results

In order to prove the reliability of this optimization algorithm, the following experiment is designed.

Record the change of the attention value within a certain period of time in two specific application scenarios, and select game and study as test scenarios to correspond to the situations of intermittent concentration and high concentration, respectively. The time is about 20 minutes. Without any

processing, about 1170 groups of attention value (one for every 1026 ms) will be collected, and 50 groups will be randomly extracted for data analysis.

Subjects should not have a history of mental illness or have taken psychiatric drugs and no history of epilepsy and other diseases that have a greater impact on brain fluctuations. On the night before the experiment, they were told to have a good rest and not to drink stimulants the day or the day before the experiment.

For the experimental environment, ensure that it is bright and quiet and try to choose a comfortable environment for the tested personnel. The volunteers were required to close their eyes and rest before starting the test. The EEG data reached a low noise and stable state, and then, the test of the corresponding scene was started at the command of the staff.

The data in Figure 6 is all after big packet optimization, where Attention\_Cal is the result of small packet optimization (blink feedback). Obviously, even without small packet optimization, the attention value is still more stable and accurate than the data in Figure 1. Big packet optimization has eliminated most of the wrong data caused by hardware, microexpressions, and environment, while small packet optimization makes the change of attention more gradual and precise, with almost no abnormal sharp rise or plunge.

Figure 7 shows the attention levels of situations. After grading, it can be seen clearly that the overall attention in study is much higher than that in game. And the grade distribution itself also conforms to the principle of normal distribution, so it can be seen that the results of this algorithm conform to the expected and subjective expectations of the volunteers. No change across two or more levels, or other abnormal changes, can be a stable reflection of the trend of attention value.

## 6. Conclusion

There are three most important steps of this algorithm, which are as follows: the time interval of the big packet data extraction setting, blink feedback mechanism, and classified output based on the removal of meaningless nulls.

The beneficial effect of this algorithm is to improve the performance of the attention value output from the TGAM without improving the real-time data transmission delay. The adverse effect of blink on EEG data is optimized and in turn feedbacks the degree of attention through the blink frequency, which greatly improves the stability and



TABLE 3: Attention classification method.

Attention level	Null value region	Boundary	Attention range	Meditation upper boundary	Meditation lower boundary
0	0-4	5-6	0-4		0
			7-8	74	28
1	9, 12, 15	18-19	10-11	77	26
			13-14	83	23
			16-17	87	19
			20-21	90	16
2	22, 25, 28	31-33	23-24	93	14
			26-27	97	11
			29-30	97	10
			34-35	97	10
3	36, 39, 42	45-46	37-38	97	13
			40-41	94	10
			43-44	94	10
			47-48	91	10
4	49, 52, 55	58-59	50-51	88	10
			53-54	88	10
			56-57	87	10
			60-61	84	10
5	62, 65, 68	71-73	63-64	84	10
			66-67	81	10
			69-70	81	16
			74-75	78	20
6	76, 79, 82	85-86	77-78	78	20
			80-81	75	24
			83-84	69	27
			87-88	69	23
7	89, 92, 95	98-100	90-91	67	27
			93-94	64	29
			96-97	58	33

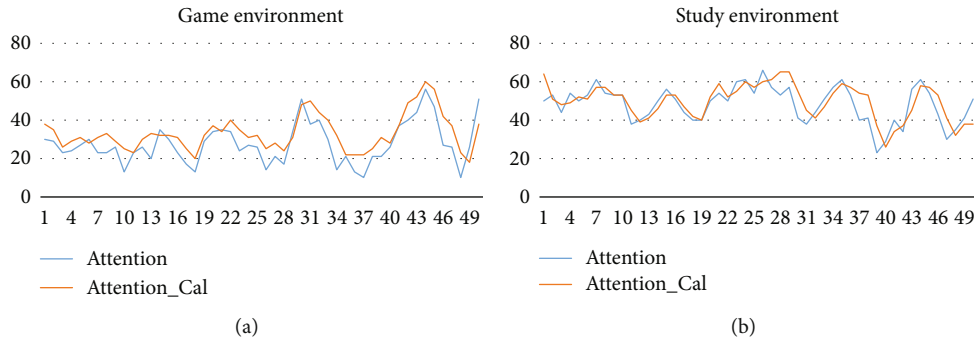


FIGURE 6: 50 groups of attention value randomly extracted in (a) game and (b) study (scenarios after big packet optimization).

accuracy. Although the effect is significant, there are still two important problems.

First, EEG data is not as objective as traditional data. For example, in the magnitude of the velocity, we only need to calculate the ratio of displacement to time. However, as the objective reflection of thinking, the strong subjectivity

and difficulty in describing thinking doomed the researchers to set subjective consciousness as the benchmark and create the analysis method of objective EEG data. Since the discovery of EEG in the 20th century, there has been a contradiction between objective data and subjective consciousness. The purpose of analyzing objective data is to

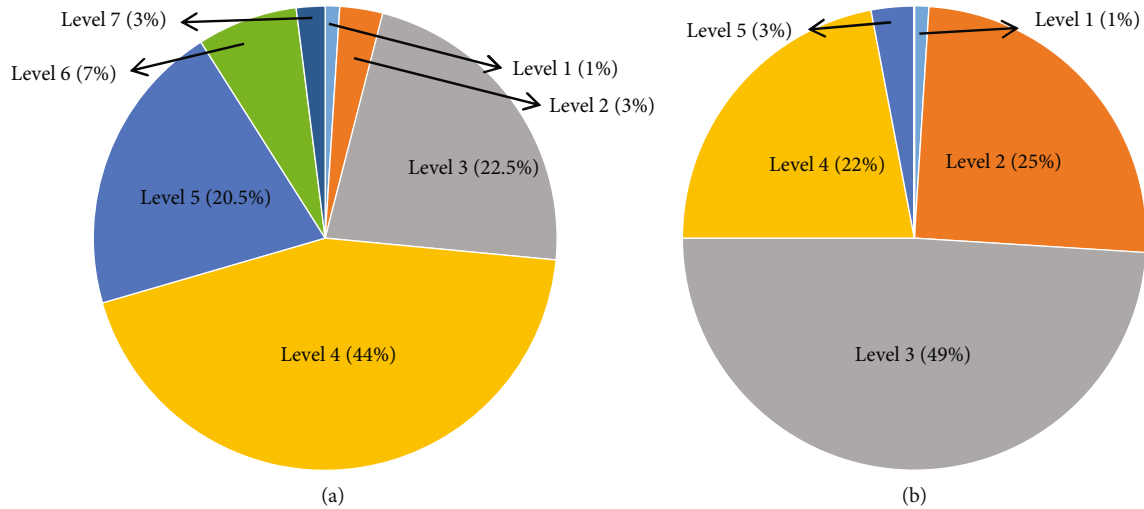


FIGURE 7: Proportion of the seven levels of attention in (a) study and (b) game.

reflect subjective consciousness, but is the existing objective data reliable enough to reflect real subjective consciousness? In the process of compiling and improving this algorithm, although the data optimization model has been fitted to the subjective consciousness of the testers as much as possible, the effect is not 100% satisfactory. This is not only a problem of this optimization algorithm but also a common problem of all EEG devices.

Second, because the encapsulate algorithm of the module is unknown, wave value analysis is almost impossible to achieve. Although we have tried to find the relationship between the eight wave value data and the attention, meditation, and raw data, we have failed in the case that the raw data analysis method is sealed. The TGAM itself is a product for developers, with more open-source data and more flexibility than a purely commercial EEG device. We believe that if the raw data analysis algorithm can also be open source, our optimization can be further improved to get better results.

Finally, the optimization algorithm is the improvement of software. In the field of electronics, the software and hardware should be combined to play the most important role. However, in the case of limited hardware conditions, this optimization algorithm also has certain practicability and promotion value.

## 7. Future Work

No matter how good the algorithm is, it needs to be verified by practical application [31], and it should be improved and perfected continuously in this process. So our next step is to explore the possibilities of our algorithm in application scenarios that require EEG analysis. Especially in VR applications, we believe that EEG analysis will make a difference.

EEG and BCI technologies are beginning to take off, and it is believed that such research will yield more brilliant results.

Furthermore, we would like to apply EEG into a more realistic practical research framework with crossmodal [32, 33] and multitask [34] in action recognition [31] and HCI in VR [35].

## Data Availability

Please contact us by email if any data used in the manuscript is demanded.

## Conflicts of Interest

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

## Acknowledgments

The study is funded by the General Program of Sichuan provincial science and Technology Department (2018JY0528). This work is part of the research supported by the National Nature Science Foundation of China under Grant No. 61602088 and No. 61976156 and the Fundamental Research Funds for the Central Universities No. Y03019023601008011. Additionally, the authors thank the Glasgow College, UESTC for the assistance and support.




## References

- [1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks et al., "Brain-computer interface technology : a review of the first international meeting," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 2, pp. 164–173, 2000.
- [2] J. J. Vidal, "Toward direct brain-computer Communication," *Annual Review of Biophysics & Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.
- [3] W. Xing-Yu, J. Jing, Z. Yu, and W. Bei, "Brain control: human-computer integration control based on brain-computer interface approach," *Acta Automatica Sinica*, vol. 39, no. 3, pp. 208–221, 2013.
- [4] Z. Lin and Z. Huang, "Research on event-related potentials in motor imagery BCI," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6, Shanghai, 2017.
- [5] E. Ventouras, M. Moatsos, C. Papageorgiou, A. Rabavilas, and N. Uzunoglu, "Independent component analysis applied to the P600 component of event-related potentials," *The 26th Annual*

- International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004, pp. 80–83, San Francisco, CA, 2004.
- [6] S. Sutton, M. Braren, J. Zubin, and E. R. John, “Evoked-potential correlates of stimulus Uncertainty,” *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965.
  - [7] J. Jin, E. W. Sellers, and X. Wang, “Targeting an efficient target-to-target interval for P300 speller brain-computer interfaces,” *Medical & Biological Engineering & Computing*, vol. 50, no. 3, pp. 289–296, 2012.
  - [8] W. G. Walter, R. Cooper, V. J. Aldridge, W. C. McCallum, and A. L. Winter, “Contingent Negative Variation : An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain,” *Nature*, vol. 203, no. 4943, pp. 380–384, 1964.
  - [9] N. Birbaumer, N. Ghanayim, T. Hinterberger et al., “A spelling device for the paralysed,” *Nature*, vol. 398, no. 6725, pp. 297–298, 1999.
  - [10] Z. Su, X. Xu, J. Ding, and W. Lu, “Intelligent wheelchair control system based on BCI and the image display of EEG,” *IEEE Advanced Information Management, Communications, Electronic & Automation Control Conference. IEEE*, 2016.
  - [11] O. A. Ruşanu, L. Cristea, M. C. Luculescu, and S. C. Zamfira, “Experimental model of a robotic hand controlled by using NeuroSky Mindwave Mobile headset,” *2019 E-Health and Bioengineering Conference (EHB)*, 2019pp. 1–4, Iasi, Romania, 2019.
  - [12] K. Banumalar, “Neural-based control of a prosthetic arm using Matlab,” *International Journal of Pure and Applied Mathematics*, vol. 118, pp. 411–417, 2018.
  - [13] T. Chaipakornwong, S. C. Djohan, and P. Sittirapaporn, “Brain exercise in elderly people by Neurosky Smarter Kit,” *2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2019pp. 490–493, Pattaya, Chonburi, Thailand, 2019.
  - [14] D. J. Lancheros-Cuesta, J. L. R. Arias, Y. Y. Forero, and A. C. Duran, “Evaluation of e-learning activities with NeuroSky MindWave EEG,” *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, 2018, pp. 1–6, Caceres, 2018.
  - [15] D. Bright, A. Nair, and D. Salvekar, “EEG-based brain controlled prosthetic arm,” *Advances in Signal Processing. IEEE*, 2016.
  - [16] J. Cheng, G. Mabasa, and C. Oppus, “Prolonged distraction testing game implemented with ImpactJS HTML5, Gamepad and Neurosky,” *2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2014, pp. 1–6, Palawan, 2014.
  - [17] S. B. Badia, H. Samaha, A. G. Morgade, and P. F. M. J. Verschure, “Exploring the synergies of a hybrid BCI - VR neurorehabilitation system,” *2011 International Conference on Virtual Rehabilitation*, 2011, pp. 1–8, Zurich, Switzerland, 2011.
  - [18] A.-M. Brouwer, J. van der Waa, and H. Stokking, “BCI to potentially enhance streaming images to a VR headset by predicting head rotation,” *Frontiers in human neuroscience*, vol. 12, 2018.
  - [19] *NeuroSky.Mindset Instruction Manua*, NeuroSky Co., 2009.
  - [20] L. Yin, C. Zhang, and Z. Cui, “Experimental research on real-time acquisition and monitoring of wearable EEG based on TGAM module,” *Computer Communications*, vol. 151, pp. 76–85, 2020.
  - [21] L. Zhang, Q. Lv, and Y. Xu, “Single channel brain-computer interface control system based on TGAM module,” *2017 10th International Congress on Image and Signal Processing, Bio-Medical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–5, Shanghai, 2017.
  - [22] D. Sulisetyo Nugroho and I. Fahrudi, “An application real-time acquiring EEG signal from single lead electrode to recognize brain activity using Neurosky sensor,” *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2019, pp. 400–404, Semarang, Indonesia, 2019.
  - [23] O. Chinbat and J. Lin, “Prosthetic arm control by human brain,” *Proceedings of the International Symposium on Computer, Consumer and Control (IS3C)*, 2018, Taiwan, 2018.
  - [24] B. Ülker, M. B. Tabakcıoğlu, H. Çizmeci, and D. Ayberkin, “Relations of attention and meditation level with learning in engineering education,” *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2017, pp. 1–4, Targoviste, 2017.
  - [25] M. Antonio and A. Alessandro, “Spontaneous eye blink rate: an index of dopaminergic component of sustained attention and fatigue,” *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 123, pp. 58–63, 2018.
  - [26] S. H. Chen, “Attention measurement by eye blinking for meditation practitioner,” *Advanced Science and Industry Research Center.Proceedings of 2015 International Conference on Automation, Mechanical and Electrical Engineering(AMEE 2015)*, pp. , 20151095–1100, 2015.
  - [27] U. Santosuosso, A. Cini, and A. Papini, “Tracing outliers in the dataset of Drosophila suzukii records with the Isolation Forest method,” *Journal of Big Data*, vol. 7, no. 1, 2020.
  - [28] M. I. Posner, “Orienting of attention,” *Quarterly Journal of Experimental Psychology*, vol. 32, no. 1, pp. 3–25, 2007.
  - [29] Y. Yaginuma, H. Yamada, and H. Nagai, “Study of the relationship between lacrimation and blink in VDT work,” *Occupational Ergonomics*, vol. 33, no. 6, pp. 799–808, 1990.
  - [30] K. R. Trivedi and R. A. Thakker, “Brainwave enabled multi-functional, communication, controlling and speech signal generating system,” in *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp. 4889–4893, Chennai, 2016.
  - [31] Y. Ji, Y. Zhan, Y. Yang, X. Xu, F. Shen, and H. T. Shen, “A Context knowledge map guided coarse-to-fine action recognition,” *IEEE Transactions on Image Processing*, vol. 29, pp. 2742–2752, 2020.
  - [32] B. Wang, Y. Yang, X. Xu, A. Hanjalic, and H. T. Shen, “Adversarial cross-modal retrieval,” *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017*, 2017, pp. 154–162, Mountain View, CA, USA, 2017.
  - [33] H. T. Shen, L. Liu, Y. Yang et al., “Exploiting subspace relation in semantic labels for cross-modal hashing,” *IEEE Transactions on Knowledge and Data Engineering*, p. 1, 2020.
  - [34] Y. Yang, Z. Ma, Y. Yang, F. Nie, and H. T. Shen, “Multitask spectral clustering by exploring intertask correlation,” *IEEE Transactions on Cybernetics*, vol. 45, no. 5, pp. 1069–1080, 2015.
  - [35] H. Liao, N. Xie, H. Li et al., “Data-driven spatio-temporal analysis via multi-modal zeitgebers and cognitive load in VR,” *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020, pp. 473–482, Atlanta, GA, USA, 2020.

## Research Article

# Detection of Solitary Pulmonary Nodules Based on Brain-Computer Interface

Shi Qiu <sup>1</sup>, Junjun Li,<sup>2</sup> Mengdi Cong <sup>3</sup>, Chun Wu,<sup>4</sup> Yan Qin,<sup>4</sup> and Ting Liang <sup>2,5</sup>

<sup>1</sup>Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China

<sup>2</sup>Department of Radiology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China

<sup>3</sup>Department of Computed Tomography and Magnetic Resonance, Children's Hospital of Hebei Province, Shijiazhuang 050031, China

<sup>4</sup>Beijing Hi-Tech Institute, Beijing 00085, China

<sup>5</sup>Department of Biomedical Engineering, The Key Laboratory of Biomedical Information Engineering of The Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710061, China

Correspondence should be addressed to Mengdi Cong; [congmd\\_hbpch@hotmail.com](mailto:congmd_hbpch@hotmail.com) and Ting Liang; [liangting31500@126.com](mailto:liangting31500@126.com)

Received 15 April 2020; Accepted 26 May 2020; Published 15 June 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Shi Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Solitary pulmonary nodules are the main manifestation of pulmonary lesions. Doctors often make diagnosis by observing the lung CT images. In order to further study the brain response structure and construct a brain-computer interface, we propose an isolated pulmonary nodule detection model based on a brain-computer interface. First, a single channel time-frequency feature extraction model is constructed based on the analysis of EEG data. Second, a multilayer fusion model is proposed to establish the brain-computer interface by connecting the brain electrical signal with a computer. Finally, according to image presentation, a three-frame image presentation method with different window widths and window positions is proposed to effectively detect the solitary pulmonary nodules.

## 1. Introduction

Pulmonary nodules are the main pulmonary lesions. Malignant pulmonary nodules may be transformed into lung cancer, which is a serious threat to human health [1, 2]. It needs many years of clinical practice for professional doctors to make accurate diagnosis. Recently, with the development of human brain research, the research of a brain-computer interface (BCI) has become more and more popular. Vansteensel et al. [3] realize cognitive control based on BCI. Tan and Nijholt [4] establish a human-computer interaction channel. Jin et al. [5] optimize the representation of BCI stimulation targets. Myrden et al. [6] use BCI technology to realize bilateral transcranial Doppler ultrasound. The validity of P300 in an EEG signal is verified by Duvinage et al. [7]. Lin and Yang [8] establish an EEG and blinking relationship to control wheelchair movement. Do et al. [9] apply BCI

technology to gait correction. Gu et al. [10] establish a semi-supervised network to realize BCI online. Sun and Zhou [11] review the development of BCI and predict the development of EEG feature extraction and classification. Mikołajewska and Mikołajewski [12] establish a brain-computer interface from the perspective of children for analysis. Nijholt [13] introduces a competition and coordination mechanism to realize BCI learning. Kumar et al. [14] extract features from EEG signals and apply them to the medical field. Jeunet et al. [15] propose the standard of BCI sequence. Liu et al. [16] establish a correlation model between the P300 signal and time to realize BCI. Thomas et al. [17] establish a deep learning network to realize BCI learning. Dong et al. [18] establish an SVM algorithm to analyze multiple EEG data. Guy et al. [19] use P300 to analyze literacy signals to assist patients. Schwemmer et al. [20] use a deep neural network to simulate people's expectations of events. Abbasi [21]

serves the paralyzed through BCI. Wan et al. [22] analyze the performance of BCI from the perspective of EEG. Sengupta et al. [23] establish a semisupervised neural network to realize the medical auxiliary treatment system. Wang et al. [24] propose an algorithm to enhance the useful signal strength of EEG. All the above algorithms are studied from different layers, and good results are obtained. However, there are few researches in the field of image recognition based on BCI.

Currently, the main problems and difficulties of image recognition based on BCI are as follows: (1) Based on the brain structure, it is difficult for people to stay focused for a long time, making EEG information overlapping. (2) There is a large amount of EEG data, so it is difficult to select effective features. (3) Because of the particularity of the medical image, how to build an effective response mechanism is an important issue.

In view of the above problems and difficulties, in this paper, we propose a detection algorithm of solitary pulmonary nodules based on a brain-computer interface. (1) The time-frequency feature fusion model is constructed to enhance the signal identification. (2) A multifeature network based on deep learning is proposed. (3) Through the training of doctors and ordinary people, an effective response mechanism is proposed.

## 2. Algorithm

To solve the above problems, we design the algorithm flow as shown in Figure 1. First, the image is blocked and displayed according to specific rules. Then the time-frequency feature fusion model is constructed to enhance the signal identification. An effective feature extraction method based on the stacking fusion model is proposed. A complete brain-computer interface is constructed to detect solitary pulmonary nodules.

**2.1. Time-Frequency Feature Fusion.** EEG signals have multiple frequencies, from which useful information needs to be extracted. The MEMD (multivariable electromagnetic mode decision) algorithm can ensure that signals from different sources get decomposition results matching each other in terms of quantity and frequency. It has advantages in processing multichannel EEG signals [25].

MEMD can expand the single variable EMD in many dimensions, realize the joint analysis of multiple vibration components of high-dimensional signals, and avoid the modal aliasing problem of standard EMD. The specific algorithm is described as follows.

**Data sampling:** calculate the set  $\{P^x(t)\}_{k=1}^K$  of mappings of the input signal  $\mathbf{V}$  along the direction vector  $\mathbf{x}$ . Find the time point  $t^x$  corresponding to the maximum value in  $\{P^x(t)\}_{k=1}^K$ . Obtain the multivariate envelope curve  $\{e^x(t)\}_{k=1}^K$ .

Calculate the mean value of envelope curve:  $m(t)$ . Calculate  $d(t) = x(t) - m(t)$  until stop if  $d(t)$  meets IMF criteria. Otherwise, it will iterate.

Hilbert transform [26] is applied to each channel of the obtained brain-computer signal to decompose the signal:

$$Y_k(t) = \frac{1}{\pi} \kappa \int_{-\infty}^{\infty} \frac{u_k(t')}{t - t'} dt, \quad (1)$$

$$Z_k(t) = u_k(t) + jY_k(t), \quad (2)$$

where  $\kappa$  is the Cauchy value and  $u(t)$  is the input signal. Rewrite Equation (2) as follows:

$$Z_k(t) = a_k(t)e^{j\theta_k(t)}, \quad (3)$$

where  $a_k(t)$  is the amplitude and  $\theta_k(t)$  is the phase.

$$\begin{cases} \theta_k(t) = \arctan \left( \frac{Y_k(t)}{u_k(t)} \right), \\ w_k(t) = \frac{d\theta_k(t)}{dt}. \end{cases} \quad (4)$$

Each component of the EEG signal is expressed as

$$x(t) = \text{Re} \left\{ \sum_{k=1}^K a_k(t)e^{j\theta_k(t)} \right\}. \quad (5)$$

The Hilbert spectrum is defined as

$$H(w, t) = \text{Re} \left\{ \sum_{k=1}^K a_k(t)e^{j \int w_k(t) dt} \right\}. \quad (6)$$

Solve the Hilbert energy spectrum (IES) and the Hilbert marginal spectrum (MS).

$$\begin{cases} \text{IES}(t) = \int_{w_1}^{w_2} H^2(w, t) dw, \\ \text{MS}(w) = \int_0^T H(w, t) dt, \end{cases} \quad (7)$$

where  $w_1$  and  $w_2$  correspond to different frequency bands of EEG.

Calculate the mobility parameter Mob in the Hjorth parameter.

$$\text{Mob} = \sqrt{\frac{\text{Var}(y'(t))}{\text{Var}(y(t))}}. \quad (8)$$

The mean value and standard deviation are calculated as the feature  $F_t$  in the time domain, and the feature  $F_s$  in the frequency domain is calculated in the same way.

Sample entropy is introduced to characterize the nonlinear dynamic coupling characteristics of EEG signals.



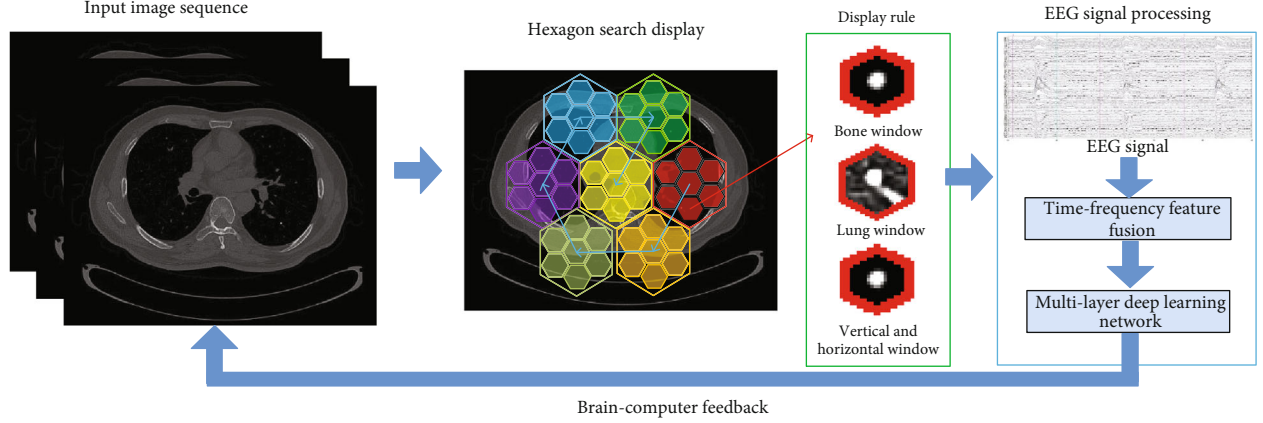


FIGURE 1: The algorithm flow chart.

The input sequences  $u(i)$  and  $v(i)$  are reconstructed into vectors:

$$\begin{cases} \mathbf{X}(i) = [u(i), \dots, u(i+m-1)], \\ \mathbf{Y}(i) = [v(i), \dots, v(i+m-1)], \\ i \in [1, N-m+1]. \end{cases} \quad (9)$$

Calculate the distance

$$d(i, j) = |u(i+k) - v(j+k)| \quad k=0 \dots m-1, \quad (10)$$

and statistical proportional mean

$$\begin{aligned} B^m(r) &= \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r), \\ B_i^m(r) &= \frac{N(d \leq R)}{N(d)}, \end{aligned} \quad (11)$$

where  $N(\cdot)$  is the number of qualified pixels and  $R$  is the conditional threshold. The cross sample entropy can be defined as

$$C(m, r) = -\ln \left( \frac{B^{m+1}(r)}{B^m(r)} \right). \quad (12)$$

Characterize the relative complexity of two sequences.

To characterize the phase coupling relationship, we calculate

$$Q = \frac{1}{N} \left| \sum_i^N \exp(j(\theta_1 - \theta_2)) \right|. \quad (13)$$

Similarly, calculate the frequency locking value:

$$S = \frac{1}{N} \left| \sum_i^N \exp(j(w_1 - w_2)\Delta t) \right|, \quad (14)$$

where  $N$  is the sampling frequency,  $\theta_1$  and  $\theta_2$  are the phase

information, and  $w_1$  and  $w_2$  are the frequency information. The single channel time-frequency characteristics of the EEG signal are as follows:

$$F = \{F_t, \cdot F_t, \cdot Q, \cdot S\}. \quad (15)$$

$F$  is used as an EEG signal feature for the following calculation.

**2.2. Multilayer Deep Learning Network.** An EEG signal has many signals, so it is difficult to extract useful information by a single learning framework. A stacking algorithm [27] can realize the efficient utilization of training data. The main idea of the algorithm is to train several different basic learners based on the initial training data set and then generate a new data set to train the next layer of learners according to the output of the primary learner as the input. The tag value of each training data is unchanged. In practice, in order to prevent overfitting, we often use cross validation or leave one method to generate the training samples of the next level of the learner with the samples not used in the training of the basic learner, and the basic learner uses different learning algorithms for generation.

In order to improve the recognition ability and generalization of the model, we use five simple heterogeneous learning devices, including SVM, random forest [28], logistic regression [29], KNN [30], and AdaBoost [31] to build a two-tier stacking integrated learning model.

The proposed algorithm in this paper adopts the deep learning framework. The training set and test set are very important. The ratio of the training set and the test set is 1 : 1.

Considering the phenomenon of overfitting, the training set of the secondary learners is obtained by using the method of half-fold cross validation, that is, each primary learner is used to train the current compromised training set in each compromise, and then the output of the current compromised verification set is predicted. When the cross validation is completed, the mapping of all the data in the original training set is completed and the secondary training set is generated.

Take the average value after each predicted test set. The execution process of the same primary learner cross validation is shown in Figure 2. First, the original training set is

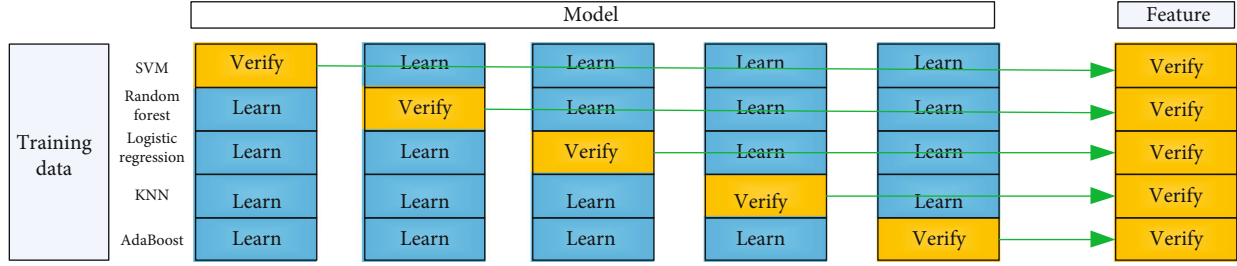


FIGURE 2: Deep learning network based on multiple features.

divided into five parts. In each iteration, four parts of the data are taken as the training set in turn, and the primary learner is trained. Then, the remaining training data and all test data sets are predicted, and the prediction results are saved. In this way, after five iterations, each primary learner is trained five times, and each data divided initially is predicted. All test data is predicted. A new training data matrix with the number of rows equal to the number of training set samples and the number of columns equal to the number of primary learners is obtained, which is the input to the secondary learner as training data. After the training of the secondary learner, the new test set is predicted and the final output is obtained.

### 3. Data Acquisition and Processing

The data studied in this paper can be divided into brain data and pulmonary image data. It includes data acquisition, annotation, and processing.

**3.1. Data Acquisition and Presentation.** The image sequence is from the International Early Lung Cancer Action Project database [32, 33], and the lung CT images were collected by the hospital. Image presentation is on a 21-inch LCD. The subjects sit in a comfortable posture, and the distance between their eyes and the center of the screen is about 80 cm. EEG signal acquisition uses a 64-channel brain product EEG. According to the international standard 10-20 system, electrode placement ensures effective coverage of all interested cortical areas. The sampling frequency of the EEG signal is 1000 Hz. Before the experiment, make sure that the impedance of each electrode is below 5 kΩ. In order to reduce the possible EMG signal crosstalk, the subjects are required to reduce head activity as much as possible after completing the EEG electrode arrangement.

**3.2. Data Annotation.** The image sequence of a solitary pulmonary nodule is labeled by two professional doctors using the independent blind marking method. The gold standard of pulmonary nodules is selected.

All the data are marked by two professional doctors in accordance with the blind mark method, and disputed annotations are arbitrated by a third expert. 10 groups of lung CT images are collected in the experiment, which include 3200 frames in total. EEG signals include 100 sets.

**3.3. Data Processing.** In order to increase the locality of the image sequence, maximize the awareness of spatial context, and minimize the vertigo of the interpreter, the hexagon

search path algorithm is adopted [34]. The hexagon search path algorithm decomposes a two-dimensional hexagon mesh into a nested set of approximate self-similar patterns with approximate hexagon symmetry. From the largest hexagon in the image to the smallest hexagon, it is similar to a Piano De Karl space to fill the whole image space. Through the hexagon search path algorithm.

The EEG signal is preprocessed as follows: (1) Resampling: reduce the EEG signal frequency to 256 Hz in order to improve processing speed and remove interference. (2) Select the common average reference surface: the average value of all electrodes is used as the common average reference in this experiment. (3) Filtering: this is done because the power frequency signal and high-frequency interference will be mixed in the EEG acquisition process, and the frequency of the EEG signal with research significance based on the oddball experimental paradigm is generally less than 50 Hz. Thus, the passband frequency of a 0.5~48 Hz filter is adopted in this experiment.

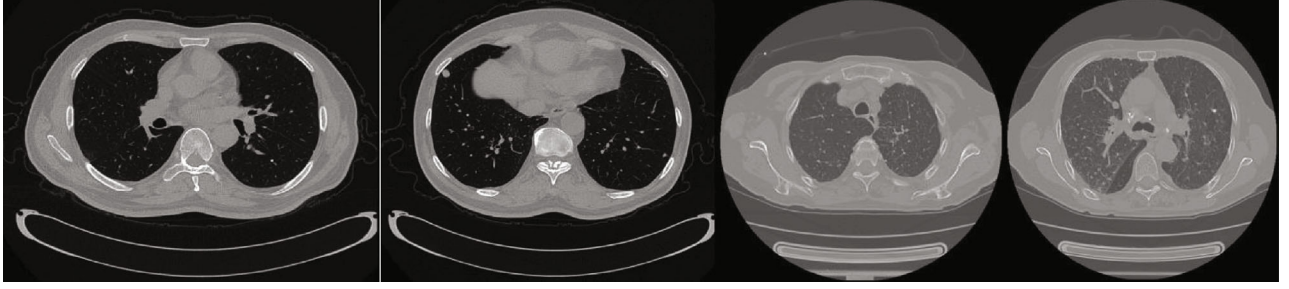
### 4. Experiment and Result Analysis

The subjects are divided into two groups, six in each group (three males and three females, average age 35 years). Group 1 consists of professional radiologists. Group 2 consists of people who had not participated in video training. There are 20 groups of image data, among which the coronal image sequence is generated by coronal sequence, as shown in Figure 3.

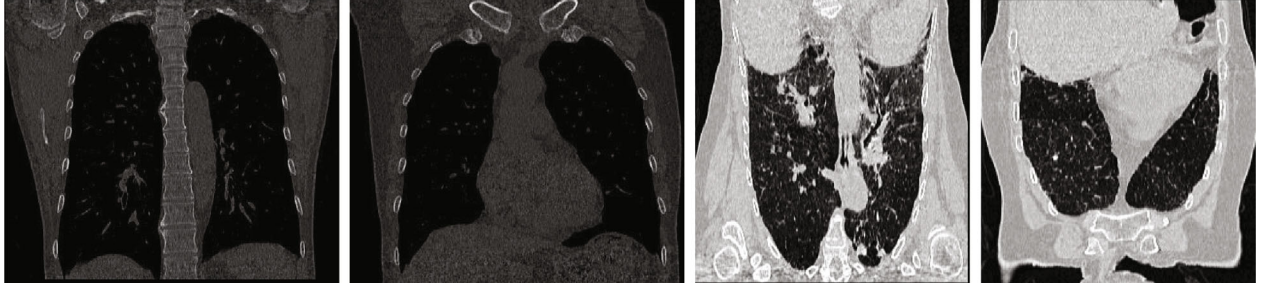
Each experimenter needs to complete two sessions, with a rest time of no more than five minutes between each session. Each session contains several blocks. Before the start of the block, the flashing red cross symbol is displayed in the center of the screen to remind the subjects to pay attention, and the two types of stimulus images appear in a proportional order.

**4.1. Saliency and Specificity Verification.** The significance of response and the specificity of brain response are the premise of brain-computer interface research. Because of the particularity of the medical image, the traditional brain-computer interface criterion cannot be directly applied to lung CT image detection. Therefore, we study the significance of test response and the specificity of brain response.

**4.1.1. Test Response Saliency.** According to the CT image data of pulmonary nodules, in order to verify whether there is significant difference in behavior of the subjects, we have



(a) Axial sequence image



(b) Coronal image sequence

FIGURE 3: Experiment data presentation.

TABLE 1: Recognition effect of each group.

Group	SEN	SPE	FPF	AUC
1	56%	92%	16%	0.89
2	17%	80%	32%	0.43

made statistics on the recognition effect of group 1 and group 2.

Sensitivity (SEN), specificity (SPE), and false positive fraction (FPF) are introduced to measure the performance of different algorithms [34]:

$$\begin{aligned}
 \text{SEN} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
 \text{SPE} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\
 \text{FPF} &= \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}},
 \end{aligned} \tag{16}$$

where TP is True Positive, FN is False Negative, FP is False Positive, and TN is True Negative.

In this paper, the parameter setting of the comparison algorithms is based on the parameter range mentioned in the corresponding article to find the optimal parameters.

From Table 1, the detection effect by specially trained doctors on pulmonary nodules is much higher than that by group 2. Furthermore, it is confirmed that there is significant difference in behavior among the pulmonary nodule test subjects.

To visually demonstrate the detection effects of different groups, an ROC curve is shown in Figure 4. We can see that the recognition effect of group 1 is much higher than that of group 2.

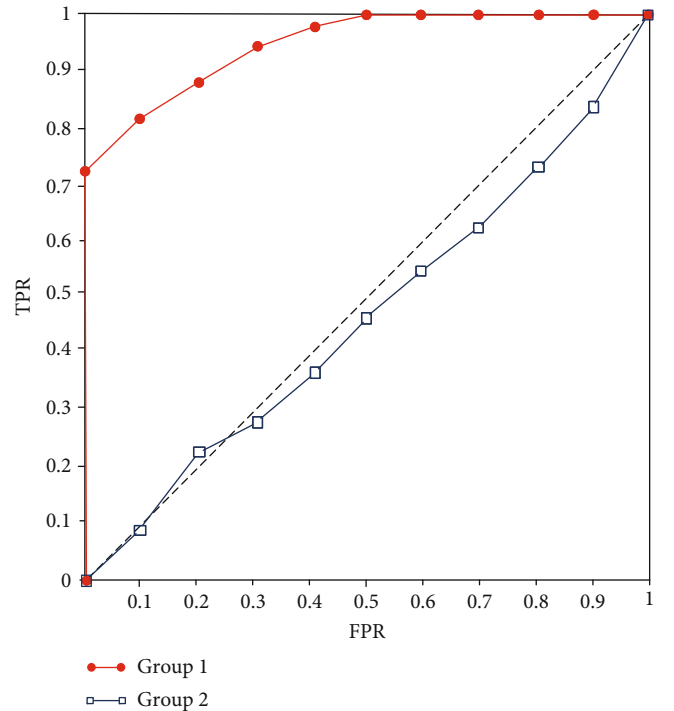


FIGURE 4: Recognition ROC figure.

**4.1.2. Brain Response Specificity.** The EEG data of subjects under relevant tasks are collected, and the specific brain response information of subjects to target and nontarget images is studied through the active participation of subjects in the target detection task. The EEG data of each subject are analyzed independently. Compared with the nontarget image, the brain activity of the prefrontal cortex area is more intense when the target image is observed,

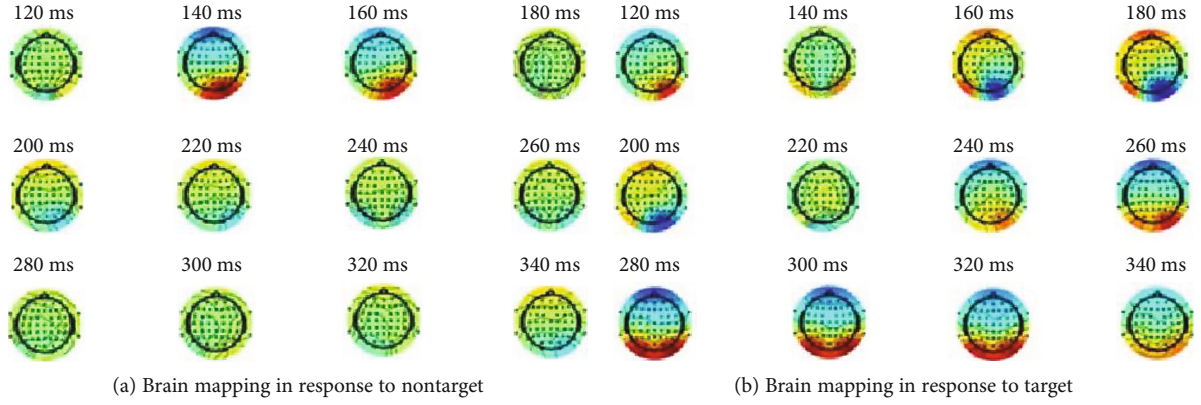


FIGURE 5: Brain response mapping.

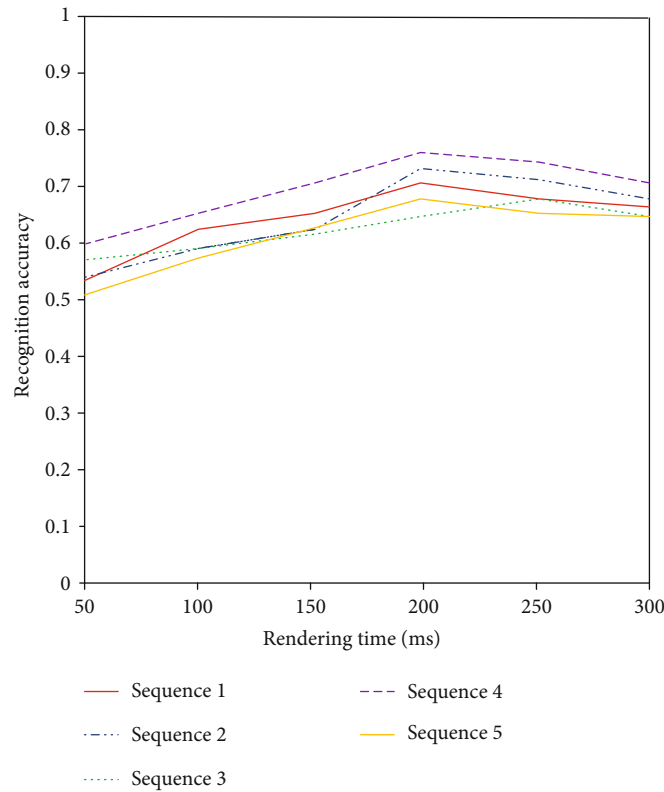


FIGURE 6: Time-recognition accuracy graph.

which is in line with P300 characteristics. That is to say, the EEG amplitude of the prefrontal cortex area would be larger than that of the normal stimulus sequence about 300 ms after the abnormal stimulus sequence is observed. The difference of the EEG amplitude is convenient to distinguish two kinds of data signals to judge the corresponding CT images of pulmonary nodules. In Figure 5, it is the brain topographic map of the target and nontarget images observed by one of the subjects in the experimental group at different time points. It can be concluded that the amplitude of the prefrontal cortex in the observation target image EEG data between 280 and 320 ms is larger than that in other time periods, while the EEG data changes little in the observation of nontarget images.

**4.2. Image Presentation.** The CT image of the lung has the characteristics of medicine and anatomy, which leads to a great difference between CT images and traditional images. In order to obtain EEG signals better, we start experiments from the time of image display, the proportion of target and background display, the way of image presentation, and the times of image repetition.

**4.2.1. Display Time.** The display time of an image will directly affect the corresponding effect on the brain. Fast sequence visual presentation is an experimental model for the detection of attention time characteristics [35]. In the fast sequence visual presentation paradigm, a series of target stimuli are placed in the background stimuli, which are fixed



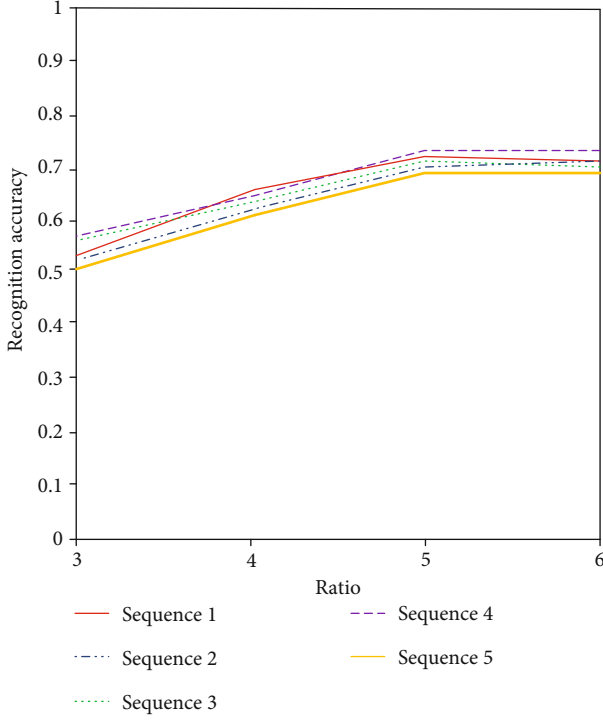


FIGURE 7: Ratio-recognition accuracy graph.

in the same specific position of the screen and presented in a specific period of time. Because of the particularity of the medical image, we need to further verify the effect of the display time. For this reason, we use five representative image sequences to select different display times to verify the effect.

The recognition accuracy of 50~250 ms is shown in Figure 6. With the increase of the presentation time, the recognition rate shows an upward trend within 50-200 ms. However, with the increase of presentation time, when it is more than 200 ms, the recognition rate will decrease due to the weakening of the brain signal. So far, in this paper, we select the optimal 200 ms as the presentation time.

**4.2.2. Target to Background Image Number Ratio.** According to the oddball paradigm [36], random presentation of two stimuli acting on the same sensory channel should ensure that there is a great difference in the probability of interclass stimulus presentation under the relevant potential. Therefore, we adjust the ratio of the number of pulmonary nodule images to the number of pulmonary nodule images under the condition of a 200 ms display. As shown in Figure 7, when the ratio of the target to the background increases, the signal stimulation produced by the brain tends to increase. When the ratio of the target to background reaches 1:5, it tends to be stable. So far, in this paper, we select the optimal ratio of 1:5 as the number of target and background images.

**4.2.3. Number of Repetitions.** Visual expert object recognition is a complex process of perception and cognition, including the dynamic interaction of low-level perception and high-level cognitive components, which involves the close coupling of perception, memory, attention, and semantics.

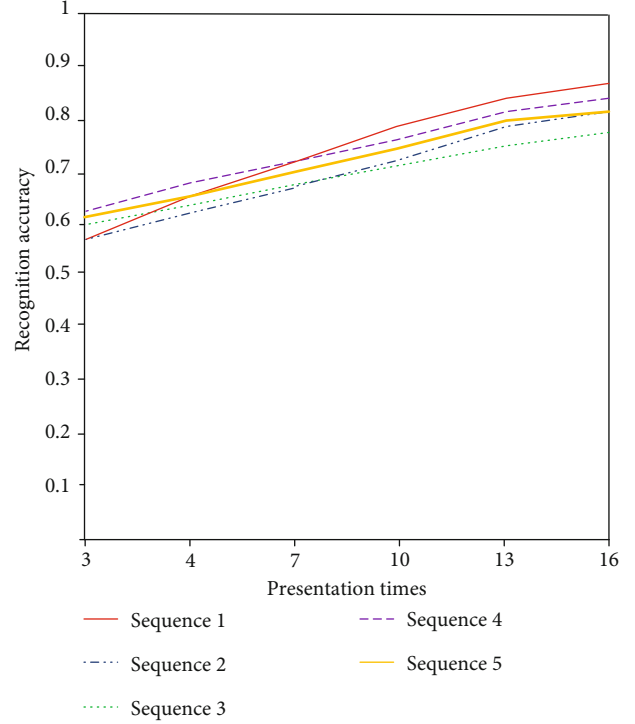


FIGURE 8: Presentation time-recognition accuracy relationship graph.

At the central level, multiple subsystems interact and work together to support visual experts to complete object recognition tasks. For medical image interpreters, each of them has accumulated rich experience through long-term and high-intensity purposeful training, and their behavior characteristics are significant. The accuracy of EEG classification can be improved by presenting the same image multiple times. For this reason, we compared the effect of different repetition times on the recognition accuracy.

As shown in Figure 8, with the increasing number of image repetitions, the accuracy of EEG interpretation increases correspondingly, and the curve rises and slows down when the image is repeated 13 times. It can be seen that for the same subject, repeated image presentation can improve the signal-to-noise ratio of EEG data and provide more effective brain response information.

**4.2.4. Display Mode.** Lung CT image data is 16 bits; an ordinary display cannot display all its information on a single frame image. To verify different image display modes, we define the following modes:

- (i) *Mode 1*: the display of window width and window position is set based on the average value of a single frame CT image.
- (ii) *Mode 2*: window width and window position display mapped by the lung window.
- (iii) *Mode 3*: based on the lung window mapping, two images are generated by floating up and down, that is, three images are displayed.



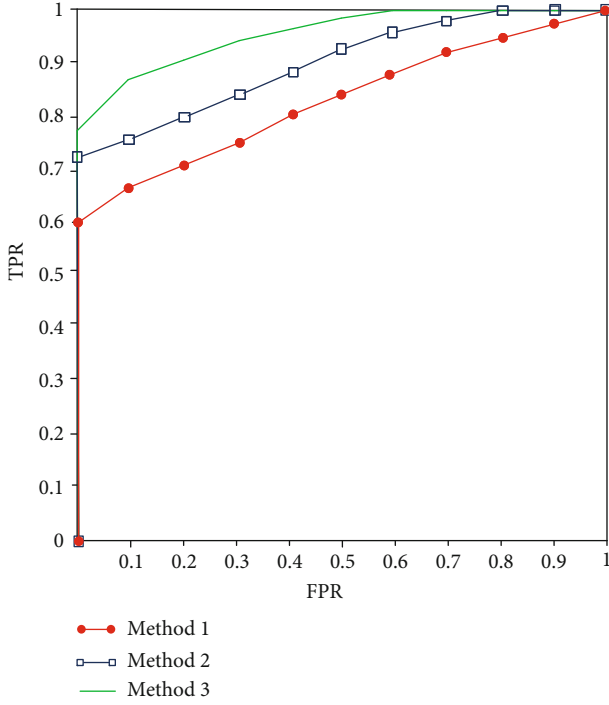


FIGURE 9: ROC curves of different methods.

The bone window, lung window, and vertical and horizontal window have a display mode with a fixed mapping range. No human-computer interaction is required. When the image is inputted, it will be automatically displayed according to the mapping rules.

As shown in Figure 9, due to the limited information obtained in Mode 1, the features of pulmonary nodules are not obvious, which lead to recognition failure. For Method 2, under the premise that the best window of the lung can be observed, the shape of the lung can be displayed better, but the recognition of pulmonary nodules with an unclear boundary still fails. Inspired by the above two methods, Mode 3 displays the information image of multiple windows and wide windows, and the tested person can make a comprehensive judgment, which has a good effect.

**4.3. Classification Effect Comparison.** To verify the proposed algorithm classification effect, we compare different algorithms, as shown in Figure 10. SVM [18] establishes a multilevel structure to decompose the input signal by wavelet, which has a certain effect on EEG signal classification. Random forest [28] integrates spatial information and polarizes information to realize signal classification. Logistic regression [29] extends the algorithm of polynomial logistic regression to the semisupervised learning of posterior class distribution to improve the classification performance. KNN [30] introduces convolution to the EEG image for classification. AdaBoost [31] combines fuzzy entropy, sample entropy, approximate entropy, and spectral entropy to realize EEG signal classification.

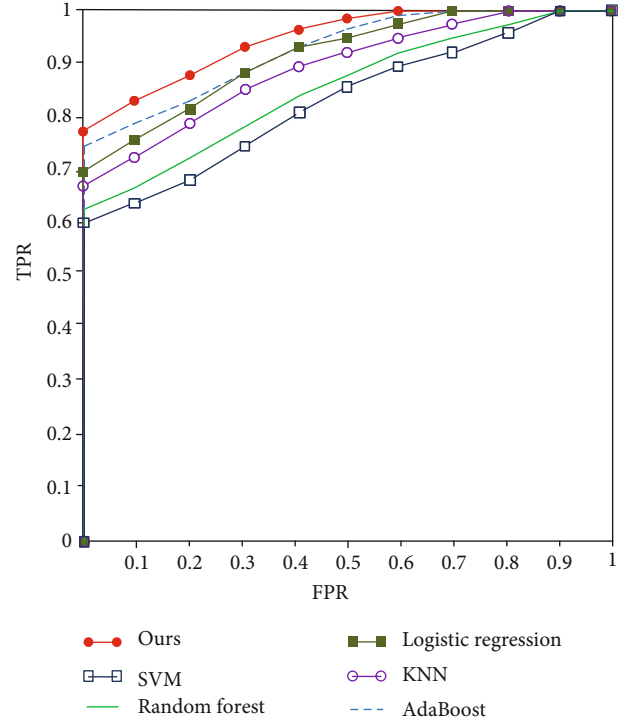


FIGURE 10: ROC curves of different algorithms.

## 5. Conclusion

We studied the process of detecting solitary pulmonary nodules by physicians from the aspects of pulmonary nodule imaging and physicians' EEG response. A time-domain and frequency-domain model is proposed to extract EEG features. A multilayer feature fusion model is proposed to verify the feasibility of the brain-computer interface in detecting pulmonary nodules.

## Data Availability

All used data is within the paper.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work is supported by the Light of West China (Grant No. XAB2016B23), Chinese Academy of Sciences; the Open Project Program of the State Key Lab of CAD&CG (Grant No. A2026), Zhejiang University; the National Natural Science Foundation of China (Grant No. 61561040); and the Shanxi National Science Foundation (2020JQ-518).

## References

- [1] V. K. Venugopal, K. Vaidhya, M. Murugavel et al., "Unboxing AI - radiological insights into a deep neural network for lung nodule characterization," *Academic Radiology*, vol. 27, no. 1, pp. 88–95, 2020.

- [2] S. Qiu, Q. Guo, D. Zhou, Y. Jin, T. Zhou, and Z. He, "Isolated pulmonary nodules characteristics detection based on CT images," *IEEE Access*, vol. 7, pp. 165597–165606, 2019.
- [3] M. J. Vansteensel, D. Hermes, E. J. Aarnoutse et al., "Brain-computer interfacing based on cognitive control," *Annals of Neurology*, vol. 67, no. 6, pp. 809–816, 2010.
- [4] D. Tan and A. Nijholt, "Brain-computer interfaces and human-computer interaction," in *Brain-Computer Interfaces, Human-Computer Interaction Series*, D. Tan and A. Nijholt, Eds., pp. 3–19, Springer, London, 2010.
- [5] J. Jin, B. Z. Allison, E. W. Sellers et al., "Optimized stimulus presentation patterns for an event-related potential EEG-based brain-computer interface," *Medical & Biological Engineering & Computing*, vol. 49, no. 2, pp. 181–191, 2011.
- [6] A. J. B. Myrden, A. Kushki, E. Sejdić, A. M. Guerguerian, and T. Chau, "A brain-computer interface based on bilateral transcranial Doppler ultrasound," *PLoS One*, vol. 6, no. 9, p. e24170, 2011.
- [7] M. Duvinage, T. Castermans, M. Petieau et al., "A P300-based quantitative comparison between the Emotiv Epoc headset and a medical EEG device," *Biomedical Engineering*, vol. 765, no. 1, pp. 2012–2764, 2012.
- [8] J. S. Lin and W. C. Yang, "Wireless brain-computer interface for electric wheelchairs with EEG and eye-blinking signals," *International Journal of Innovative Computing, Information and Control*, vol. 8, no. 9, pp. 6011–6024, 2012.
- [9] A. H. Do, P. T. Wang, C. E. King, S. N. Chun, and Z. Nenadic, "Brain-computer interface controlled robotic gait orthosis," *Journal of Neuroengineering and Rehabilitation*, vol. 10, no. 1, p. 111, 2013.
- [10] Z. Gu, Z. Yu, Z. Shen, and Y. Li, "An online semi-supervised brain-computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 9, pp. 2614–2623, 2013.
- [11] S. Sun and J. Zhou, "A review of adaptive feature extraction and classification methods for EEG-based brain-computer interfaces," in *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 1746–1753, Beijing, China, 2014.
- [12] E. Mikołajewska and D. Mikołajewski, "The prospects of brain — computer interface applications in children," *Open Medicine*, vol. 9, no. 1, pp. 74–79, 2014.
- [13] A. Nijholt, "Competing and collaborating brains: multi-brain computer interfacing," in *Brain-Computer Interfaces, Intelligent Systems Reference Library*, A. Hassanien and A. Azar, Eds., pp. 313–335, Springer, Cham, 2015.
- [14] S. Kumar, V. Kumar, and B. Gupta, "Feature extraction from EEG signal through one electrode device for medical application," *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, 2015, pp. 555–559, Dehradun, India, 2015.
- [15] C. Jeunet, E. Jahanpour, and F. Lotte, "Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study," *Journal of Neural Engineering*, vol. 13, no. 3, article 036024, 2016.
- [16] J. C. Liu, H. C. Chou, C. H. Chen, Y. T. Lin, and C. H. Kuo, "Time-shift correlation algorithm for P300 event related potential brain-computer interface implementation," *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3039454, 22 pages, 2016.
- [17] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 234–239, Banff, AB, Canada, 2017.
- [18] E. Dong, C. Li, L. Li, S. du, A. N. Belkacem, and C. Chen, "Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain-computer interfaces," *Medical & Biological Engineering & Computing*, vol. 55, no. 10, pp. 1809–1818, 2017.
- [19] V. Guy, M. H. Soriani, M. Bruno, T. Papadopoulos, C. Desnuelle, and M. Clerc, "Brain computer interface with the P300 speller: usability for disabled people with amyotrophic lateral sclerosis," *Annals of Physical and Rehabilitation Medicine*, vol. 61, no. 1, pp. 5–11, 2018.
- [20] M. A. Schwemmer, N. D. Skomrock, P. B. Sederberg et al., "Meeting brain-computer interface user performance expectations using a deep neural network decoding framework," *Nature Medicine*, vol. 24, no. 11, pp. 1669–1676, 2018.
- [21] J. Abbasi, "New deep brain stimulation target for mood disorders," *JAMA*, vol. 321, no. 6, pp. 537–537, 2019.
- [22] X. Wan, K. Zhang, S. Ramkumar et al., "A review on electroencephalogram based brain computer interface for elderly disabled," *IEEE Access*, vol. 7, pp. 36380–36387, 2019.
- [23] P. Sengupta, M. R. S. John, and S. S. Sridhar, "Classification of conscious, semi-conscious and minimally conscious state for medical assisting system using brain computer interface and deep neural network," *Journal of Medical Robotics Research*, vol. 4, no. 2 article 1942004, p. 10, 2020.
- [24] K. Wang, M. Xu, Y. Wang, S. Zhang, L. Chen, and D. Ming, "Enhance decoding of pre-movement EEG patterns for brain-computer interfaces," *Journal of Neural Engineering*, vol. 17, no. 1, article 016033, 2020.
- [25] N. Rehman and D. P. Mandic, "Multivariate empirical mode decomposition," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 466, no. 2117, pp. 1291–1302, 2010.
- [26] D. W. Vogt, M. Erkintalo, and R. Leonhardt, "Terahertz frequency domain spectroscopy using Hilbert transformation," in *2019 Conference on Lasers and Electro-Optics Europe & European Quantum Electronics Conference (CLEO/Europe-EQEC)*, p. 1, Munich, Germany, 2019.
- [27] T. Song, Z. Fei, M. Yankowitz et al., "Switching 2D magnetic states via pressure tuning of layer stacking," *Nature Materials*, vol. 18, no. 12, pp. 1298–1302, 2019.
- [28] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 38–53, 2015.
- [29] T. V. N. Prabhakar, G. Xavier, P. Geetha, and K. P. Soman, "Spatial preprocessing based multinomial logistic regression for hyperspectral image classification," *Procedia Computer Science*, vol. 46, pp. 1817–1826, 2015.
- [30] X. Tang, T. Wang, Y. Du, and Y. Dai, "Motor imagery EEG recognition with KNN-based smooth auto-encoder," *Artificial Intelligence in Medicine*, vol. 101, p. 101747, 2019.
- [31] J. Hu, "Automated detection of driver fatigue based on AdaBoost classifier with EEG signals," *Frontiers in Computational Neuroscience*, vol. 11, p. 72, 2017.
- [32] S. Qiu, D. Wen, J. Feng, and Y. Cui, "Lung nodules detection in CT images using Gestalt-based algorithm," *Chinese Journal of Electronics*, vol. 25, no. 4, pp. 711–718, 2016.

- [33] SIMBA Public Database, “International Early lung Cancer Action Project [DB/OL],” 2015, <http://www.via.cornell.edu/lungdb.html>.
- [34] S. Qiu, J. Luo, S. Yang, M. Zhang, and W. Zhang, “A moving target extraction algorithm based on the fusion of infrared and visible images,” *Infrared Physics & Technology*, vol. 98, pp. 285–291, 2019.
- [35] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, “Brain activity-based image classification from rapid serial visual presentation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 5, pp. 432–441, 2008.
- [36] R. Verleger and K. Śmigajewicz, “Do rare stimuli evoke large P3s by being unexpected? A comparison of oddball effects between standard-oddball and prediction-oddball tasks,” *Advances in Cognitive Psychology*, vol. 12, no. 2, pp. 88–104, 2016.

## Research Article

# The EEG-Based Attention Analysis in Multimedia m-Learning

Dan Ni , Shuo Wang, and Guocheng Liu

*The School of Information Engineering, Jilin Engineering Normal University, Changchun, 130000 Jilin, China*

Correspondence should be addressed to Dan Ni; 249962743@qq.com

Received 27 March 2020; Accepted 9 May 2020; Published 10 June 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Dan Ni 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, research on brain-computer interfaces has been increasing in the field of education, and mobile learning has become a very important way of learning. In this study, EEG experiment of a group of iPad-based mobile learners was conducted through algorithm optimization on the TGAM chip. Under the three learning media (text, text + graphic, and video), the researchers analyzed the difference in learners' attention. The study found no significant difference in attention in different media, but learners using text media had the highest attention value. Later, the researchers studied the attention of learners with different learning styles and found that active and reflective learners' attention exhibited significant differences when using video media to learn.

## 1. Introduction

The industrial revolution 4.0, also known as the digital revolution, has driven a massive transformation in the world [1]. Currently, the use of digital technology has occurred in all aspects of life and various age groups. People have been accustomed to accessing various knowledge from the Internet.

The rapid development in digital technology has also created new challenges to conventional education systems at various levels of education, from basic education to higher education [2]. The popularity of mobile terminals and the rapid development of the mobile Internet have provided a carrier for the development of mobile learning (m-learning) and have made mobile learning a hot research topic in education informatization [3]. The size and weight of tablets, as well as improved screen quality, have made reading from mobile devices more acceptable [4]. Mobile learning has attracted much attention as a new teaching and learning model. Mobile learning (m-learning) has become increasingly popular in universities, and more college students have access to smartphones, tablets, and other mobile devices [5]. Digital technology has also changed students' tendency to learn. Today, students tended to demand the freedom to decide what they want to learn and when and how they want to learn [1]. A common factor in mobile learning was the

learning media, and people had an interest in further analysis of the selection of appropriate media for educational content to achieve higher learning results [6]. There was a wide range of perspectives on the use of information technology for achieving learning effectiveness, ranging from those who asserted that media did not influence learning effectiveness to those who believed that the decisions made regarding media would influence learning effectiveness [7].

Today, common types of contents were accessed via mobile devices, including videos and texts such as pdf, audio, or video files or a combination of these file types evidenced in e-books and online articles. Different kinds of content had the potential to support learning through both verbal and visual demonstrations that might motivate students to learn [7]. A study showed that there might be an interaction between media choice and other variables, which not only affected perceived usefulness but also affected the learning effect of educational programs [8].

Most research on mobile learning focused on learner motivation and attitudes [9]. Hwang and Wu [10] pointed out that the limitations of mobile devices also affected students' learning attention and cognitive load. Solving the problem of attention was a topic worth exploring in mobile learning, but the research on this problem was insufficient [11].

After entering the 21st century, the increasingly mature brain imaging technology has once again promoted cognitive science to move forward. Using neuroscience research methods and techniques to study classical cognitive problems has become a new trend in academia. Therefore, people's research on cognition has also achieved a breakthrough from macrobehavior to the microneural connection [12]; EEG was one of the research tools used to measure human brain activity. Students' learning involves brain activities of information input and processing because the use of EEG to measure students' learning status is a good choice.

Therefore, this study focused on the differences in learners' attention when using three different learning media for mobile learning based on EEG.

The purpose of the research is to use EEG technology to measure mobile learners' attention differences in different learning media environments and then to clarify the impact of different media on mobile learners' attention, to provide corresponding suggestions for learners, teachers, and e-learning resource developers, which were related to the trend of improving students' mobile learning, to support individuals to better improve mobile learning performance. Researchers were motivated to design and research a mobile learning experience with different media learning materials, which would be supported by tablet devices. In this process, we took learning style theory as a theoretical perspective to compare the differences of the learners' attention.

The research questions are as follows:

- (1) When learners used three media for mobile learning, which case has the highest learner attention value?
- (2) When learners used three media for mobile learning, would gender, education affect learners' attention?
- (3) When learners use three media for mobile learning, was there a difference in attention among students with different learning styles?

## 2. Literature Review

**2.1. Mobile Learning.** Mobile learning has currently no more unified concept. Researchers had different definitions of mobile learning from different perspectives. From a technical point of view, mobile learning was a kind of learning that could be done by learners at any time and any place with the help of mobile computing devices that could effectively present learning content and provide two-way communication between teachers and students [13]. From the learner's point of view, mobile learning was that learners used mobile devices to learn anywhere and anytime [14]. Researchers of China believed that mobile learning referred to a new form of learning that used wireless mobile communication network technology and wireless mobile communication devices to obtain educational information and resources and educational services [15]. Definitions of mobile learning emphasised mobility, access [16], immediacy [17], ubiquity [18], and convenience [17].

In summary, mobile learning covered the following points: any place, any time, any form, mobile network, and

equipment support. In short, mobile learning referred to a kind of learning model that could achieve access to digital learning resources and educational information anytime and anywhere, with the help of the seamless wireless network, and portable mobile communication equipment, and facilitate communication and interaction.

**2.1.1. Attention in Mobile Learning.** In e-learning and mobile learning, lack of student attention has become an important problem [11], and some scholars have confirmed that attention was closely related to academic performance [19]. Although e-learning or mobile learning had the advantage of not being limited by time and place, there were also related issues caused by the non-face-to-face learning environment of teachers and students, especially the screen size of mobile learning devices and the characteristics of their representation methods [20]. Attention was still an urgent problem. Assessing students' attention status in an e-learning environment was more difficult than assessing them in face-to-face instruction [21]. With the advancement in the evaluation of human physiological signals, e-learning research increasingly used physiological signals to determine students' level of attention [19]. In recent years, some researchers have started using EEG detection tools [22, 23] for empirical research on student attention and attention during multimedia learning [24].

**2.1.2. The Learning Style in Mobile Learning.** There were many different definitions of learning styles, among which were more influential: Kinsella [25] proposed that learning style referred to the partial way in which learners were naturally employed in the process of information and information processing.

Studies have shown that learning styles and learning environments had a positive effect on learners' academic performance [26]. Learning style was also a factor that affects the quality of e-learning [27] and greatly affected learners' academic performance [28]. Learning styles also affected mobile learning [11].

The most widely used in mobile learning research was the learning style scale developed by Felder and Solomon (1991), which was divided into four dimensions (information input, information processing, content understanding, and perception) (see Table 1). Learners had different characteristics in each dimension, and each dimension had two attributes (visual/verbal, sensing/intuitive, active/reflective, and sequential/global). ILS was compared with other learning style questionnaires; the results showed that this questionnaire was more comprehensive, and the results are better [29]. This scale was widely used in e-learning research.

The learning style was added in this study to confirm whether students with different learning styles had different attention when using different media for mobile learning.

**2.2. Learning Media.** The influence of the media on learning has been a research topic that researchers pay attention to. Liu et al. [31] studied the impact of different media on e-learning content, including the impact of text and video on users' acceptance of e-learning and user attention.



TABLE 1: Brief description of the types of ILS.

One polarity	Opposite polarity
(i) Active learners prefer rushing in and doing	(i) Reflective learners prefer to reflect before starting
(ii) Sensing learners prefer facts and prefer using well-known relationships	(ii) Intuitive learners prefer to discover possibilities and relationships
(iii) Visual learners prefer pictures and visual material	(iii) Verbal learners prefer written and spoken text
(iv) Sequential learners tend to learn material in steps	(iv) Global learners absorb material often randomly without necessarily seeing the connections

This table was drawn from [30].

They found that the richness of content was positively correlated with user focus.

According to cognitive theory, learning was the acquisition or reorganization of the cognitive structures through which humans processed and stored information [32]. It focused on information and content delivery in mobile learning using multimedia learning (dual code, cognitive load theory): images, audio, video, text, and animations [33].

The main focus of media debate was whether one media naturally promotes learning more effectively than the other [34–36]. The answer to this question was mainly divided into two opposing views represented by Clark and Kozma. Clark [37] believed that the type of media did not affect learning; learning was only affected by the way the media was used. This view suggested that video-based learning materials were not necessarily more effective than text, because how text-based or audio-based applications were designed should promote equal levels of learning [38]. In contrast, Kozma [39] believed that different types of media “had special characteristics, which made them more or less suitable for some types of learning tasks.” A set of computer-based experiments were carried out, which used three different media combinations (text only, text and diagrams, spoken text, and diagrams) to study learners’ understanding achieved in complex areas (statistics) in a real e-learning environment [30].

This research was to provide the learning content of three media in a mobile learning environment and analyze the differences in learners’ attention.

**2.3. Attention and EEG.** EEG was a process used to record brain wave activity and was called the “mind window” [40]. In human-computer interface research, EEG was a method for making psychophysiological measurements to check the relationship between psychological processes and physical processes. In general, EEG was measured by recording the voltage of the electrodes on the scalp. The electrodes were placed at designated positions allocated on the head [41]. In recent years, EEG technology has also been widely used in other fields, especially in computer interface design and computer game development. Brain-computer interface research focused on using EEG activities to control external devices, such as robots and virtual environments [42].

Currently, as a research tool, more and more researchers used portable EEG for educational research and have achieved many research results [43–45]. To a certain extent, investment in attention could also promote brain information processing and coding so that learners could obtain bet-

ter academic performance [46]. Since it was difficult to measure attention using self-reporting tools, many studies have used electroencephalography (EEG) as a tool to measure changes in attention status [47]. Attention was the most commonly used evaluation index of portable EEG technology in education research [48]. If we could use EEG technology scientifically, it could be an effective tool for detecting and processing brain signals for educational purposes. More precisely, when using EEG technology in some innovative ways, it could capture brain signals and process them to determine learners’ learning and memory during learning [49]. Studies identified that the EEG data was used successfully in detecting the learners learning style and learning preferences and the correlation between them [50].

### 3. Research Methodology

**3.1. Participants.** In this study, 30 college students from a university of science and technology in China were randomly recruited by snowball sampling, all of whom were right-handed. Because the test data of two participants were unreliable, they were ignored. Among the 28 participants, 19 were male, 9 were female, and the age was between 19 and 31 years old ( $M_{age} = 24$ ,  $SD = 2.78$ ). Among them, 13 were undergraduates and 15 were postgraduates. According to the investigation of the participants, they had no mental diseases such as epilepsy, depression, and hyperactivity disorder or did take psychoactive drugs for a long time. At present, they neither had used any drugs to change their thinking nor had any history of head injury or brain injury. The experimenter introduced the scope and procedure of the experiment to the participants and informed them that the experiment would not cause any risk to their health, to ensure that the participants could participate in the experiment voluntarily and sign the informed consent before the experiment. All participants had more than 5 years of mobile device experience, and 27 people had more than 2 years of experience in using tablet computers, which showed that they are familiar with the operation of mobile devices.

Because this experiment was based on the learning of mobile devices, the participants needed to ensure that they can see the learning content, so they were required to wear appropriate glasses when necessary to ensure that the learning content is clear and visible, and the test materials were in Chinese, that is, the native language of the participants. The background characteristics of the participants are shown in Table 2.

TABLE 2: Background characteristics of participants.

	N	Percentage (%)
Gender		
Male	19	67.9
Female	9	32.1
Educational background		
Undergraduate	13	46.4
Graduate	15	53.6
Age		
Under 20	2	7.1
21-25	20	71.4
Above 26	6	21.4
Tablet experience		
Less than 1 year	1	3.6
2-4 years	15	53.6
Above 5 years	12	42.9

TABLE 3: Experimental material design.

Media forms	Contents	Estimated learning time
Text	Artificial intelligence	5 min
Text + graph	Virtual reality	5 min
Video	5G technology	5 min

**3.2. Stimulus Materials.** To ensure that the experimental content can be scientific and reasonable, the researcher invited two experts with rich experience in mobile learning and EEG. After repeated discussion, research, and trial, the final experimental materials were determined to choose the Xinhua News Agency website, the most influential online media in China and the Chinese website with global influence (<http://www.xinhuanet.com>) and determined three declarative knowledge with similar themes, and the same difficulty was presented in three different media forms: text, text + graph, and video. This was a combination of multimedia demonstrations used in many e-learning environments. Based on determining the content, the researcher arranged five college students to take a test. After feedback, the content was similar and the difficulty was the same, and the required learning time was close. The media types and contents selected in this experiment are shown in Table 3.

### 3.3. Research Instruments

**3.3.1. EEG Device and Algorithm Optimization.** In this study, the MindWave mobile headset produced by the NeuroSky company was selected as a tool to record EEG data signals. This equipment had a great influence on students' attention training. Previous studies [19] have confirmed that NeuroSky's mindset headset was effective and reliable enough based on the correlation between bird watching scores. Also, Rebolledo-Mendez et al. [51] also found that NeuroSky's mindset headphones measured a positive correla-



FIGURE 1: MindWave mobile headset.

tion between attention measures and self-reported attention levels through the second life assessment exercise. The results showed that the attention value measured by NeuroSky's mind headset had satisfactory validity and reliability for recognizing the attention of learners in learning activities. The new generation of the MindWave mobile headset was more perfect based on a mindset headset. The device ran well, had good measurement accuracy, and had its software development kit (SDK), which was convenient for software developers to design applications [52].

The MindWave mobile headset was composed of an ear clip and a sensor arm (see Figure 1). The reference electrode and the ground electrode of the earphone were located on the ear clamp, and the EEG electrode was located on the sensor arm and the left forehead above the eyes. The EEG signal of the forehead could be detected in real time. These EEG signals could show the changes in people's consciousness after the complex e-sense TM calculation. By quantifying the psychological state of the subjects as the value of attention, which could be used to analyze the degree of attention of learners, it could be divided into attention and relaxation, and the parameters were between 0 and 100. The device also had noise-filtering technology, through a complex algorithm to eliminate interference from other electronic equipment or daily living environment.

The MindWave mobile headset used the TGAM module to process and output brain wave spectrum, EEG signal quality, original brain wave, and two NeuroSky e-sense parameters: attention and relaxation detection. The NeuroSky system was composed of dry electrodes. The interface with the human body only needed a simple dry contact. A single EEG channel had three contact points: EEG (EEG acquisition point), ref (reference point), and GND (ground point). The original EEG data was output at 512 Hz.

As the attention data calculated by the chip's algorithm still needed to be improved in the application, such as large fluctuations and delay (see Figure 2), at the same time, the data was divided into many meaningless null regions. This

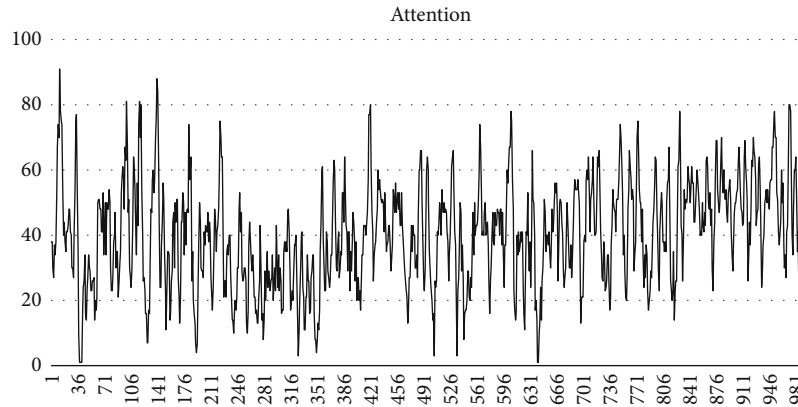


FIGURE 2: Attention data calculated by the chip's algorithm.

increases analysis cost. To ensure that scientific and accurate data was collected in this experiment, the researchers optimized the algorithm. The main methods included the following: (1) a checksum check was added, that is, checking whether the data value of the data packet and the checksum data used for detection are consistent, (2) cross-checking with the previous data packet was performed, and (3) the attention was stepped and the average value of the two continuously updated attention values was used to make the jittered data smoother.

**3.3.2. Index of the Felder-Solomon Learning Style (ILS).** The ILS questionnaire was composed of forty-four items requiring students to choose from two options, a or b, compulsorily. Each number was referred to any of the four scopes or dimensions which included the following: active and reflective, sensing and intuitive, visual and verbal, and sequential and global. For the scoring, summing up the number of *a* and *b* responses for each dimension formed scores which range from 1 to 11. Lower scores were subtracted from the higher score of either a or b.

**3.3.3. The Participant Questionnaire.** Participants' questionnaire collected demographic information about participants' gender, education background, age, participants' experience of using tablet computers and mobile devices, and learning time spent on mobile devices every day. At the same time, participants also had to answer the content test.

### 3.4. Experiment Design and Procedure

**3.4.1. Experiment Design.** Participants in the experiment read three different media types of learning content on the tablet computer, and each of them freely chose the reading order according to their wishes. Each reading content was about 5 minutes, and after the experiment, they filled in the test questions.

**3.4.2. Experiment Procedure.** The experiment was carried out in a very quiet research room. The curtain was put down to avoid direct sunlight, and the indoor light source was turned on. After confirming that there was no problem with the

tablet computer, MindWave mobile headset, and data storage terminal (computer), two research assistants explained and demonstrated the experiment. The participants completed the following tasks.

Each participant signed an informed consent form and found a suitable position to sit down, facing the tablet, correctly wore the EEG headset with the help of the researchers, and confirmed that the device was correctly placed, and the participants were told not to move as much as possible. The participants closed their eyes in a relaxed way (starting from the stable EEG data), opened their eyes, and began to read at their own pace after a few seconds. After each reading, they should answer the corresponding test questions. During this period, they were not allowed to return to the article to find the answers. When one content test was over, the participants closed their eyes and had a rest for about 15 seconds, then moved to the next one, and so on, until the three content tests were over (see Figure 3).

After the test, each participant would receive a survey about learning style and demographic information, followed by a reading behavior survey, and then, the researchers reported the results to each participant and thanked them for their participation. The whole process lasted 30-40 minutes, depending on how fast participants read each paragraph.

**3.5. Data Collection and Analysis.** According to the calculation method of the learning style scale, the researcher made statistics on the collected learning style questionnaire. After putting each answer in the corresponding question, the researcher can finally sum up the category of student learning style, which is represented by number + a/b. Each calculation result was finally recorded in an Excel spreadsheet.

The data collected by EEG is directly transmitted to the computer through Bluetooth and saved in Excel. Based on the relevant previous research, the average value of the participants' attention in the data information was selected for analysis and research.

The independent sample *t*-test in SPSS 25.0 statistical software was mainly used for data analysis.



FIGURE 3: EEG measurement process. The participant was being tested while using an iPad.

#### 4. Experimental Results

In this section, we analyzed the attention from the EEG experiment and answered the following questions:

- (1) Were there differences in the learner's attention in the context of three learning media (text, text + graphic, and video)?
- (2) Did the gender and education affect learner attention on learning media?
- (3) Did learning styles affect learners' attention differences in different media?

**4.1. Difference Comparison of Attention Values of Learners in the Three Media Contexts.** From the results, we could get that the average attention value of the text media ( $M = 55.43$ ,  $SD = 12.82$ ) was higher than others, including text + graph ( $M = 51.32$ ,  $SD = 10.83$ ) and video ( $M = 51.18$ ,  $SD = 11.55$ ) (see Table 4). It could be seen that there was no significant difference in the attention of media form when using text + graph or video, and the attention of text media showed a higher mean, which could prove that text media had a certain promotion effect on the attention of the learners.

**4.2. Difference Comparison of Attention Values of Learners on Gender.** Through the independent sample  $t$ -test, there was no significant difference of attention on gender when using text ( $t = -0.941$ ,  $p = 0.356$ ), text + graph ( $t = 0.475$ ,  $p = 0.639$ ), and video ( $t = 1.261$ ,  $p = 0.219$ ), but we found that both male and female learners have a higher attention value on text media than the other two media. Meanwhile, the average male attention value was significantly higher than the female's (see Table 5).

TABLE 4: Learners' attention value in the three media contexts.

	$N$	$M$	$SD$
Attention (text)	28	55.43	12.816
Attention (text + graph)	28	51.32	10.829
Attention (video)	28	51.18	11.554

**4.3. Difference Analysis of Attention Values of Learners with Different Education Experience.** As shown in Table 6, there was no significant difference of learners' attention with different education experience when using text ( $t = -1.532$ ,  $p = 0.138$ ), text + graph ( $t = -1.472$ ,  $p = 0.153$ ), and video ( $t = -1.746$ ,  $p = 0.093$ ). Graduates and undergraduates had a higher attention value ( $M = 58.80$ ,  $51.54$ ) when using text media to learn than text + graph ( $M = 54.07$ ,  $48.15$ ) or video ( $M = 54.60$ ,  $47.23$ ). Moreover, the average attention value of graduate students was higher than undergraduates'.

**4.4. Difference Comparison of Attention Values of Learners with Different Learning Styles.** The first axis of Felder's learning style is active and reflective, so students are divided into two categories based on the survey results. As shown in Table 7, through the independent sample  $t$ -test, the results showed that the attention difference was not significant between the active learners and reflective learners when using text or text + graph. However, there was a significant difference in attention between active and reflective learners when using video media ( $t = -2.308$ ,  $p < 0.05$ ); it indicated that when using video media to learn in a mobile learning context, the attention of active learners ( $M = 46.15$ ,  $SD = 9.23$ ) significantly was lower than that of reflective learners ( $M = 55.53$ ,  $SD = 11.86$ ).



TABLE 5: Difference comparison of attention values of learners on gender.

	Gender	<i>N</i>	Mean	SD	<i>t</i>	<i>p</i>
Attention (text)	Male	19	57.00	14.22	0.941	0.356
	Female	9	52.11	9.02		
Attention (text + graph)	Male	19	52.00	11.12	0.475	0.639
	Female	9	49.89	10.68		
Attention (video)	Male	19	53.05	11.21	1.261	0.219
	Female	9	47.22	11.91		

TABLE 6: Difference comparison of attention values of learners with different education experience.

	Education	<i>N</i>	Mean	SD	<i>t</i>	<i>p</i>
Attention (text)	Undergraduate	13	51.54	13.78	-1.532	0.138
	Graduate	15	58.80	11.31		
Attention (text + graph)	Undergraduate	13	48.15	10.65	-1.472	0.153
	Graduate	15	54.07	10.57		
Attention (video)	Undergraduate	13	47.23	9.54	-1.746	0.093
	Graduate	15	54.60	12.35		

TABLE 7: Difference comparison of attention values of learners with active and reflective learning styles.

	Types	<i>N</i>	<i>M</i>	SD	<i>t</i>	<i>p</i>
Concentration (text)	Active	13	53.54	13.14	-0.720	0.478
	Reflective	15	57.07	12.75		
Concentration (text + graph)	Active	13	50.54	8.53	-0.360	0.722
	Reflective	15	52.00	12.76		
Attention (video)	Active	13	46.15	9.23	-2.308	0.029*
	Reflective	15	55.53	11.86		

\* $p < 0.05$ .

TABLE 8: Difference comparison of attention values of learners with sequential and global learning styles.

	Types	<i>N</i>	<i>M</i>	SD	<i>t</i>	<i>p</i>
Attention (text)	Sequential	14	55.93	11.82	0.203	0.841
	Global	14	54.93	14.17		
Attention (text + graph)	Sequential	14	53.07	11.26	0.851	0.403
	Global	14	49.57	10.50		
Attention (video)	Sequential	14	52.50	13.64	0.598	0.555
	Global	14	49.86	9.36		

Since almost all the research subjects belong to the perception type and the visual type, no attention difference analysis was conducted.

In the group of sequential and global learning styles, the data (see Table 8) showed that whether in text ( $t = 0.203$ ,  $p = 0.841$ ), text + graph ( $t = 0.851$ ,  $p = 0.403$ ), or video ( $t = 0.598$ ,  $p = 0.555$ ), there was no significant difference in learner attention between sequential and global learning styles.

## 5. Conclusions and Discussion

**5.1. Research Conclusions and Contributions.** This paper has made an empirical study on the attention of learners on different learning media in a mobile learning environment through questionnaires and psychophysiological methods (EEG recording). This study improved the stability and accuracy of the attention data output by the TGAM chip without increasing the real-time data transmission delay, optimized



the impact of blinking on EEG data, and improved the expressiveness of the data. The researchers have chosen 28 college students as the participants and the NeuroSky MindWave headset as the tool to measure attention based on EEG. Through the EEG experiment, we have analyzed the attention difference of students on three media forms (text, text + graph, and video) and have investigated the influence of gender, education, iPad experience, and learning style factors on a student's attention.

The learners expressed greater attention in text media, which possibly resulted from the fact that text media had only words that were easy to concentrate; however, the difference did not reach statistical significance. Also, the male's attention was greater in the three media than the female's, graduate students' attention was greater than that of undergraduates, and participants who used an iPad for more than five years showed stable attention.

According to learning styles, there was a significant difference in attention on video between active learners and reflective learners though there was no significant difference between sequential and global learners. The reason for the results might be related to the nature of the experimental material, which might require further research by using more sensitive learning style tests.

According to this study, in mobile learning, teachers or instructional designers could provide more text media materials for college students, rather than other multimedia. Teachers or instructional designers should appropriately improve the attention of female college students in mobile learning. Experiments have also shown that graduate students' attention was much greater than that of undergraduates. Therefore, appropriate design or strategies would be used to improve the undergraduate's learning attention in mobile learning. In mobile learning, according to the results of learning style measurement, teachers or instructional designers should provide appropriate learning media for students; in particular, reflective learners were more suitable for learning with video media than active learners.

The results of this study confirmed the relationship between media and attention using traditional questionnaires and psychophysiological measures. The results also showed that psychophysiological measures could be used to study learning behavior.

The study had some contributions to related literature. First, the study confirmed that learners' attention is different when using different media in a mobile learning environment, especially video media. Although previous studies have investigated this relationship, they have not focused on mobile learning environments.

Another contribution of this study was the use of EEG recording for research. In the existing studies, it was not sufficient to measure the impact of media on learners in mobile learning only by a questionnaire. This study suggested the use of EEG recordings to research mobile learning, rather than the traditional paper-and-pencil questionnaire.

This study provided a way to study the effect of learning attention. It was necessary for educators or developers of e-learning systems to design appropriate media learning materials based on learners' learning styles.

**5.2. Limitations and Suggestions for Future Research.** The research has certain limitations but also provides some suggestions for future research. First of all, three types of learning media were regarded as the important premise of affecting learners' attention in a mobile learning environment. However, many factors might be considered in future research.

Secondly, compared with the objective tools of EEG, the method of the learning style questionnaire was subjective; thus, this study is not completely objective.

Thirdly, this study used a NeuroSky MindWave headset to measure the attention. Future research should try to analyze the original EEG data (alpha, beta, gamma, delta, and theta), which could lead to more accurate experimental data.

Fourth, experiments were complex and easily affected by the surrounding environment; in order to ensure the cycle of the experiment, the sample size was small, which was a limitation of this study. As the sample size was very limited, the results of statistical analysis were not easy to explain. In EEG studies, the sample size was usually small [42]. In future studies, we would try to increase the sample size and confirm whether the results are valid.

Fifth, the data used in this study was the average value of attention. In most EEG experiments, the average value of attention was selected. In future research, we would try to follow the value of attention at specific time research to explore more effective research methods.

## Data Availability

If the original data of this study is needed, please contact the author by email (249962743@qq.com).

## Consent

Informed consent was obtained from all individual participants included in the study references.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

Thanks are due to all participants and researchers during the experiment. This study was supported by the Education and Teaching Reform Foundation of the Education Department of Jilin Province, China, in 2016 (Cultivating Talents of Information Technology by Industry-University-Research Cooperation in the Background of the Internet Plus Era).

## References

- [1] A. Collins and R. Halverson, *Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America*, Teachers College Press, 2018.
- [2] A. Larasati, A. M. Hajji, and A. N. Handayani, "Preferences analysis of engineering students on choosing learning media using support vector machine (SVM) model," *Advances in*

- Social Science, Education and Humanities Research*, vol. 242, pp. 57–59, 2019.
- [3] M. L. Chen, *Influence factors research on mobile learning users' continuance usage*, [M.S. thesis], Jiangsu Normal University, 2013.
  - [4] E. Siegenthaler, P. Wurtz, and R. Groner, "Improving the usability of e-book readers," *Journal of Usability Studies*, vol. 6, pp. 25–38, 2010.
  - [5] H. Oz, "Prospective English teachers' ownership and usage of mobile devices as m-learning tools," *Procedia - Social and Behavioral Sciences*, vol. 141, pp. 1031–1041, 2014.
  - [6] V. Sahasrabudhe and S. Kanungo, "Appropriate media choice for e-learning effectiveness: role of learning domain and learning style," *Computers & Education*, vol. 76, pp. 237–249, 2014.
  - [7] I. Reythav and D. Wu, "Mobile collaborative learning: the role of individual learning in groups through text and video content delivery in tablets," *Computers in Human Behavior*, vol. 50, pp. 520–534, 2015.
  - [8] D. H. Jonassen, K. Peck, and B. G. Wilson, *Learning with Technology: A Constructivist Approach*, Merrill, Upper Saddle River, NJ, USA, 1999.
  - [9] G. J. Hwang and C. C. Tsai, "Research trends in mobile and ubiquitous learning: a review of publications in selected journals from 2001 to 2010," *British Journal of Educational Technology*, vol. 42, no. 4, pp. E65–E70, 2011.
  - [10] G. J. Hwang and P. H. Wu, "Applications, impacts and trends of mobile technology-enhanced learning: a review of 2008–2012 publications in selected SSCI journals," *International Journal of Mobile Learning and Organisation*, vol. 8, no. 2, pp. 83–95, 2014.
  - [11] T. Lu and X. Yang, "Effects of the visual/verbal learning style on concentration and achievement in mobile learning," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 14, no. 5, 2018.
  - [12] Y. Y. She and W. C. Du, "Research progress in educational neuroscience," *Researches in Open Education*, vol. 17, no. 4, pp. 12–22, 2011.
  - [13] B. E. Solstad, J. Aloka, K. Odingo, and D. Alexzander, "Mobile education-a glance at the future[EB/OL]," 2007, November 2018, <http://www.dye.no/articles/a-glance-at-the-futere/introduction.html>.
  - [14] T. Chabra and J. Figueiredo, "How to design and deploy and held learning," 2002, November 2018, [http://www.empoweringtechnologies.net/eLearning/eLearning\\_exPov5\\_files/frame.htm](http://www.empoweringtechnologies.net/eLearning/eLearning_exPov5_files/frame.htm).
  - [15] C. L. Ye, F. Y. Xu, and J. Xu, "A review of mobile learning research," *E-education Research*, vol. 3, pp. 12–19, 2004.
  - [16] D. Parsons and H. Ryu, "A framework for assessing the quality of mobile learning," in *In Proceedings of the international conference for process improvement. Research and education*, pp. 17–27, Southampton, UK, 2006.
  - [17] H. Kynäslähti, "In search of elements of mobility in the context of education," in *Mobile Learning*, pp. 41–48, IT-press, Helsinki, 2003.
  - [18] A. Kukulska-Hulme, M. Sharples, M. Milrad, I. Arnedillo-Sanchez, and G. Vavoula, "Innovation in mobile learning," *International Journal of Mobile and Blended Learning*, vol. 1, no. 1, pp. 13–35, 2009.
  - [19] C. M. Chen and S. H. Huang, "Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance," *British Journal of Educational Technology*, vol. 45, no. 5, pp. 959–980, 2014.
  - [20] M. Sharples, *Big Issues in Mobile Learning*, Report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative, University of Nottingham, UK, 2006.
  - [21] C. M. Chen, J. Y. Wang, and C. M. Yu, "Assessing the attention levels of students by using a novel attention aware system based on brainwave signals," *British Journal of Educational Technology*, vol. 48, no. 2, pp. 348–369, 2017.
  - [22] A. Drigas, R. E. Ioannidou, G. Kokkalia, and M. D. Lytras, "ICTs, mobile learning and social media to enhance learning for attention difficulties," *Journal of Universal Computer Science*, vol. 20, no. 10, pp. 1499–1510, 2014.
  - [23] D. J. Lancheros-Cuesta, J. L. R. Arias, Y. Y. Forero, and A. C. Duran, "Evaluation of e-learning activities with NeuroSky MindWave EEG," in *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6, Cáceres, 2018.
  - [24] E. Jamet, M. Gavota, and C. Quaireau, "Attention guiding in multimedia learning," *Learning and Instruction*, vol. 18, no. 2, pp. 135–145, 2008.
  - [25] K. Kinsella, "Understanding and empowering diverse learners in ESL classrooms," *Learning Styles in the ESL/EFL Classroom*, pp. 170–194, Heinle, Boston, MA, USA, 1995.
  - [26] C. Babadoğan, "Öğretim stili odaklı ders tasarımı geliştirme (in Turkish)," *Milli Eğitim Dergisi*, vol. 147, pp. 61–63, 2000.
  - [27] S. Marković and N. Jovanović, "Learning style as a factor which affects the quality of e-learning," *Artificial Intelligence Review*, vol. 38, no. 4, pp. 303–312, 2012.
  - [28] F. Dağ and A. Geçer, "Relations between online learning and learning styles," *Procedia - Social and Behavioral Sciences*, vol. 1, no. 1, pp. 862–871, 2009.
  - [29] R. M. Felder and B. A. Soloman, *Index of Learning Styles*, 1991, <http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSPage.html>.
  - [30] J. L. Alty, A. Al-Sharrah, and N. Beacham, "When humans form media and media form humans: an experimental study examining the effects different digital media have on the learning outcomes of students who have different learning styles," *Interacting with Computers*, vol. 18, no. 5, pp. 891–909, 2006.
  - [31] S.-H. Liu, H.-L. Liao, and J. A. Pratt, "Impact of media richness and flow on e-learning technology acceptance," *Computers & Education*, vol. 52, no. 3, pp. 599–607, 2009.
  - [32] T. L. Good and J. E. Brophy, *Educational Psychology: A Realistic Approach*, Longman, White Plains, NY, USA, 4th edition, 1990.
  - [33] N. O. Keskin and D. Metcalf, "The current perspectives, theories and practices of mobile learning," *Turkish Online Journal of Educational Technology-TOJET*, vol. 10, no. 2, pp. 202–208, 2011.
  - [34] V. Carter, "Do media influence learning? Revisiting the debate in the context of distance education," *Open Learning: The Journal of Open, Distance and e-Learning*, vol. 11, no. 1, pp. 31–40, 2006.
  - [35] N. B. Hastings and M. W. Tracey, "Does media affect learning: where are we now?," *TechTrends*, vol. 49, no. 2, pp. 28–30, 2005.
  - [36] R. E. Mayer, "Multimedia Learning," *Psychology of Learning and Motivation*, vol. 41, pp. 85–139, 2002.

- [37] R. E. Clark, "Media will never influence learning," *Educational Technology Research and Development*, vol. 42, no. 2, pp. 21–29, 1994.
- [38] N. Maniar, E. Bennett, S. Hand, and G. Allan, "The effect of mobile phone screen size on video based learning," *Journal of Software*, vol. 3, no. 4, pp. 51–61, 2008.
- [39] R. B. Kozma, "Will media influence learning? Reframing the debate," *Educational Technology Research and Development*, vol. 42, no. 2, pp. 7–19, 1994.
- [40] P. L. Nunez and R. Srinivasan, "A theoretical basis for standing and traveling brain waves measured with human EEG with implications for an integrated consciousness," *Clinical Neurophysiology*, vol. 117, no. 11, pp. 2424–2435, 2006.
- [41] L. E. Nacke, S. Stellmach, and C. A. Lindley, "Electroencephalographic assessment of player experience," *Simulation & Gaming*, vol. 42, no. 5, pp. 632–655, 2011.
- [42] C. C. Wang and M. C. Hsu, "An exploratory study using inexpensive electroencephalography (EEG) to understand flow experience in computer-based instruction," *Information & Management*, vol. 51, no. 7, pp. 912–923, 2014.
- [43] M. Y. Ma and C. C. Wei, "A comparative study of children's concentration performance on picture books: age, gender, and media forms," *Interactive Learning Environments*, vol. 24, no. 8, pp. 1922–1937, 2015.
- [44] R. Shadiev, Y. M. Huang, and J. P. Hwang, "Investigating the effectiveness of speech-to-text recognition applications on learning performance, attention, and meditation," *Educational Technology Research and Development*, vol. 65, no. 5, pp. 1239–1261, 2017.
- [45] R. Shadiev, T. T. Wu, and Y. M. Huang, "Enhancing learning performance, attention, and meditation using a speech-to-text recognition application: evidence from multiple data sources," *Interactive Learning Environments*, vol. 25, no. 2, pp. 249–261, 2017.
- [46] E. F. Smithson, R. Phillips, D. W. Harvey, and M. C. H. J. Morrall, "The use of stimulant medication to improve neurocognitive and learning outcomes in children diagnosed with brain tumours: a systematic review," *European Journal of Cancer*, vol. 49, no. 14, pp. 3029–3040, 2013.
- [47] J. C.-Y. Sun, "Influence of polling technologies on student engagement: an analysis of student motivation, academic performance, and brainwave data," *Computers & Education*, vol. 72, pp. 80–89, 2014.
- [48] J. Xu and B. Zhong, "Review on portable EEG technology in educational research," *Computers in Human Behavior*, vol. 81, pp. 340–349, 2018.
- [49] N. B. A. Rashida, M. N. B. Taibb, S. B. Liasb et al., "Summative EEG-based assessment of the relations between learning styles and personality traits of openness," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 98–104, 2013.
- [50] D. Deenadayalan, A. Kangaammal, and B. K. Poornima, "EEG based learner's learning style and preference prediction for E-learning," in *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2018 2nd International Conference, pp. 316–320, Palladam, India, 2018.
- [51] G. Rebolledo-Mendez, I. Dunwell, E. Martínez-Mirón et al., "Assessing Neurosky's usability to detect attention levels in an assessment exercise," in *International Conference on Human-Computer Interaction*, pp. 149–158, Springer, 2009.
- [52] K. Dobosz and P. Wittchen, "Brain-computer interface for mobile devices," *Journal of Medical Informatics & Technologies*, vol. 24, 2015.

## Research Article

# A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI

Qi Xiong <sup>1,2</sup>, Xinman Zhang <sup>1</sup>, Wen-Feng Wang <sup>3</sup> and Yuhong Gu <sup>4</sup>

<sup>1</sup>School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710000, China

<sup>2</sup>International College, Hunan University of Arts and Sciences, Changde 415000, China

<sup>3</sup>School of Electronic and Electrical Engineering, Shanghai Institute of Technology, Shanghai 200235, China

<sup>4</sup>Shihezi Medical School, Shihezi 832000, China

Correspondence should be addressed to Wen-Feng Wang; wangwenfeng@sit.edu.cn

Received 23 January 2020; Revised 28 February 2020; Accepted 10 March 2020; Published 27 May 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Qi Xiong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we present a parallel framework based on MPI for a large dataset to extract power spectrum features of EEG signals so as to improve the speed of brain signal processing. At present, the Welch method has been widely used to estimate the power spectrum. However, the traditional Welch method takes a lot of time especially for the large dataset. In view of this, we added the MPI into the traditional Welch method and developed it into a reusable master-slave parallel framework. As long as the EEG data of any format are converted into the text file of a specified format, the power spectrum features can be extracted quickly by this parallel framework. In the proposed parallel framework, the EEG signals recorded by a channel are divided into  $N$  overlapping data segments. Then, the PSD of  $N$  segments are computed by some nodes in parallel. The results are collected and summarized by the master node. The final PSD results of each channel are saved in the text file, which can be read and analyzed by Microsoft Excel. This framework can be implemented not only on the clusters but also on the desktop computer. In the experiment, we deploy this framework on a desktop computer with a 4-core Intel CPU. It took only a few minutes to extract the power spectrum features from the 2.85 GB EEG dataset, seven times faster than using Python. This framework makes it easy for users, who do not have any parallel programming experience in constructing the parallel algorithms to extract the EEG power spectrum.

## 1. Introduction

EEG is a recorded signal of electrical activity of the brain, which is collected from the scalp through electrodes. The application of EEG has important practical value in medical treatment, military, sports, and the intelligence fields, which has been widely recognized by all the researchers. So far, scientists from various disciplines have achieved good results in this field. For instance, scientists from American Wadsworth Centre help the paralyzed people input 36 characters via signals instead of using their own fingers. These signals correspond to specific brain activity. In China, scientists from Tsinghua University have designed an automatic dialling system. The system is connected to the computer for real-time dialling by interpreting the thinking mode of the brain as corresponding numbers [1–6]. The

research of pattern recognition of the EEG signal includes the following steps: data collection, data storage, data processing, data classification, and recognition as shown in Figure 1.

In order to record EEG signals, several electrodes need to be placed on the scalp. Traditional device usually has 20 electrodes. However, recently, EEG devices with as many as 256 electrodes have been used, as shown in Figure 2(a). The increase in the number of electrodes enables recording of huge data thereby making the data processing stage in Figure 1 more important and complicated. This not only consumes a lot of computer resources but also leads to poor data extraction thereby directly affecting the accuracy of classification. There are many signal processing methods which can be used to extract EEG features with good discrimination. These methods include time domain analysis,



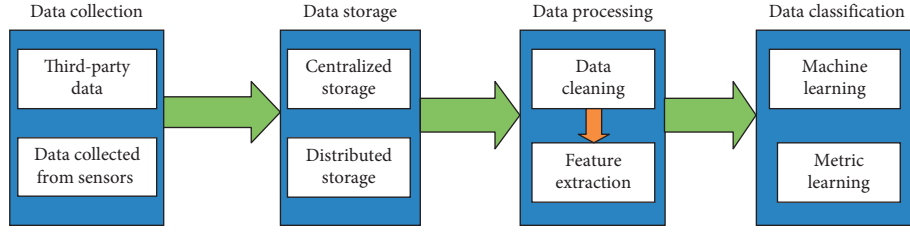


FIGURE 1: Steps of pattern recognition of the EEG signal.

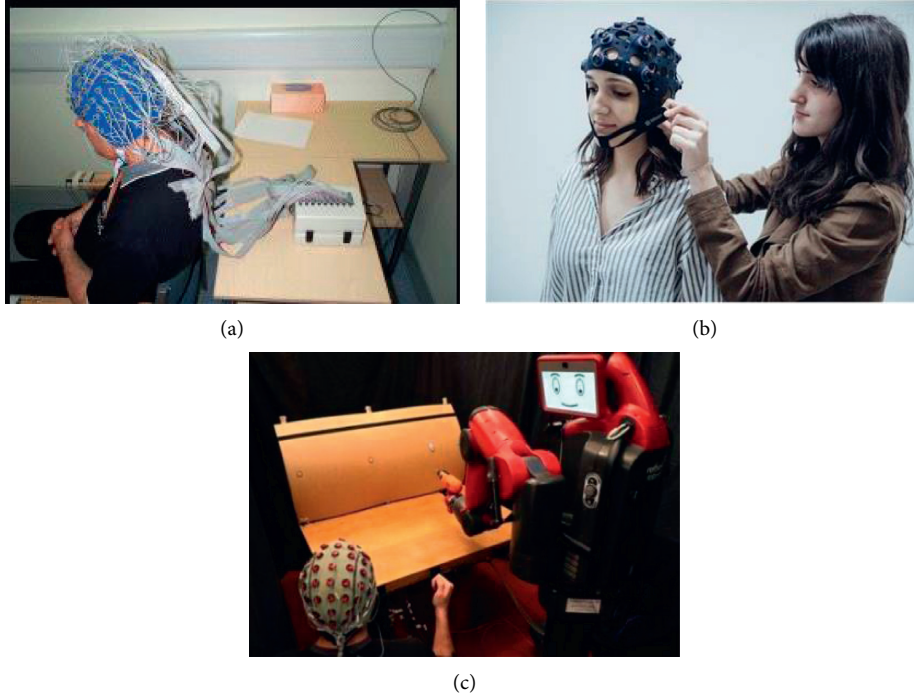


FIGURE 2: (a) 256-electrode device, (b) wearable EEG device, and (c) EEG device for disable.

frequency domain analysis, and time-frequency analysis. A comparative analysis of different approaches to spectral signal representation was performed in the paper [7]. These approaches include power spectral density (PSD) techniques, atomic decompositions, time-frequency ( $t$ - $f$ ) energy distributions, and continuous and discrete wavelet approaches, from which band power features can be extracted and used in the framework of motor imagery (MI) classification. It is pointed out that, among all the feature types of EEG signals, PSD approaches demonstrate to be the most consistent, robust, and effective in extracting the distinctive spectral patterns for accurate discrimination between left and right MI-induced EEGs.

At present, many methods are used to calculate the PSD of signals. The Welch method is one of the most popular, in which users calculate the PSD of the EEG signal in Python or Matlab environment. The function `Scipy.signal.welch` is used in Python and `pwelch` is used in Matlab. In case of small amount of data, PSD of the EEG signal can be quickly obtained by those two functions. But, from Figures 2(b) and 2(c), we can see that, with rapid developments in science and information technology and

the upcoming of 5G era, wearable devices will widely be preferred. EEG signals of human activities can be collected by these devices. We can collect the current working or learning status of some specific people (such as drivers and students) via analysis of their EEG signals [8, 9]. If we calculate the PSD in the Matlab or Python environment, it will take a long time which we cannot bear it. Therefore, we focus on how to calculate the PSD of EEG signals by parallel approaches.

Parallel approaches can improve processing speeds but will need technologies that support distributed computations. Nowadays, there are two commonly different frameworks for big data analysis. One is Apache Spark, and the other is OpenMP/MPI. The two computational frameworks are compared and analyzed in literature [10]. Apache Spark has the advantage of good data management. OpenMP/MPI is faster than Apache Spark by an order of magnitude. Motivated by this analysis, we present a parallel framework for the large dataset to extract power spectrum features of EEG signals, which can be implemented in Linux and MPI environment.

The main contributions of this paper are threefold:



- (1) According to the principle of Welch algorithm, we propose a parallel framework of Welch algorithm, PFWelch, to compute PSD of EEG. The architecture of PFWelch is based on the master-slave mode. The EEG signals recorded by each channel are divided into  $N$  overlapping data segments. Then, the  $N$  segments are computed by the master and slave node in parallel. The results are collected and summarized by the master node. The final Welch PSD results of each channel are saved in the text file, which can be read and analyzed by Microsoft Excel.
- (2) A middle file of the specified format was used. There are many kinds of EEG datasets in the world; different datasets have different file formats. In order to process any kind of EEG data by the proposed parallel framework, EEG data need to be converted into middle files. The relationship between the middle file and the parallel framework is depicted in Figure 3.
- (3) A comparative experiment was designed. In the experimental stage, we first run the function pwelch in the Matlab environment to extract PSD features from the EEG signal as a baseline for comparison and subsequently run PFWelch on the Ubuntu platform. The results show that the PFWelch have the same result as function pwelch. After this, we run the Python function Scipy.signal.welch in the same environment with PFWelch. The experimental result shows that the proposed parallel framework is 7 times faster than using Python.

This paper is organized as follows. Section 2 takes a brief overview of the principle of the Welch method, and then presents a serial algorithm of the Welch method and the proposed parallel framework of the Welch method. In Section 3, we present experimental results and analysis. Discussion and conclusions are presented in Section 4.

## 2. Materials and Methods

**2.1. Welch's Method.** The power spectral density (PSD) exhibits how the power is contained in a signal in the frequency domain. Welch's method and the multitaper approach have shown the best performance among the PSD estimators [11]. The Welch algorithm [12] is exhibited in Figure 4.

From Figure 4, we can demonstrate the Welch algorithm in the following mathematical form.

The input signal  $x[n]$ ,  $n = 0, 1, \dots, N-1$  is split into many overlapping segments. In most cases, an overlap of 50% is applied when the input signal is divided into segments. Let the length of each segment be  $L$  and the total number of segments be  $N_s$ . The formula for the data in the  $i$ th segment is as follows:

$$x_i = x \left[ i \times \frac{L}{2} + n \right], \quad \text{where } n = 0, \dots, L-1, i = 0, 1, 2, \dots, N-1. \quad (1)$$

The procedure of the segmentation is illustrated in Figure 5.

The relationship of the sampling length  $N$ , number of overlapping points  $N_D$ , number of segments  $N_s$ , and segmental length  $L$  is

$$N = L + (L - N_D)(N_s - 1). \quad (2)$$

A smooth window  $w(n)$  is applied to each segment. Generally, we usually use the Hamming window. The formula of the Hamming window for each segment is as follows:

$$w(n) = 0.54 - 0.46 \cos \left[ \frac{2n\pi}{L} \right], \quad (3)$$

where  $n = 0, 1, 2, \dots, L-1$ ,  $L$  denotes the length of each segment. Figure 6 displays a 256-point Hamming window in the time domain and frequency domain with Matlab.

The purpose of the window function is to prevent the spectral leakage [13]. Figure 7(a) shows the spectrum leakage of the original signal. Figure 7(b) exhibits that the spectrum leakage can effectively be reduced by the Hamming window.

From formula (3), we can get formula (4) for the  $i$ th segment of data after being windowed:

$$W_i = x_i(n) \times w(n). \quad (4)$$

Fourier transform of each windowed segment is computed. The formula is as follows:

$$A_i(k) = x_i(n)w(n)e^{-j((2\pi)/N)nk}, \quad (5)$$

where  $A_i$  is the Fourier transform result of the  $i$ th windowed segment,  $i = 0, 1, \dots, L-1$ .

The periodogram of each windowed segment is computed by using the following formula:

$$\phi_i = \frac{1}{LU} |A_i(k)|^2, \quad (6)$$

where  $U = (1/L) \sum_{n=0}^{L-1} w^2(n)$  denotes the mean power of the window  $w(n)$ .

So,  $LU = \sum_{n=0}^{L-1} w^2(n)$  denotes the energy of the window function  $w(n)$  with length  $L$ .

Finally, we can get the PSD by the Welch method which is the average of those periodograms, i.e.,

$$S(k) = \frac{1}{L} \sum_{i=0}^{L-1} \phi_i(k). \quad (7)$$

**2.2. Serial Algorithm of the Welch Method.** In order to design a good parallel program, it is necessary to understand the traditional serial algorithm. According to the above description of the Welch method, the serial algorithm of the Welch method is given as Algorithm 1.

### 2.3. Proposed Parallel Framework of the Welch Method

**2.3.1. Program Structure.** According to the steps of serial algorithms of the Welch method in Section 2.2, it can be seen that this algorithm can be implemented in parallel with MPI. The structure of the parallel algorithm is master-slave and is demonstrated as Figure 8.

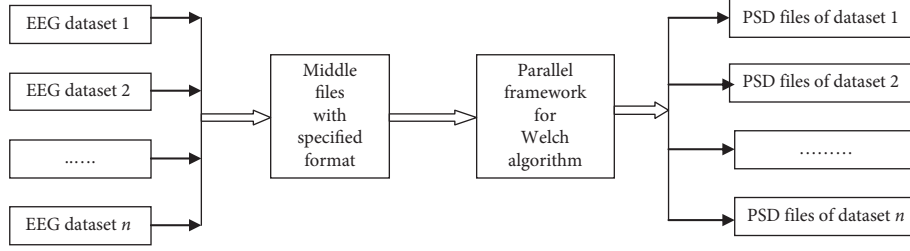


FIGURE 3: Relationship between the middle file and the parallel framework.

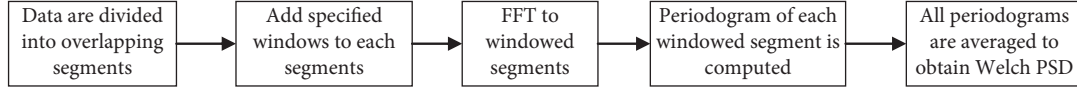


FIGURE 4: Welch PSD algorithm.

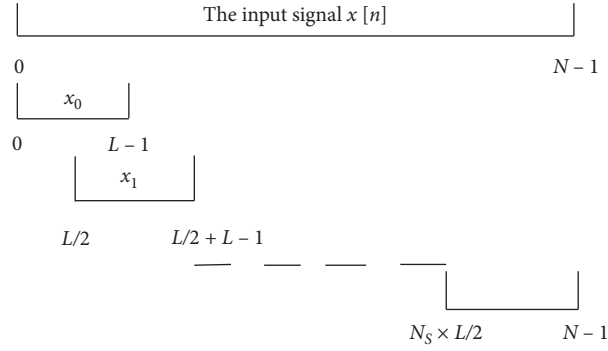


FIGURE 5: Illustration of signal segmentation.

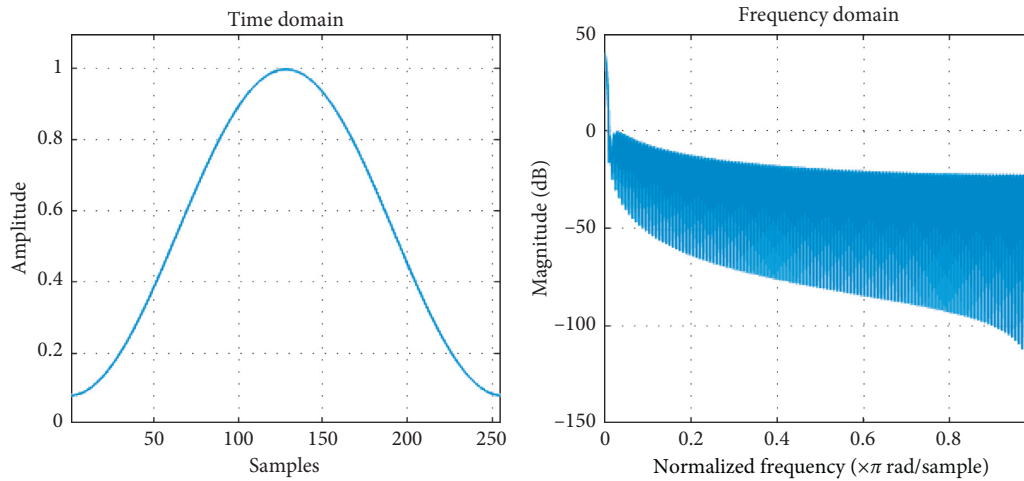


FIGURE 6: Hamming window in the time and frequency domain with Matlab.

The structure of PFWelch algorithm in Section 2.2 contains the following stages: input, split, map, reduce, and output stage. One channel data are split into seven segments, which are allocated to four nodes according to certain rules. All nodes are responsible for computing. In the reduce stage, the master process is responsible for

receiving the result from slave processes and compute the final PSD. Since the parallel implementation of fast Fourier transform has been maturely developed [14–20], we need not give the relevant details. The parallel algorithm of the Welch method is described in Algorithm 2.

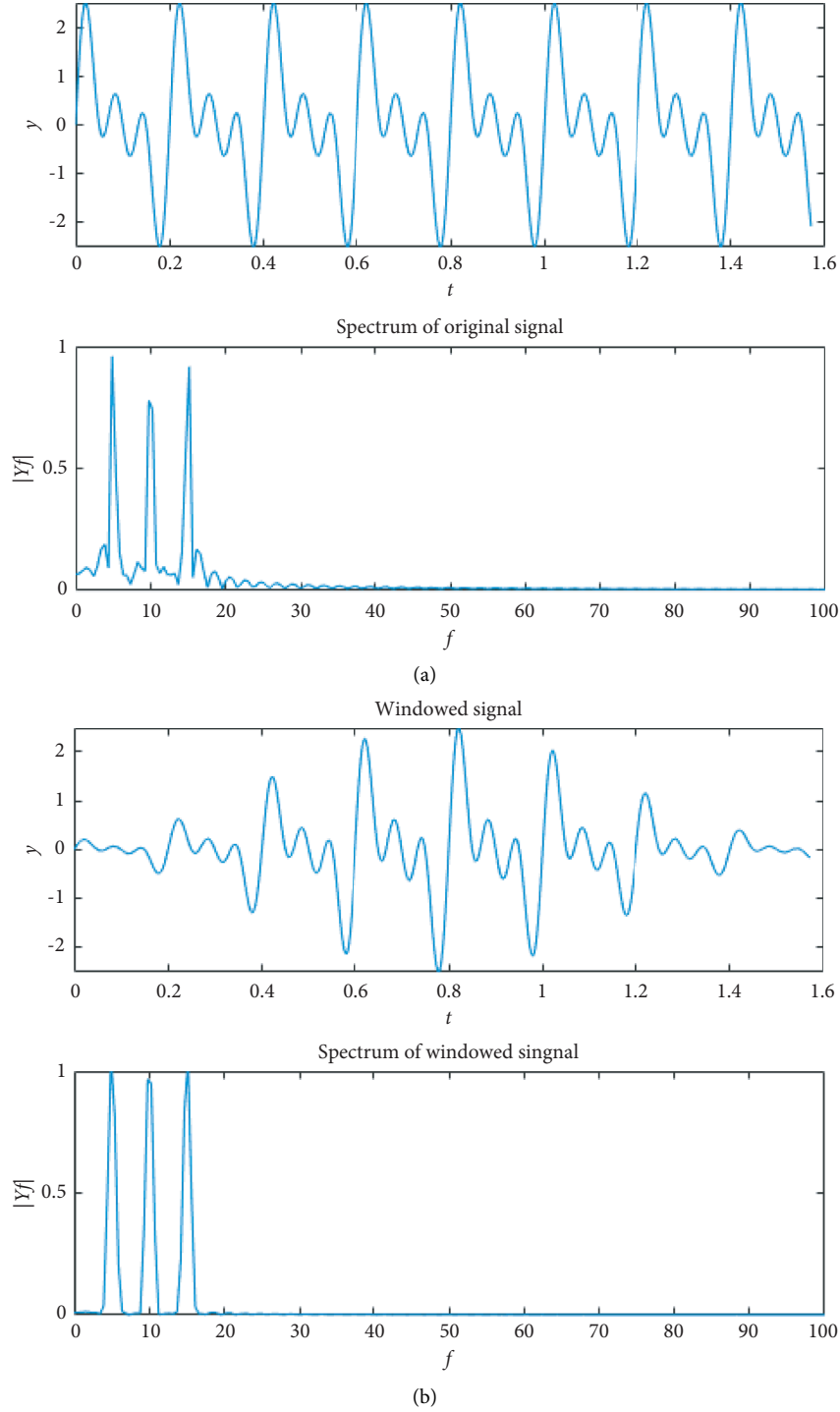


FIGURE 7: (a) The spectrum leakage of the original signal and (b) the windowed signal.

**2.3.2. Distribution of Tasks.** The key point of the parallel program is how to cooperate with each node [21]. How to distribute tasks evenly to each computing node is one of the major factors affecting the performance of the parallel program. From Figure 8, the EEG data of each channel are divided into 7 segments ( $n\_segs=7$ ). The length of each segment is 64. The allocation scheme is

shown in Table 1 when the number of nodes is 4 (size = 4).

According to Table 1, it is convenient to calculate the signal range that each processor needs to compute. For example, processor 0 needs to handle segments 0 and 4. From formula (1), it can be calculated that the signal range is  $[0, 63]$  and  $[128, 191]$ , respectively.

**Input:**

EEG\_file: EEG Signal File  
 welch\_len: the length you want to welch  
 seg\_len: length of each segment  
 Fs: sampling rate

**Output:**

- Wel\_psd: the Welch estimate of PSD
- (1) Initialization:  $over\_len = seg\_len/2$ ; //Length of overlap between the successive segments
  - (2) Data are read from EEG\_file and stored into the array  $welch[n]$ ;
  - (3) Calculate how many segments:  $n\_segs$
  - (4) for index in range (0, seg\_len)
  - (5)   Create Hamming window  $win[i]$ ;
  - (6)    $u += win[index] * win[index]$ ; //Compute the accumulated value of  $u$ ;
  - (7)   index++;
  - (8) end for
  - (9) for seg\_start in range (1, x\_len-seg\_len+1)
  - (10)    $seg\_end = seg\_start + seg\_len - 1$ ; //seg\_start and seg\_end is start point and end point of each data segment, respectively.
  - (11)   apply Hamming window to the data of  $welch[n]$  between  $[seg\_start, seg\_end]$ ;
  - (12)   Do FFT for windowed data. The result is stored in variable  $fft\_x$ ;
  - (13)   Calculating the square of the amplitude of the  $fft\_x$ . The result is stored in variable  $pgram$ ;
  - (14)   Summate power spectra  $Pxy += pgram$ ;
  - (15)    $start\_seg += seg\_len - overlap$ ;
  - (16) end for
  - (17)  $Wel\_psd = Pxy / (n\_ffts \times Fs \times u)$ ;

ALGORITHM 1: Serial algorithm of the Welch method.

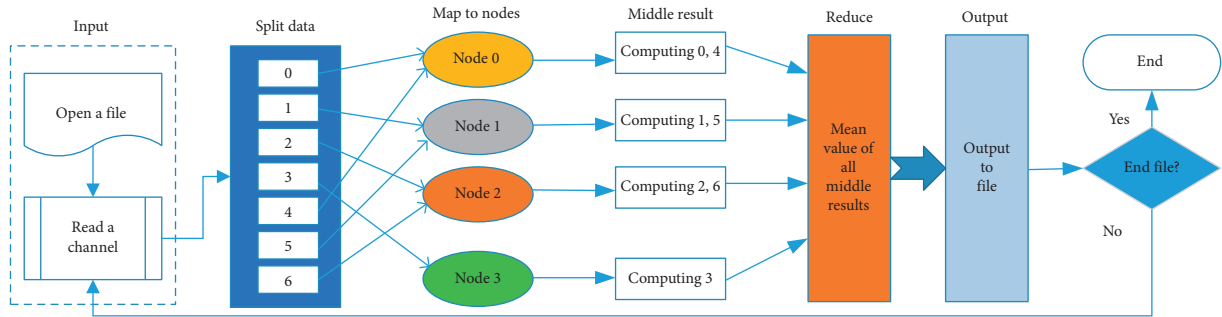


FIGURE 8: Program structure of PFWelch.

### 3. Experimental Results and Analysis

**3.1. Dataset.** In the experiment, the data were downloaded from the website <http://kdd.ics.uci.edu/databases/eeg/eeg.html>. This dataset arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It includes measurements from 64 electrodes placed on the scalp sampled at 256 Hz. There are three versions of the EEG data set: small data set, large data set, and full data set. The full data set contains all 120 trials for 122 subjects. The entire set of data is compressed to about 700 MB [22]. When it was uncompressed, its size is approximately 2.8 GB. After all files were converted into middle files, there were in all 11,058 files.

**3.2. Testing Method and Environment.** There are three different performances which need to be tested:

- (1) *Accuracy.* The result of MATLAB is compared with that of the parallel framework.

- (2) *Speed.* The running time of Python program is compared with that of the parallel framework.

- (3) *Speedup.* The parallel framework was run with different nodes. The running time was recorded, and speedup can be obtained.

The testing environment of hardware is shown in Table 2, and the testing environment of software is shown in Table 3.

**3.3. Experimental Results.** PSD computation of one EEG channel was conducted in the Matlab by using the Welch function, i.e., `pwelch(data, hamming(64), 32, 64, 256)`. The `pwelch` parameters mean that sample data are split up into segments, each segment has 64 data points, the overlap of neighboring segments is 32, Hamming window is applied, and the sampling rate is 256 Hz. Figure 9 shows the result of the Welch method in three different window functions in Matlab.

## Input:

EEG\_file: EEG Signal File  
 welch\_len: the length you want to welch  
 seg\_len: length of each segment  
 Fs: sampling rate

## Output:

Wel\_psd: the Welch estimate of PSD  
 Begin  
 (1) Initialization: MPI\_Init;  
     rank  $\leftarrow$  MPI\_Comm\_rank;  
     size  $\leftarrow$  MPI\_Comm\_size;  
     overlap\_len = seg\_len/2;  
 (2) Read a EEG\_file  
 (3) Data from EEG\_file are stored into the array welch[n];  
 (4) Calculate how many segments: n\_segs  
 (5) for node = 1 to size par-do//Parallel process  
 (6)   apply Hamming window to welch [n]  
     //n between [startPos, stopPos]; startPos and stopPos is starting point and end point of certain data segment, respectively.  
 (7)   Do FFT for windowed data. The result is stored in variable fft\_x;  
 (8)   Calculating the square of the amplitude of the fft\_x. The result is stored in variable pgram;  
 (9)   Summate power spectra Pxy += pgram;  
 (10) end for  
 (11) if (rank == 0)//master process  
 (12)   Receive Pxy from each process and compute average of all periodogram.  
 (13) else//slave processes  
 (14)   Send Pxy to master process  
 END

ALGORITHM 2: Parallel algorithm of the Welch method.

TABLE 1: Task allocation.

Label of the segment	0	1	2	3	4	5	6
Label of nodes	0	1	2	3	0	1	2

TABLE 2: Testing environment of hardware.

Nodes	CPU	Memory	Type
4	Intel(R) core(TM) i5-4460	8 GB	Desktop computer

TABLE 3: Testing environment of software.

Operating system	Parallel environment	Matlab
Ubuntu 16.04	C + Mpcih 3.3.1 on Ubuntu	R2018a on Win10

We performed the same test in the parallel framework. The PSD results were stored in the text file which can be accessed by Microsoft Excel and is shown in Figure 10.

Comparison between Figures 9 and 10 indicates that the results obtained by both Matlab and PFWelch are consistent, which corroborates the correctness of PFWelch. To test the time performance, we first used different number of nodes in different segments to calculate the PSD of one EEG file with PFWelch (in this experiment, we regard different CPU cores as different nodes). Time consumption (measured in seconds) is recorded in Table 4. For comparison, the time cost in Python environment by the serial algorithm is also listed in Table 4. As we have not used Matlab Linux version, we carried out such comparative experiments with Python.

Second, we use different number of nodes with 7 segments to calculate the PSD of all EEG files with PFWelch. The time cost is recorded in Table 5.

Speedup can evaluate the time performance of PFWelch. The definition of speedup is  $\text{speedup} = T_s/T_p$ , where  $T_s$  indicates the time of serial operations and  $T_p$  indicates the time of parallel operations. The speedups are shown in Figure 11. Figure 11 demonstrates that, as the number of nodes increases, the speedup also increases. But, in different cases, the increase is not the same. For all EEG files, Figure 11 shows that, as the number of nodes increases to a certain value, the speedup slowly increases. The reason is the time it takes includes files open and close. This time cannot be changed when the number of nodes is increased.

For a single EEG file, the PFWelch shows the best performance when the signal is split into 7 segments. This is because it has the maximum speedup when the number of nodes is 4. Therefore, we divided the signal into 7 segments to calculate the PSD of all EEG files in Figure 11. If the signal is split into 15 segments, the speedup grows basically in a linear manner, but its speedup is the lowest. Figure 12 clearly reveals the relationship between the number of segments and the speedup.

It can be seen that, by increasing the number of nodes, the speedup is improved. But, merely increasing the number of nodes cannot improve the speedup. From Figure 12, we



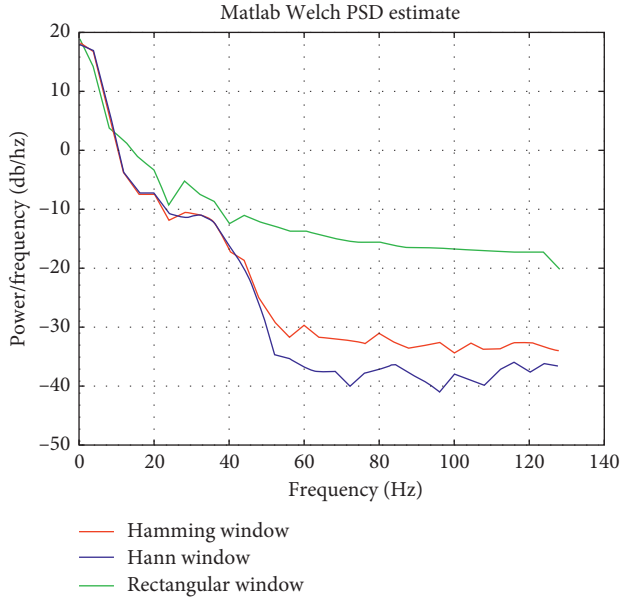


FIGURE 9: Welch result in Matlab.

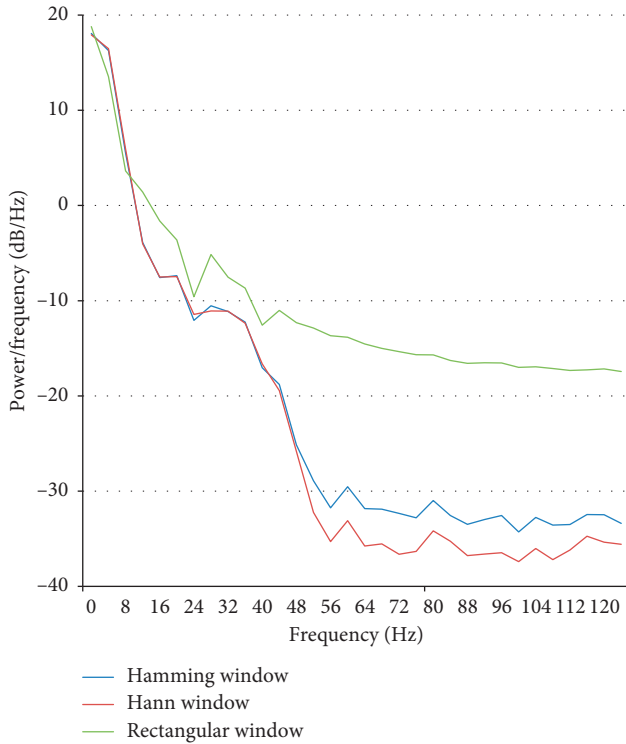


FIGURE 10: Result of PFWelch.

TABLE 4: Time cost of one EEG file.

Segments	Nodes				Python
	1	2	3	4	
7	0.0543	0.0413	0.0382	0.0365	0.228
3	0.075	0.06	0.056	0.056	0.235
15	0.064	0.054	0.05	0.047	0.224

TABLE 5: Time cost of all EEG files.

Nodes	1	2	3	4	Python
Time cost (s)	478.778	354.008	342.715	337.966	2451

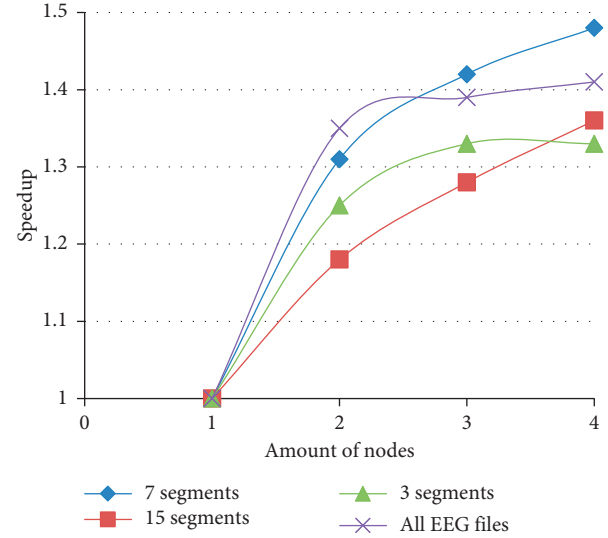


FIGURE 11: Speedup of PFWelch.

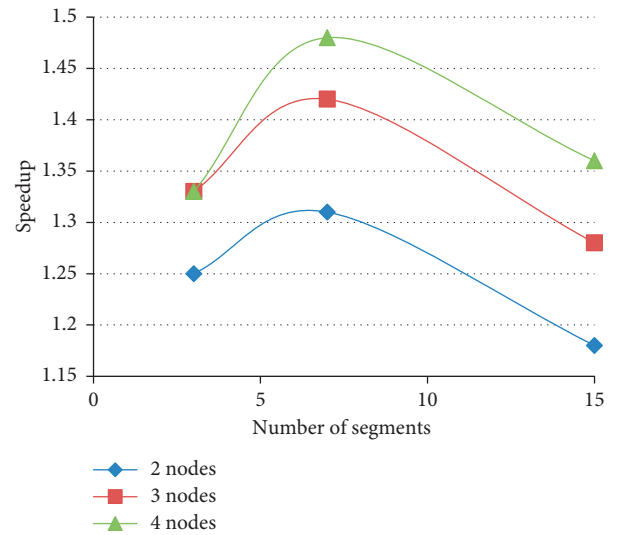


FIGURE 12: Relationship between the number of segments and speedup.

can see that the speedup decreases when the number of segments increases to 15. When the number of segments is 7 and the number of nodes is 4, we get the best speedup performance.

#### 4. Discussion and Conclusion

At present, although there are several methods which can extract features from the EEG signal, the PSD is still one of the most important methods. But, while dealing with a large

number of EEG data, it is necessary to improve the calculation speed of PSD. Therefore, the parallel framework proposed in this paper is used to solve the problem of taking long time to extract PSD features from the EEG dataset in big data environment. The framework is based on C+MPI language and is completed by the master-slave mode. Compared with the traditional serial Welch method, this framework divides the signal into  $N$  segments and distributes them evenly to different nodes on which PSD can be calculated in parallel. In the experiment, the values of  $N$  are 3, 7, and 15, respectively. The number of nodes is from 1 to 4. For the given data set, we found that, although the speedup can be improved by increasing the nodes, the speedup performance will decrease if there are too many segments. Experiments show that the best performance of the parallel framework can be achieved only when the number of nodes and segments is reasonably selected. The speed of PFWelch is 7 times faster than using Python in the same hardware and operating system platform.

Because of the powerful function of MPI, this framework can be deployed not only on the cluster but also on the desktop computer, which is very convenient for the users. The experimental results also corroborate that the framework is correct, efficient, and has a good practical value. It can be applied to extract all kinds of EEG datasets with a little modification. Researchers who are interested in PFWelch can download the source code from <https://github.com/abxcq>.

## Data Availability

All the data utilized in our research can be accessed from <http://kdd.ics.uci.edu/databases/eeg/eeg.html>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Scientific Research Project of Education Department of Hunan Province (no. 17A148), the National Natural Science Foundation of China (nos. 61673316 and 11601339), and the Natural Science Foundation of Hunan Province (no. S2019JJSSLH0130).

## References

- [1] B. He, L. Astolfi, P. A. Valdes-Sosa et al., "Electrophysiological brain connectivity: theory and implementation," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 7, pp. 2115–2137, 2019.
- [2] Y. Fang, M. Chen, and X. Zheng, "Extracting features from phase space of EEG signals in brain-computer interfaces," *Neurocomputing*, vol. 151, no. 10, pp. 1477–1485, 2015.
- [3] M. H. Shi, C. L. Zhou, J. Xie et al., "Electroencephalogram-based brain-computer interface for the Chinese spelling system: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 3, 2018.
- [4] B. Yan, S. He, and K. Sun, "Design of a network permutation entropy and its applications for chaotic time series and EEG signals," *Entropy*, vol. 21, no. 9, p. 849, 2019.
- [5] Y. Dai, X. Zhang, Z. Chen, and X. Xuebin, "Classification of electroencephalogram signals using wavelet-CSP and projection extreme learning machine," *Review of Scientific Instruments*, vol. 89, no. 7, Article ID 074302, 2018.
- [6] S. He, K. Sun, and R. Wang, "Fractional fuzzy entropy algorithm and the complexity analysis for nonlinear time series," *The European Physical Journal Special Topics*, vol. 227, no. 7–9, pp. 943–957, 2018.
- [7] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, 2008.
- [8] Y. Gu, E. Cleeren, J. Dan et al., "Comparison between scalp EEG and behind-the-ear EEG for development of a wearable seizure detection system for patients with focal epilepsy," *Sensors*, vol. 18, no. 2, p. 29, 2017.
- [9] J. W. Ahn, Y. Ku, D. Y. Kim, J. Sohn, J.-H. Kim, and H. C. Kim, "Wearable in-the-ear EEG system for SSVEP-based brain-computer interface," *Electronics Letters*, vol. 54, no. 7, pp. 413–414, 2018.
- [10] J. L. Reyes-Ortiz, L. Oneto, and D. Anguita, "Big data analytics in the cloud: spark on hadoop vs MPI/openMP on beowulf," *Procedia Computer Science*, vol. 53, pp. 121–130, 2015.
- [11] M. G. Jahromi, H. Parsaei, A. Zamani, and D. W. Stashuk, "Cross comparison of motor unit potential features used in EMG signal decomposition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 1017–1025, 2018.
- [12] K. K. Parhi and M. Ayinala, "Low-complexity welch power spectral density computation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 1, pp. 172–182, 2014.
- [13] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.
- [14] M. Ayinala, M. Brown, and K. K. Parhi, "Pipelined parallel fft architectures via folding transformation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 6, pp. 1068–1081, 2012.
- [15] T. Cho, H. Lee, J. Park, and C. Park, "A high-speed low-complexity modified radix-2<sup>5</sup> FFT processor for gigabit WPAN applications," in *Proceedings of the IEEE International Symposium on Circuits & Systems*, IEEE, Beijing, China, May 2013.
- [16] Z. Wang, L. G. Yeo, W. Li, Y. Yan, Y. Ting, and M. Tomisawa, "A novel FFT processor for OFDM UWB systems," in *Proceedings of the APCCAS IEEE Asia Pacific Conference on Circuits & Systems*, IEEE, Eindhoven, Netherlands, April 2006.
- [17] S. Qiao, Y. Hei, B. Wu, and Y. Zhou, "An area and Power efficient FFT processor for UWB systems," in *Proceedings of the 2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Shanghai, China, September 2007.
- [18] S. N. Tang, J. W. Tsai, and T. Y. Chang, "A 2.4-GS/s fft processor for ofdm-based wpan applications," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 6, pp. 451–455, 2010.
- [19] M. Garrido, K. K. Parhi, and J. Grajal, "A pipelined FFT architecture for real-valued signals," *IEEE Transactions on*

- Circuits and Systems I: Regular Papers*, vol. 56, no. 12, pp. 2634–2643, 2009.
- [20] M. Ayinala, K. K. Parhi, Parhi, and K. Keshab, “FFT architectures for real-valued signals based on Radix-2{3} and Radix-2{4} algorithms,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 9, pp. 2422–2430, 2013.
- [21] L. Xiao-Ping, A. N. Zhu-Lin, and Z. Li-Ping, “Master-slave parallel genetic algorithm framework on MPI,” *Acta Simulata Systematica Sinica*, vol. 16, no. 9, pp. 1938–1827, 2004.
- [22] <http://kdd.ics.uci.edu/databases/eeg/eeg.html>.

## Research Article

# EEG Signal and Feature Interaction Modeling-Based Eye Behavior Prediction Research

Pengcheng Ma and Qian Gao 

*School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan Shandong 250353, China*

Correspondence should be addressed to Qian Gao; [qiangao0603@163.com](mailto:qiangao0603@163.com)

Received 25 February 2020; Accepted 20 April 2020; Published 16 May 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Pengcheng Ma and Qian Gao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, with the development of brain science and biomedical engineering, as well as the rapid development of electroencephalogram (EEG) signal analysis methods, using EEG signals to monitor human health has become a very popular research field. The innovation of this paper is to analyze the EEG signal for the first time by building a depth factorization machine model, so that on the basis of analyzing the characteristics of user interaction, we can use EEG data to predict the binomial state of eyes (open eyes and closed eyes). The significance of the research is that we can diagnose the fatigue and the health of the human body by detecting the state of eyes for a long time. On the basis of this inference, the proposed method can make a further useful auxiliary support for improving the accuracy of the recommendation system recommendation results. In this paper, we first extract the features of EEG data by wavelet transform technology and then build a depth factorization machine model (FM+LSTM) which combines factorization machine (FM) and Long Short-Term Memory (LSTM) in parallel. Through the test of real data set, the proposed model gets more efficient prediction results than other classifier models. In addition, the model proposed in this paper is suitable not only for the determination of eye features but also for the acquisition of interactive features (user fatigue) in the recommendation system. The conclusion obtained in this paper will be an important factor in the determination of user preferences in the recommendation system, which will be used in the analysis of interactive features by the graph neural network in the future work.

## 1. Introduction

The electrical activity of the cerebral cortex is recorded by detecting electrodes, and the potential amplitude is taken as the vertical axis and the time as the horizontal axis to form a map that can reflect the spontaneous and rhythmic electrical activity of different parts of the brain with time, which is called EEG [1]; its full name is electroencephalogram. In 1924, Hans Berger, a German psychiatrist, first discovered and recorded the regular electrical activity of the human brain [2]. In 1933, Berger's research was affirmed by the famous British physiologist E.D. Adrian. Since then, EEG has developed rapidly, and the detection methods which include multilead regular electroencephalogram (spontaneous EEG and evoked EEG), sleep EEG, dynamic EEG, video EEG, and other detection methods have occurred [3–5]. The

structure of the human brain is very complex and sophisticated. From the microscopic point of view, it is mainly composed of billions of neurons. The EEG signal consists of two parts, one is the pyramidal neurons in the cortex and the other one is the postsynaptic potential difference of the vertical dendrites. The bioelectric signal of the central nervous system is used to transmit, store, and process various physiological activity information, so as to control the human behavior [6]. As an action related to EEG, the eye state (open and close eyes) can be identified by observing the characteristic changes of EEG signals.

As the key technology of brain computer interface (BCI), EEG can be applied in five stages [7]. Among them, there are three important contents for signal processing, which are the preprocessing, feature extraction, and classification of the EEG signal; all above have been widely studied [8, 9]. As early

as the end of the 19th century, British physiologist Richard Caton took the lead in using a galvanometer to capture the weak current signal on the surface of the animal's cerebral cortex [10]. Through the study of the captured electrical signals of animals, the researchers found that in the state of quiet without external stimulation, the captured current signal waveform showed rhythmic oscillation. In recent years, with the development of computer science and technology, advanced technologies such as X-ray electronic computed tomography and positron emission computed tomography have been introduced in EEG research [11]. EEG is also commonly used to monitor a patient's sleep, anesthesia depth, and fatigue. This article is to monitor the state of eyes (fatigue degree) through EEG.

A large number of scholars are trying to detect the fatigue state through EEG. The intelligent transportation laboratory (University of Pennsylvania) and NHTSA [12] use a brain wave tester and head motion detector to test and analyze the EEG and eye characteristic parameters of the driver under a fatigue state and finally determine PERCLOS (ratio of cumulative time of eye closure to unit time) as the test index of the driver's physiological fatigue degree evaluation. Bergasa et al. [13] (Australia) tested several nondrivers, and based on the changes of the brain wave in a nonfatigue state, they further analyzed the characteristics of brain wave changes in five stages of nonfatigue, near fatigue, moderate fatigue, doze, and antifatigue (wake up from fatigue). Yoshihiro Takei et al. (Japan University of Technology, Zhipu) obtained the steering angle signal of drivers when turning in the process of driving simulation on the basis of the simulation experiment test. They first processed the steering angle signal with the help of Fourier transform theory and wavelet theory. Then, they determined the wavelet transform value of the steering angle signal of drivers in different mental states by using nonlinear theory (chaos theory), so as to judge the driver's fatigue level. Finally, the fast transform algorithm (FFT, fast wavelet transform) is discussed to realize the real-time evaluation of driver fatigue state. According to four kinds of typical brain waves ( $\delta$  wave, frequency is 1-3.5 Hz;  $\alpha$  wave, frequency is 7.5-12.5 Hz;  $\beta$  wave, frequency is 12.5-30 Hz; and  $\theta$  wave, frequency is 3.5-7.5 Hz) and their changes, we can reflect people's mental state. Relevant research shows that the change of  $\alpha$  wave is the best reflection of human fatigue [14]. W-Bang et al. collected four kinds of typical brain waves, obtained the corresponding brain wave entropy using the thermodynamic entropy theory, and evaluated the fatigue degree of drivers according to the change of the calculated value [15]. Yan et al. [16] conducted experimental tests with the KT98-2000A dynamic electroencephalograph. First, the driver (healthy, not taking any irritating drugs) under nonfatigue and fatigue states was collected to engage in various driving operations (starting, shifting, controlling brain waves during movement, steering...). Next, the power spectrum is obtained, and the power spectrum density estimates of various brain waves are calculated. Finally, the average power spectral density ratios of four typical brainwaves when the driver performs various driving operations such as starting, shifting, braking, and steering under nonfatigue and fatigue conditions are obtained.

Most of the research mentioned above are to judge the fatigue state of the human body by directly recognizing the characteristics of EEG. In this paper, the feature data is extracted from EEG by wavelet transform technology, and then, the eye state is predicted by a classification algorithm, compared with traditional classification algorithms,  $k$ -nearest neighbor algorithm [17–19], naive Bayes algorithm [20–22], artificial neural network (ANN) [23], and linear discriminant analysis [24–27]. In this paper, we build a FM+LSTM model to predict the eye state. The model is combined with the linear interaction characteristics of FM and the nonlinear interaction characteristics of LSTM. It can greatly improve the accuracy of eye state classification and prediction. On the basis of this research, we can also infer the fatigue degree of the human body by monitoring the eye state of the human body for a long time, which can help us to recommend user behavior. For example, it can help us to detect whether the user is in the state of work overload in working time or not by combining the proposed eye state classification prediction method and our previous interaction characteristic research results. Meanwhile, it can also help us to detect whether the fatigue state of users is caused by non-working reasons in nonworking time or not. Thus, more accuracy user behavior recommendations can be got.

## 2. Related Work

The purpose of this paper is to predict the eye state (open and closed eyes) by EEG data, which is a typical binary classification problem. Because the traditional binary classification has many shortcomings in some aspects, and its performance is poor, we use the FM algorithm which is widely used in the click rate prediction of recommendation system, as well as a good binary classification algorithm. The advantages of the FM algorithm are as follows: (1) FM model can carry out reasonable parameter trajectory in very sparse data; (2) the complexity of the FM model is linear, the optimization effect is good, and it does not need to rely on a support vector like support vector machines (SVM); and (3) FM is a general model, which can be used in any case where the eigenvalue is real. Other factorization models can only be used in some cases where the input data is relatively fixed. In addition, we also use LSTM, which is a deep network model with good generalization ability and nonlinear mapping ability. Compared with the recurrent neural network, LSTM solves the gradient explosion problem; therefore, LSTM is a strong classification prediction model.

In the experimental part of this paper, after extracting the corresponding features through wavelet transform technology, one hot coding processing is carried out on these data, which will not only increase the number of input data but also generate a large number of sparse data (a large number of zero data appears in the data set). According to the idea of integrated learning, referring to the DeepFM [28] model (a highly efficient click rate prediction model of recommendation system), we integrate FM and LSTM in parallel, and the FM+LSTM model is constructed, which has the characteristics of automatic low-order feature combination, linear data connection, and sparse matrix friendliness. LSTM can



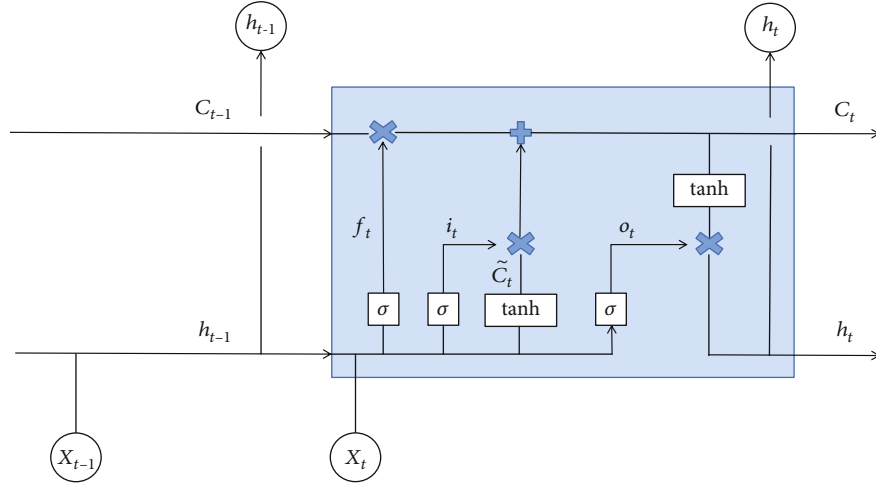


FIGURE 1: Internal structure of LSTM.

solve the gradient explosion problem and has the characteristics of achieving high-order feature combination and nonlinear mapping. In addition, EEG data is obtained from sequential testing in a time series, which is very suitable for the prediction and classification of long-term and short-term memory models. Therefore, the proposed depth factorization machine model is suitable for two classification problems that determine eye behavior. The depth factorization machine model combines the advantages of FM and LSTM, and it is also applicable to two classification problems for judging the eye state. In the following section, we will introduce the FM model and the LSTM model briefly.

**2.1. Introduction to FM.** The factorization machine is a new model proposed by Rendle in 2010, which uses the idea of the implicit factor model and matrix decomposition for reference [29]. The problem of data sparsity can be overcome to some extent by using the idea of matrix decomposition to solve the optimal quadratic parameters. At present, many implicit factor models have been used for rating prediction in recommended fields, such as the common matrix decomposition, and the representative algorithm is singular value decomposition [30]. The main disadvantage of these algorithms is that they are only suitable for specific input data types, and the optimization algorithm is only proposed for the current task, which does not have generality. In different task scenarios, they cannot be directly migrated and extended horizontally. The factorization machine model is a general model, which can change the shortcomings of a traditional matrix decomposition algorithm. The FM only needs to change the form of the input eigenvector to simulate the common matrix decomposition model, while the traditional matrix decomposition model has to define model expression and optimization method separately for each specific task. Therefore, FM effectively avoids this kind of malpractice.

**2.2. Introduction of LSTM.** Because the artificial neural network can model the nonlinear process, it can solve a series of complex problems such as classification, clustering, dimensionality reduction, regression, and structural predic-

tion. With the revolution of the computer industry, the exponential improvement of computer computing power, and the explosive growth of data in recent years, the deep-seated artificial neural network which needs a lot of computing power has been widely used. From the most classical perceptron, there are many kinds of neural network models. The typical models are the Convolutional Neural Network (CNN), recurrent neural network (RNN), and RNN variant LSTM. Different neural network models have different scenarios.

As mentioned above, the cyclic neural network or RNN is a kind of neural network for processing sequence data [31]. The network structure of LSTM was proposed by Hochreiter and Schmidhuber in 1997 [32], and then, this kind of network became very popular. Many people solved many practical problems based on the network structure of LSTM, and now, LSTM is still widely used. The recurrent neural network is a chain loop structure, and the network structure of LSTM is basically the same structure, but LSTM has a more complex structure in the network; therefore, it can deal with long-term dependence. LSTM has three gates to control, which are input gate, forgetting gate, and output gate. The input gate controls the input of the network, the forgetting gate controls the memory unit, and the output gate controls the output of the network. The most important one is the forgetting gate. The function of the forgetting gate is to decide which memories will be preserved and which memories will be forgotten. It is precisely because the function of forgetting gate "LSTM" has a long-term memory function. For a given task, the forgetting gate can be used to learn how many previous memories it can retain, which makes the network have the ability of learning autonomously without human interference.

The following explains the network flow process of LSTM from the specific internal structure. The internal structure of the LSTM is shown in Figure 1. Next, we will explain their internal operation mode and how to represent the three gates mentioned above.

Input gate  $i_t$  controls how much information can flow into memory cells. Forgetting gate  $f_t$  controls how much

information in the memory cells of the last moment can be accumulated into the memory cells of the current moment. Output gate  $o_t$  controls how much information in the memory cells of the current time can flow into the current hidden state  $h_t$ .

*Step one:* use the forgetting gate to decide what information to discard from the cell state and calculate the attenuation coefficient. Read  $h_{t-1}$  and  $x_t$  and output a value between 0 and 1 to each number in cell loading  $C_{t-1}$ . This determines what information we discard from the state of the cell. Since the output of sigmoid is 0 to 1, 1 indicates “full reservation” and 0 indicates “complete abandonment.”

*Step two:* update the information. First, the sigmoid layer is the “input gate layer,” which determines what value we will update. Then, the tanh layer creates a new candidate vector.

*Step three:* update the time of old cell status  $C_{t-1}$ , which is updated to  $C_t$ .

*Step four:* the output gate decides what value to output. The output at this time is calculated according to the  $C_t$  state of the third part.

**2.3. Methodology and Contribution.** The research content of this paper is to detect the eye state (open eyes, closed eyes) through EEG data. The purpose is to detect the eye state for a long time, so as to diagnose the fatigue and health status of individuals. In fact, this is a two classification problem. Specifically, in this paper, we extract feature data from EEG by wavelet transform technology and then build a FM+LSTM model to classify and predict the eye state. This model combines the linear interaction characteristics of FM and the nonlinear interaction characteristics of LSTM, which can greatly improve the classification efficiency. The main contributions of this paper can be summarized as follows:

- (1) The feature is extracted from EEG by wavelet transform:
  - (i) The first-order low-frequency signal features of EEG data are obtained by wavelet transform [33], and there are 14 EEG values
  - (ii) The extracted data features are used as the basis for our classification and prediction
- (2) Build a FM+LSTM model to classify and predict the eye state:
  - (i) This model is the first that we have established and applied to the study of eye behavior classification and prediction based on EEG
  - (ii) The model is composed of FM and short-term memory neural network in a parallel form, which combines the advantages of both models with the characteristics of automatic low-order feature combi-

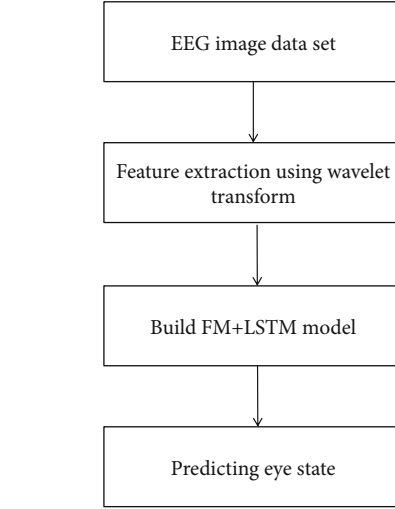


FIGURE 2: The framework of the proposed method.

nation, linear data connection, and sparse matrix friendliness. LSTM can be used to solve the problem of gradient explosion and realize high-order feature combination and nonlinear mapping, which is suitable for the classification and prediction of eye behavior in this paper

- (3) In this paper, the model is tested on the real data set, and through the comparison of the experimental results, it is found that the FM+LSTM model established in this paper has better classification and prediction efficiency than other classification models

### 3. EEG Signal-Based FM+LSTM Model

**3.1. Theoretical Framework.** In the experimental part of this paper, we will first introduce the theoretical framework of the experimental content, as shown in Figure 2.

**3.2. Feature Extraction of EEG by Wavelet Transform.** The EEG signal is a nonlinear and nonstationary random weak signal, with a relatively weak amplitude. Generally, the amplitude of the EEG signal is not more than  $300 \mu V$ . Continuous wavelet transform (CWT) [33–36] essentially convolutes the EEG signal with the translation dilation wavelet weight function with localization in a time and frequency domain, so as to decompose the signal into various components in different times or frequency. Let the signal  $f(t)$  of EEG be the square integrable function, which is recorded as  $f(t) \in L^2(R)$ . Convolute the EEG signal  $f(t)$  with the mother wavelet function, and call Equation (1) the continuous wavelet transform of the EEG signal  $f(t)$ :

$$WT_f(a, \tau) = \langle f(t), \psi_{a,\tau}(t) \rangle = a^{-1/2} \int_R f(t) \overline{\psi\left(\frac{t-\tau}{a}\right)} dt. \quad (1)$$

In the above formula,  $WT_f(a, \tau)$  is the wavelet coefficient of the EEG signal, scale  $a$  controls the expansion and contraction of wavelet function, and translation amount  $\tau$  controls the translation of wavelet function. The scale corresponds to frequency (inverse ratio), and translation  $\tau$  corresponds to time.

If the Fourier transform  $\widehat{\psi}(w)$  of the parent wavelet function satisfies the constant resolution constraint

$$C_\psi = \int_0^\infty \frac{|\widehat{\psi}(w)|^2}{w} dw < \infty, \quad (2)$$

then there is a reconstruction formula for the continuous wavelet transform of the EEG signal, and its expression is shown in

$$f(t) = \frac{1}{C_\psi} \int_0^\infty \frac{da}{a^2} \int_R WT_x(a, \tau) \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) d\tau. \quad (3)$$

In the above formula,  $C_\psi$  is the permissive condition of  $\psi(t)$  and  $WT_x(a, t)$  is the wavelet transform coefficient. Since the wavelet base has two parameters of scale and displacement, the expansion of the wavelet base means that one-hour function is projected on the two-dimensional time-scale phase plane. And because of the characteristics of the basic body of wavelet, the function projection to wavelet transform is conducive to extract some features.

Because of the invariable window and the excellent characteristics of local analysis, the characteristics of the EEG signal in different frequencies can be analyzed by wavelet transform in different time scales.

**3.3. FM+LSTM Model Construction Method.** For a classification and prediction system, it is very important to learn the EEG feature combination behind the real eye state. Among the features extracted from EEG, low-order combined features (including one-dimensional and two-dimensional feature data) or high-order combined features (multidimensional feature data) may affect the final classification and prediction results. Considering the disadvantages of traditional classification methods, it includes the following:

- (1) The decision tree model easily leads to overfitting and ignores the correlation between data
- (2) The random forest model is prone to overfitting when the data noise is large
- (3) The logistic regression model cannot combine features and depends on artificial feature combination
- (4) The prediction process of the SVM model depends on the support vector in the training sample

We can see that the traditional classification algorithm cannot complete the classification task well, and therefore, we use the FM in this paper. The factorization machine model extracts the feature combination through the implicit variable inner product of each one-dimensional feature,

which can extract the combined feature actively, keeps independent of the training sample but not rely on the support vector, and will not easily be affected by the noise data and appear the phenomenon of overfitting. However, in theory, FM can model higher-order feature combinations; in fact, only the second-order feature combination is used because of the complexity of calculation. In this paper, we solve the problem of high-order feature combination by multilayer neural networks which can learn the nonlinear complex relationship. We further use LSTM, which is a variation of RNN. LSTM contains the characteristics of a traditional neural network which is composed of the input layer, hidden layer, and output layer. LSTM also inherits all the advantages of RNN. It can adjust the parameters through back propagation. Moreover, LSTM can make the neurons in the hidden layer communicate with each other and establish the connection between the characteristic data. Not only that, the LSTM also solves the problem that the circulating neural network is prone to gradient explosion. Therefore, the LSTM contains the advantages of many neural networks, and its performance has been improved. And it also has a good performance in the classification problem.

Based on the advantages of the FM model and LSTM, we build the FM+LSTM model on the basis of the DeepFM [28] model to solve the classification problem. It effectively combines the advantages of FM and LSTM in feature learning: It can extract low-order combined features and high-order combined features at the same time. In the FM+LSTM model, we use a parallel approach to build our depth factorization machine model. Through the depth factorization machine model, on the one hand, FM can be used to extract the features from the first-order features and second-order features which are by the combination of the pair first-order features; on the other hand, LSTM can be used to extract features from the higher-order features which is formed by the input first-order features. Specifically, the characteristics of the FM+LSTM model include

- (1) being able to process sparse data
- (2) combining the FM model and the LSTM model, learning low-order feature combination and high-order feature combination at the same time

To sum up, we explained the principle of the FM+LSTM model and the advantages of the model. Next, we will show the mathematical principle of the model.

First, we use the FM [29]. The second-order expression of the algorithm is shown in

$$y(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n w_{ij} x_i x_j, \quad (4)$$

where  $n$  represents the feature dimension,  $x_i$  represents the feature value of the first feature, and  $w_i$  and  $w_{ij}$  are the coefficients of the primary term and the secondary term of the factorization machine model, respectively. A symmetric matrix  $w$  is composed of all the parameters  $w_{ij}$  of quadratic

terms. At this time, the matrix can be decomposed into the following forms:  $w = v^T v$ , where the  $i$ th column of  $v$  is the corresponding vector expression of the  $i$ th feature, and at this time, the coefficient of quadratic terms can be expressed as the inner product of two vectors, namely,  $w_{ij} = \langle v_i, v_j \rangle$ . In this case, the expression of the second-order model is as shown in

$$y_{\text{FM}} = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j, \quad (5)$$

where  $v_i$  is the vector expression of the  $i$ th feature  $k$ , dimension  $k$ , and  $\langle \cdot, \cdot \rangle$  is the vector dot product. At this time, the coefficients of the quadratic terms  $x_h x_i$  and  $x_i x_j$  are no longer independent. Specifically, the coefficients of  $x_h x_i$  and  $x_i x_j$  are  $\langle x_h, x_i \rangle$  and  $\langle x_i, x_j \rangle$ , which have the same term  $v_i$ . Then, all the samples with nonzero feature combination of  $x_i$  can be used to learn  $v_i$ .

In the LSTM model [32], the output is controlled by the forgetting gate, input gate, and output gate. The forgetting gate is used to determine what information is discarded from the cell state, and the attenuation coefficient is calculated as shown in formula (6). Read  $h_{t-1}$  and  $x_t$ , and output a value between 0 and 1 to each number in cell loading  $C_{t-1}$ . This determines what information we discard from the state of the cell. Since the output of sigmoid is 0 to 1, 1 indicates “full reservation” and 0 indicates “complete abandonment”:

$$f_t = \sigma(W_f \cdot [H_{t-1}, x_t] + b_f). \quad (6)$$

In the above formula,  $W_f$  is the weight value,  $t$  is the number of input data,  $t$  is the range of  $t = \{0, 1, \dots, T\}$ ,  $x_t$  is the  $t$ th input data,  $H_{t-1}$  is the output result of  $t-1$  neuron,  $b_f$  is the deviation value, and  $\sigma(\cdot)$  is the activation function sigmoid, which is defined as shown in

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (7)$$

Use the input gate to decide what value we are going to update. Then, the tanh layer creates a new candidate vector. The calculation equation is as follows:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c). \end{aligned} \quad (8)$$

In the above formula,  $\sigma(\cdot)$  is the activation function sigmoid,  $W_i$  and  $W_o$  are weight values,  $b_i$  and  $b_o$  are deviation values, and  $\tanh(\cdot)$  is the activation function. Its definition is as shown in

$$f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (9)$$

When updating the old cell status,  $C_{t-1}$  is updated to  $C_t$ . The calculation formula is shown in

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \quad (10)$$

Use the output gate to decide what value to output. The output at this time is calculated according to the  $C_t$  state of the third part. The calculation process is shown in

$$\begin{aligned} O_t &= \sigma(W_o [h_{t-1}, x_t] + b_o), \\ H_t &= O_t * \tanh(C_t). \end{aligned} \quad (11)$$

In the above formula,  $\sigma(\cdot)$  is the activation function sigmoid,  $W_o$  is the weight value, and  $b_o$  is the deviation value.

The output results of each nerve unit can be obtained through the above steps. Finally, the output results of the neuron (the  $t$ th) can be obtained under the effect of output and transmission of different cycle units:

$$y_{\text{LSTM}} = \sigma(W_t \cdot H_t + b_t). \quad (12)$$

In the above formula,  $\sigma(\cdot)$  is the activation function sigmoid,  $W_t$  is the weight value, and  $b_t$  is the deviation value.

Finally, we combine the output of FM and LSTM to get the prediction results as shown in

$$y_{\text{prediction}} = \text{sigmoid}(y_{\text{FM}} + y_{\text{LSTM}}). \quad (13)$$

We use logloss as the loss function of the depth factor decomposition machine model constructed in this paper, and then use gradient optimization algorithm to adjust the parameters, and finally get the optimal parameters.

We have introduced the mathematical principle of the depth factor decomposition machine model in detail. Next, we will show the overall flow chart of building the FM+LSTM model, as shown in Figure 3.

## 4. Experiment and Execution Results

**4.1. Classified Forecast.** In the task of classification and prediction, we compare the effect of the model used in this paper with the following benchmark models.

- (1) *Decision Tree (DT) Algorithm* [37]. It is a method of approaching the value of discrete function, which is a typical classification method. Firstly, the data is processed, the readable rules and decision trees are generated by using an induction algorithm, and then the new data is analyzed by using decision. In essence, the decision tree is a process of data classification through a series of rules.
- (2) *Random Forest (RF) Algorithm* [38]. Random forest is an algorithm that integrates multiple trees through the idea of integrated learning. Its basic unit is the decision tree, and its essence belongs to a big branch of machine learning—ensemble learning method.

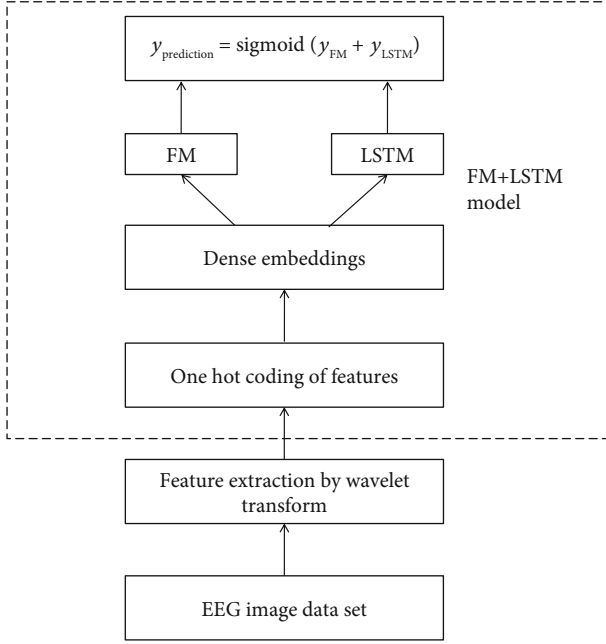


FIGURE 3: Overall flow chart of the FM+LSTM model.

- (3) *Logistic Regression (LR) Algorithm* [39]. It is used to deal with the regression problem when the dependent variable is classified variable. The common problem is the binomial or binomial distribution problem. It can also deal with the multiclassification problem. In fact, it belongs to a classification method.
- (4) *Support Vector Machines* [40]. SVM finds a hyper-plane to divide the data into one class and other classes, which is a two-class classification model. The separation interval is the largest and different from the perceptron.

**4.2. Prediction Scheme.** In order to evaluate the prediction efficiency of the classification prediction task, we use common evaluation indicators: accuracy, precision, recall, and F1-measure. Specifically, the definitions of accuracy, precision, recall, and F1-measure are as follows [26]:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FN + FP}, \\
 \text{Precision} &= \frac{TP}{TP + FP}, \\
 \text{Recall} &= \frac{TP}{TP + FN}, \\
 \text{F1-Measure} &= \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}.
 \end{aligned} \tag{14}$$

TP is the number of correctly recognized closed eyes, FP is the number of incorrectly recognized closed eyes, TN is the number of correctly recognized open eyes, and FN is the number of incorrectly recognized open eyes. The accuracy rate is the ratio of the number of samples correctly classified by the classifier to the total number of samples in a given test

TABLE 1: Classification prediction results of various methods.

Algorithm/model	Accuracy	Precision	Recall	F1-measure
DT	0.82	0.82	0.82	0.82
RF	0.88	0.88	0.88	0.88
LR	0.64	0.64	0.62	0.63
SVM	0.56	0.59	0.50	0.54
FM	0.76	0.80	0.71	0.75
LSTM	0.89	0.90	0.82	0.86
FM+LSTM	0.93	0.94	0.90	0.92

data set, the accuracy is the ratio of the number of correctly predicted samples in all predictions, the recall rate is the correctly predicted positive samples in all actual prediction proportion, and the F1-measure is the harmonic average of the exact value and the recall rate. These four evaluation schemes help us compare the efficiency of different classification models in the EEG data set for eye behavior prediction, and the bigger the results of the four evaluation methods, the closer the prediction results to the actual situation.

In this paper, the accuracy, precision, recall rate, and F1-measure are used to evaluate the prediction efficiency of different models. They are all classic evaluation methods of binary prediction which are suitable for the eye behavior (open and close eyes) classification prediction based on EEG data. Among them, accuracy is the ratio of the number of samples correctly classified by the classifier to the total number of samples in a given test data set, accuracy is the proportion of the number of samples correctly predicted as positive in all predictions, recall rate is the proportion of samples correctly predicted as positive in all actual predictions, and F1-measure is the harmonic average of the accuracy and recall rate. These four evaluation schemes help us to compare the efficiency of different classification models in EEG data set for eye behavior prediction, and the bigger the results of the four evaluation methods, the closer the prediction results to the actual situation.

**4.3. Comparison of Classified Prediction Results.** The following table is the classification prediction performance of the model used in this paper and other benchmark models on the EEG\_EYE data set, and the specific results are shown in Table 1.

As shown in Table 1, we sorted out the classification and prediction results of various classifiers on the EEG data set. In order to compare the effects of various classifier models more intuitively, we drew a histogram as shown in Figure 4.

In the traditional classification prediction model, the logical regression model has the advantages of easy use and understanding but has the disadvantages of not being able to extract the combination features actively; the decision tree model has the advantages of not being sensitive to the missing value and not being able to process the data type and the conventional type attributes at the same time but has the disadvantages of low antioverfitting ability and being easily disturbed by the noise data; the random forest model has the advantages of strong antioverfitting ability and balance



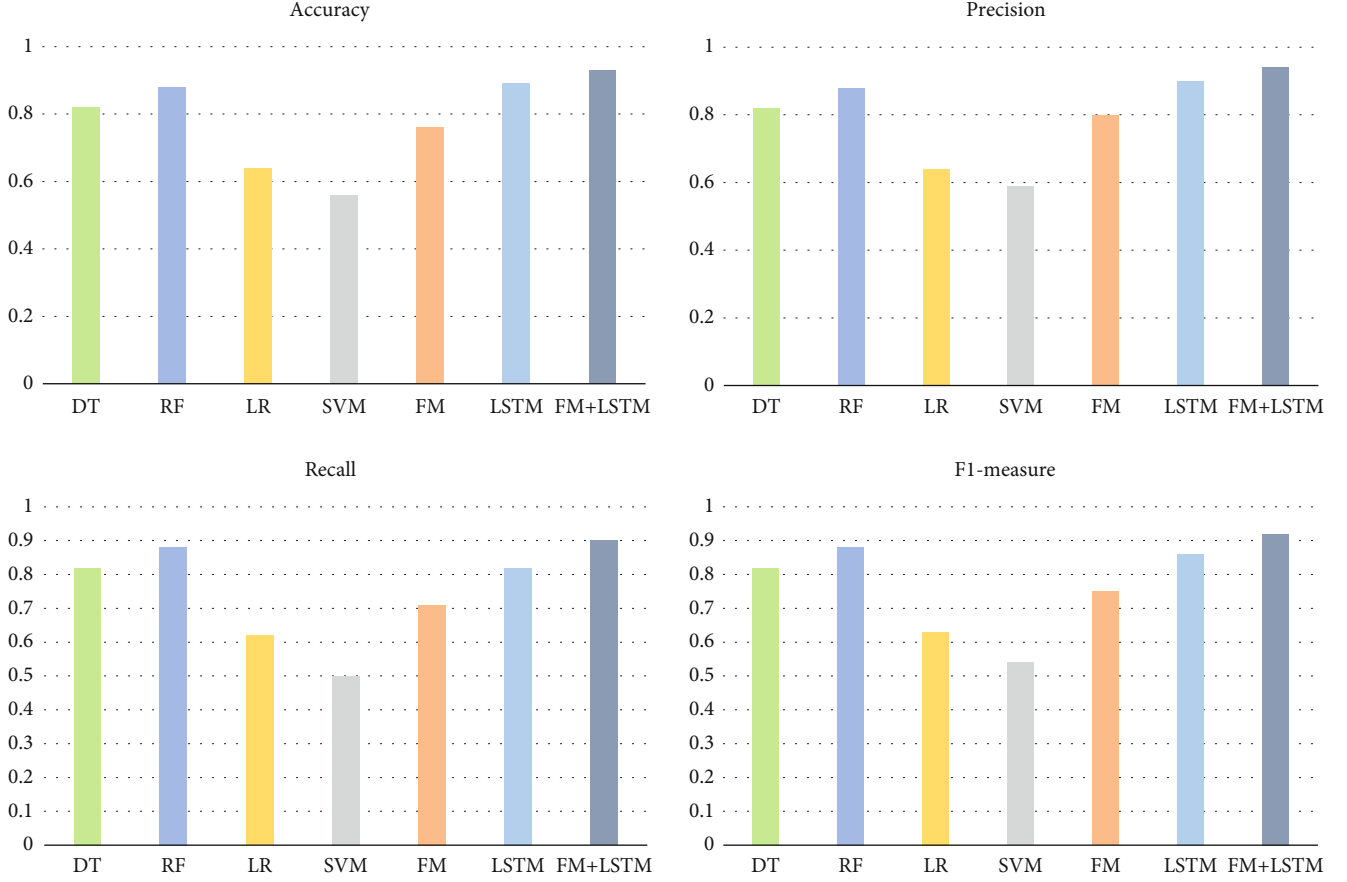


FIGURE 4: Histogram of prediction results of each classification model.

error, but it is also vulnerable to the interference of noise data; the support vector machine model has the advantages of finding the global minimum value through the convex optimization method, but it has the disadvantages of relying too much on support vector and processing large-scale training data; the factorizer model has the advantages of high learning efficiency and processing large-scale sparse data, but it has the disadvantages of only extracting the second-order feature combination; the LSTM model has the advantages of processing large-scale sparse data and solving the problem of long sequence dependence, but it can only extract the high-order feature combination and cannot combine the high-order feature with the low-order feature.

Specifically, by combining the classification results of Table 1 and Figure 2, we can get the following conclusions:

- (1) In the test set of EEG\_EYE data, the overall classification effect of the factorization machine model is better than that of the logistic regression model and the support vector machine model; this shows that the factorization model has the characteristics of extracting the combined features actively and the characteristics of keeping independent of training samples, which makes it better than the above logistic regression model and support the vector machine model in classification and prediction results

- (2) In the test set of EEG\_EYE data, the overall classification effect of LSTM is better than that of the decision tree model and random forest model, which shows that the LSTM has no serious overfitting linearity due to the noise interference in the test set, and therefore, the classification and prediction effect is better than the other two classification models

- (3) In the test set of EEG\_EYE data, the model of FM+LSTM is better than other models in classification and prediction. This is because the model can extract low-order combined features and high-order combined features at the same time, which makes up for the disadvantage that LSTM can only extract high-order feature combinations

In summary, the FM+LSTM model established by us can not only extract the combined features actively but also reduce the noise interference on the simulation. In addition, it can extract both low-order combined features and high-order combined features; therefore, its performance is better than other models.

## 5. Future Work

In this paper, we established a FM+LSTM model to predict the eye behavior (open or close eyes) of the user by EEG data,

so as to infer whether the user is in a fatigue state. Simulation results show that the proposed method achieved good classification and prediction results. The direct significance of this work is that we can judge whether the user is in the fatigue state through the EEG data. Furthermore, the fatigue state is also an important feature of the user in the recommendation system, which can affect the user's preference and evaluation of the items. Therefore, in the future work, first, we plan to further study the user's preference and evaluation of the different items based on the proposed research achievement of the eye's fatigue state which is an important feature factor related to the context, so as to improve the efficiency of recommendation. Second, we plan to use the attention mechanism to distinguish the influence of various features on eye behavior, so as to enhance the interpretability of the classification prediction result. In addition, based on the research results of this paper, we plan to further use the graph neural network model to study the complex interaction between different features in a more flexible and clearer way, so as to obtain more efficient and more accurate recommendation results.

## 6. Conclusions

The purpose of this paper is to predict the eye state (open and closed eyes) by EEG data, which is a typical dichotomous problem. The significance of this study is to diagnose the fatigue and the health of the eyes by detecting the eye state for a long time. And on the basis of this research, we can also infer the fatigue degree of the human body by monitoring the eye state of the human body for a long time, so that we can help in user behavior recommendation based on this inferential. Firstly, we use wavelet transform technology to extract the features of EEG, and then, we use a classifier to classify and predict. Because the traditional binary classification has many shortcomings in some aspects, and its performance is poor, we use the FM algorithm which is widely used in the click rate prediction of the recommendation system and binary classification algorithm. In addition, we also use LSTM, which is a deep network model with good generalization ability and nonlinear mapping ability. According to the idea of integrated learning, referring to the DeepFM [28] model (a highly efficient prediction model of click through rate of the recommendation system), we integrate FM and LSTM in parallel and build the FM+LSTM model, which combines the advantages of FM and LSTM and is suitable for the two classification problems of the eye state in this paper. And its performance in real data set is better than other classifier models. Through this research, we can also infer the fatigue degree of the human body by monitoring the eye state of the human body for a long time, so that we can help in user behavior recommendation based on this inference, which will be the focus of our next research. In this paper, the FM+LSTM model is established to solve the problem of eye behavior classification and prediction. The conclusion obtained in this paper will be an important factor in the determination of user preferences in the recommendation system, which will be used in the analysis of interactive features by the graph neural network in the future work.

## Data Availability

The data set used in this experiment is collected by Oliver rothler from Germany, Stuttgart and Baden Wurttemberg state cooperative University (DHBW). The data set is composed of EEG value and a value indicating eye state, which belongs to the field of life medicine. All the data are the result of continuous EEG measurement with Emotiv. The measurement duration is 117 seconds. During the EEG measurement, the eye state is detected by the camera, and then manually added to the file after analyzing the video frame. "1" means that the eyes are closed, and "0" means that the eyes are open. For more information about the data, please visit: <http://archive.ics.uci.edu/ml/datasets/EEG+eye+state>.

## Conflicts of Interest

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

## Acknowledgments

This work was supported partly by the National Natural Science Foundation of China (61702292 and 61703219), International Cooperation Program for Key Professors of 2017 by the Department of Education of Shandong Province, and Qilu University of Technology (Shandong Academy of Sciences) Young Doctor Cooperative Funding (2017BSH2012).

## References

- [1] M. X. Cohen, "Where does EEG come from and what does it mean?," *Trends in Neurosciences*, vol. 40, no. 4, pp. 208–218, 2017.
- [2] A. Gevins, "Hans Berger was right: what I have learned about thinking from the EEG in the past twenty years," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 1, pp. 5–6, 1997.
- [3] A. M. Coenen, "Neuronal activities underlying the electroencephalogram and evoked potentials of sleeping and waking: implications for information processing," *Neuroscience & Biobehavioral Reviews*, vol. 19, no. 3, pp. 447–463, 1995.
- [4] K. E. Crowley and I. M. Colrain, "A review of the evidence for P2 being an independent component process: age, sleep and modality," *Clinical Neurophysiology*, vol. 115, no. 4, pp. 732–744, 2004.
- [5] S. Y. Tseng, R. C. Chen, F. C. Chong, and T. S. Kuo, "Evaluation of parametric methods in EEG signal analysis," *Medical Engineering and Physics*, vol. 17, no. 1, pp. 71–78, 1995.
- [6] P. X. Wang and Z. J. Zhang, "Study on the recognition algorithm of epileptic carbuncle characteristic wave of EEG signal under wavelet transform," in *Information and computer (theoretical Edition)*, pp. 63–65, China Optical Society, 2017.
- [7] M. Z. Ilyas, P. Saad, and M. I. Ahmad, "A survey of analysis and classification of EEG signals for brain-computer interfaces," in *2015 2nd International Conference on Biomedical Engineering (ICoBE) IEEE*, Penang, Malaysia, March 2015.
- [8] S. Motamedi-Fakhr, M. Moshrefi-Torbati, M. Hill, C. M. Hill, and P. R. White, "Signal processing techniques applied to human sleep EEG signals—a review," *Biomedical Signal Processing and Control*, vol. 10, pp. 21–33, 2014.

- [9] N. R. Tambe and A. Khachane, "Mood based E-learning using EEG," in *2016 International Conference on Computing Communication Control and automation (ICCCUBEA) IEEE*, Pune, India, Aug. 2016.
- [10] R. M. Huang, S. H. Du, Z. Y. Chen, Z. Zhang, and Y. Zhou, "Pattern recognition of epileptic and carbuncle EEG based on support vector machine," *Journal of Biomedical Engineering*, vol. 30, no. 5, pp. 919–924, 2013.
- [11] M. Xi and G. H. Zhu, "Multiscale permutation and its application in epilepsy and carbuncle attack recognition," *Journal of Biomedical Engineering*, vol. 32, no. 4, pp. 751–756, 2015.
- [12] Q. Wu and W. W. Yu, "Integrating a HMM-based driver fatigue recognition model into smart vehicle space for ubiquitous computing," *Key Engineering Materials*, vol. 467–469, pp. 23–30, 2011.
- [13] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 63–77, 2006.
- [14] P. H. Gander, N. S. Marshall, I. James, and L. L. Quesne, "Investigating driver fatigue in truck crashes: Trial of a systematic methodology," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 9, no. 1, pp. 65–76, 2006.
- [15] J. W. Bang, J. S. Choi, E. C. Lee, K. R. Park, and M. Whang, "Noise reduction of EEG signal based on head movement estimation by using frontal viewing camera," *Sensor Letters*, vol. 10, no. 5, pp. 1241–1246, 2012.
- [16] S. Yan, J. Q. Wei, and Y. H. Wu, "Research on feature extraction of brain wave in drivers' sleepy state," *Chinese Journal of Biomedical Engineering*, vol. 24, no. 1, pp. 110–113, 2005.
- [17] X. Y. Wu, S. H. Wang, and Y. D. Zhang, "Overview of k nearest neighbor algorithm theory and application," *Computer Engineering and Application*, vol. 53, no. 21, pp. 1–7, 2017.
- [18] P. Soucy and G. W. Mineau, "A simple KNN algorithm for text categorization," in *A simple KNN algorithm for text categorization*, San Jose, CA, USA, Nov.-Dec. 2001.
- [19] K. Murphy, B. van Ginneken, A. M. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical Image Analysis*, vol. 13, no. 5, pp. 757–770, 2009.
- [20] F. Tahernezhad-Javazm, V. Azimirad, and M. Shoaran, "A review and experimental study on the application of classifiers and evolutionary algorithms in EEG-based brain-machine interface systems," *Journal of Neural Engineering*, vol. 15, no. 2, p. 021007, 2018.
- [21] J. Machado and A. Balbinot, "Executed movement using EEG signals through a naive, Bayes classifier," *Micromachines*, vol. 5, no. 4, pp. 1082–1105, 2014.
- [22] R. M. Mehmood, R. Du, and H. J. Lee, *Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors*, IEEE Access, 2017.
- [23] P. Dande and P. Samant, "Acquaintance to artificial neural networks and use of artificial intelligence as a diagnostic tool for tuberculosis: a review," *Tuberculosis*, vol. 108, pp. 1–9, 2018.
- [24] T. Kim and J. Kittler, "Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 318–327, 2005.
- [25] Y. Lu, H. Jiang, and W. Liu, "A review of EEG signal classifier based on deep learning," in *Proceedings of Information Science and Cloud Computing — PoS(ISCC 2017)*, Guangzhou, China, March 2018.
- [26] Z. H. Zhou, *Machine Learning*, Tsinghua University Press, Beijing, 2016.
- [27] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [28] H. Guo, R. TANG, Y. Ye, Z. Li, and X. He, "DeepFM: a factorization-machine based neural network for CTR prediction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia, 2017.
- [29] S. Rendle, "Factorization machines," in *IEEE International Conference on Data Mining IEEE*, Sydney, NSW, Australia, Dec. 2010.
- [30] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, *Application of dimensionality reduction in recommender system-a case study*, Minnesota Univ Minneapolis Dept of Computer Science, 2000.
- [31] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [33] T. Wang, "Multi-focus fusion algorithm for noisy images," *Optics and Precision Engineering*, vol. 19, no. 12, pp. 2977–2984, 2011.
- [34] C. L. Zhen, *Random Noise Suppression of Seismic Signal Based on Wavelet Transform*, Donghua University of technology, 2018.
- [35] J. Morlet, G. Arens, E. Fourgeau, and D. Giard, "Wave propagation and sampling theory-part II:sampling theory and complex waves," *Geophysics*, vol. 47, no. 2, pp. 222–236, 1982.
- [36] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square Integrable wavelets of constant shape," *SIAM Journal on Mathematical Analysis*, vol. 15, no. 4, pp. 723–736, 1984.
- [37] J. Zhang and J. Cao, "Decision tree algorithm for big data analysis," *Computer Science*, vol. 43, no. S1, pp. 374–379, 2016.
- [38] S. H. Zhang, J. X. Liu, and L. F. Kong, "Occlusion detection based on depth image using random forest," *Acta Optica Sinica*, vol. 34, no. 9, pp. 197–208, 2014.
- [39] P. Komarek, *Logistic regression for data mining and high-dimensional classification*, pp. 29–36, 2004.
- [40] S. F. Ding, B. J. Qi, and H. Tan, "A review of support vector machine theory and algorithm," *Journal of University of Electronic Science and Technology*, vol. 40, no. 1, pp. 2–10, 2011.

## Research Article

# Association between Timing of Surgical Intervention and Mortality in 15,813 Acute Pancreatitis

Lan Lan <sup>1</sup>, Jiawei Luo,<sup>1</sup> Xiaoyan Yang,<sup>1</sup> Dujiang Yang,<sup>2</sup> Mengjiao Li,<sup>1</sup> Fangwei Chen,<sup>3</sup> Nianyin Zeng <sup>4</sup>, and Xiaobo Zhou<sup>5</sup>

<sup>1</sup>West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>2</sup>Department of Gastrointestinal Surgery, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>3</sup>West China School of Public Health, Sichuan University, Chengdu 610041, China

<sup>4</sup>Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China

<sup>5</sup>School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston 77030, USA

Correspondence should be addressed to Lan Lan; [lanl@scu.edu.cn](mailto:lanl@scu.edu.cn)

Received 23 February 2020; Revised 6 April 2020; Accepted 30 April 2020; Published 16 May 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Lan Lan 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.

**Objective.** In order to find the quantitative relationship between timing of surgical intervention and risk of death in necrotizing pancreatitis. **Methods.** The generalized additive model was applied to quantitate the relationship between surgical time (from the onset of acute pancreatitis to first surgical intervention) and risk of death adjusted for demographic characteristics, infection, organ failure, and important lab indicators extracted from the Electronic Medical Record of West China Hospital of Sichuan University. **Results.** We analyzed 1,176 inpatients who had pancreatic drainage, pancreatic debridement, or pancreatectomy experience of 15,813 acute pancreatitis retrospectively. It showed that when surgical time was either modelled alone or adjusted for infection or organ failure, an L-shaped relationship between surgical time and risk of death was presented. When surgical time was within 32.60 days, the risk of death was greater than 50%. **Conclusion.** There is an L-shaped relationship between timing of surgical intervention and risk of death in necrotizing pancreatitis.

## 1. Introduction

Indications for surgical intervention of acute pancreatitis (AP) are secondary infections of the pancreas, secondary infections, or compression symptoms, mainly including the pancreas or peripancreatic symptoms or necrosis of secondary infections and organ failure. It is well known that early debridement is associated with higher morbidity and mortality, and recommendations are to delay by at least 4 weeks after the acute pancreatitis episode. Recommendations of guidelines for surgical timing of necrotizing pancreatitis from United States, United Kingdom, Italy, Finland, and Japan are delayed as far as possible, without recommendations for individuals [1–6]. Those recommendations lacking details could result in a large difference in the selection of best surgical timing in practice.

Previous studies [7, 8] on the timing of surgical interventions mostly calculated the time from admission. The time of admission of each patient was susceptible to a variety of factors, such as economic factors and availability of medical resources. Because of the big difference of the time before admission, there is often a certain error based on the time of admission. A more reasonable evaluation should be calculated from the onset of AP (the time of onset of abdominal pain). Additionally, previous studies were mostly qualitative research. A prospective study of 223 patients with well-defined early and late intervention with a subgroup analysis with multiorgan failure and infected necrosis was used [8]. However, they cannot continuously give the risk of death corresponding to a certain point in time. Infection and organ failure have been used as key factors in determining whether or not to undergo surgery and are considered the determi-

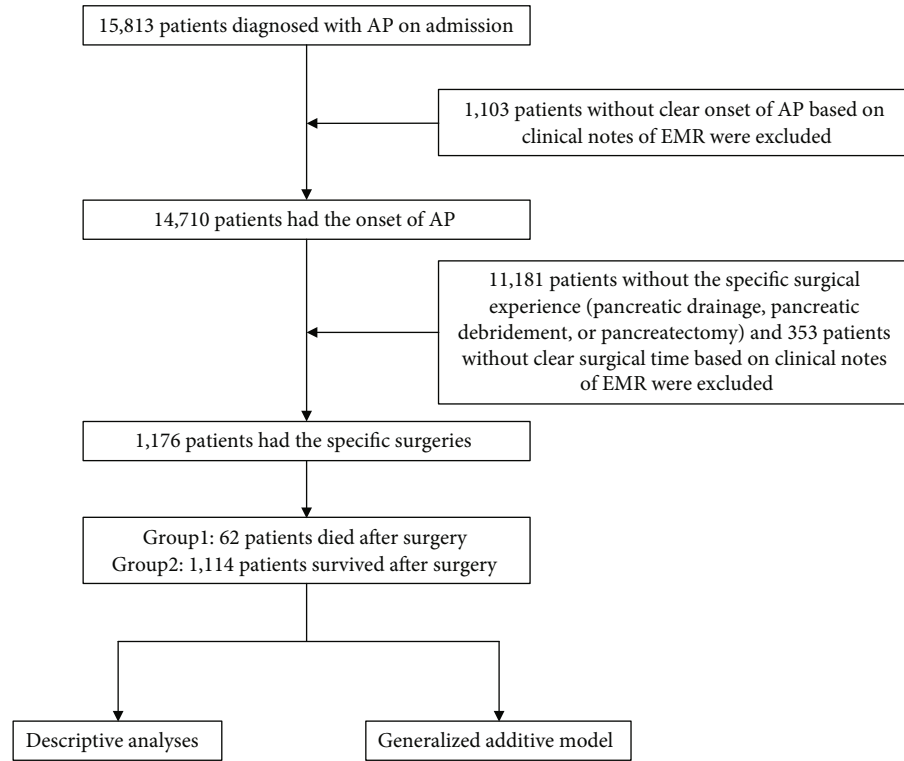


FIGURE 1: Flow diagram of this study.

nants of mortality for the patients with necrotizing pancreatitis [9, 10]. It was observed that organ failure was more likely to determine mortality in AP [11, 12]. While a prospective cohort study from the Netherlands showed that there were no associations between infection, onset of organ failure, duration of organ failure, and mortality in the patients with necrotizing pancreatitis [13]. What is more, pancreatic amylase is one of the criteria for the diagnosis of AP [14]. High-density lipoprotein within 48 hours after admission is a good predictor of the severity of AP [15], so that the effect of severity can be adjusted by early high-density lipoprotein. White blood cell count on admission is a good indicator of infection, and it can be used to adjust the impact of infection on mortality [16]. Creatinine is the diagnostic criteria for renal failure [14]. Collecting this information prospectively is labor intensive, which often results in a small sample size. Therefore, it is critical that this information can be obtained from Electronic Medical Record (EMR) without extra cost to be researched based on a large sample size.

Therefore, we applied a generalized additive model to quantitate the relationship between surgical time (from the onset of AP to first surgical intervention) and risk of death for 15,813 inpatients diagnosed with AP from EMR, as well as adjusting for demographic characteristics, infection, organ failure, and important lab indicators.

## 2. Materials and Methods

**2.1. Study Setting and Population.** The surgical approach for necrotizing pancreatitis can be classified into three categories: drainage, pancreatectomy, and pancreatic necrotic tissue

removal plus extensive drainage [5]. Therefore, we defined study patients as follows: (1) diagnosis with AP on admission based on ICD codes (ICD-9: 577.0, ICD-10: K85) and (2) having at least one surgical intervention experience including pancreatic drainage, pancreatic debridement, or pancreatectomy in a same encounter. At the beginning, 15,813 patients diagnosed with acute pancreatitis were included. After extracting surgical records of the patients, 1,176 patients were included finally (see Figure 1). This study retrospectively collected data of patients with AP and followed the STROBE guidelines [17] for observational studies. The research protocol was approved by the ethics review board of West China Hospital of Sichuan University, and the need for informed consent was waived owing to the retrospective nature of the study.

**2.2. Data Collection and Definitions.** After admission, all patients diagnosed with AP from West China Hospital of Sichuan University initially received traditional treatment. The etiology for patients was main biliary, alcohol abuse, and others. When abdominal pain, severe clinical deterioration, or development of clinical signs of sepsis persisted or recurred, the CECT was performed. Patients with confirmed or suspected infected necrosis were advised to receive surgical intervention based on the CT results. Then, experienced surgeons discussed the case with the radiologist to decide the type and time for surgical intervention, which delayed as much as possible after four weeks from the onset. When patients had persistent clinical manifestations of sepsis, prompt surgical intervention was considered. The data were retrospectively extracted from EMR of West China Hospital



of Sichuan University from 2010 to 2018, including demographic characteristics, lab tests, vital signs, and death information. If it was positive for the bacteria in the pancreas or peripancreatic drain, pus, or secretions, the patient was infected. Respiratory failure was defined as the partial pressure of oxygen in blood gas analysis less than 60 mmHg or the use of a ventilator. Circulatory failure was defined as diastolic blood pressure less than 60 mmHg or systolic blood pressure less than 90 mmHg and the use of vasoactive drugs. Kidney failure was defined as creatinine greater than  $177 \mu\text{mol/L}$ . The time from the onset of AP to admission was asked by physicians. The lab test results were extracted from the laboratory information system, and the clinical events (vital signs information, etc.) were extracted from the nursing system.

**2.3. Statistical Analysis.** We used a regular expression [18] to extract the patients who had specific surgical intervention experience and the onset of AP from the clinical notes of EMR in the patients diagnosed with AP on admission. We explored the difference between died and survived inpatients diagnosed with AP after the specific surgical intervention. The baselines of the two groups were compared, including important lab indicators, infection, and organ failure. *t*-test and Chi-square test were used to evaluate the difference between the two groups.

Considering that the relationship between many clinical factors and risk of death are often not linear and the generalized additive model [18] allows each variable to be put in the model in different nonlinear forms, the generalized additive model was used to explore the association between the timing of surgical intervention and risk of death, controlling the potential confounding factors like infection and organ failure. We assumed that the death of the patients obeys the Bernoulli distribution. The formula of the generalized additive model is as follows:  $g(Y_i) = \alpha + f(x_{1i}) + f(x_{2i}) + \dots$ , where  $Y$  is death or not,  $\alpha$  is the intercept,  $x$  is the independent variable,  $i$  indicates the  $i$ th patient, and  $f$  is the nonlinear function of independent variable.  $f$  is a smooth cubic spline regression function formulated as  $s(\cdot)$  in this study. The backfitting method was used to evaluate the model, and the hyperparameter was selected by the Akaike information criterion (AIC). Based on the adjustment of demographic characteristics and important lab indicators, we first adjusted for infection, secondly adjusted for organ failure, and finally modelled surgical time lonely. When a variable with missing values was to be used, the patient with the missing value was deleted. It is statistically significant if the  $P$  value is less than 0.05. All data analyses were done in the R software.

### 3. Results

**3.1. Baseline Characteristics.** In this study, 1,176 patients with a mean age of  $45.57 \pm 12.72$  years and 780 (66.33%) males who had surgical intervention (pancreatic drainage, pancreatic debridement, or pancreatectomy) in 15, 813 patients diagnosed with AP on admission were analyzed. The number of patients with respiratory failure, circulatory failure, and kidney failure before surgical intervention was 36 (3.06%),

522 (44.39%), and 171 (14.54%), respectively. There were 463 (39.37%) patients infected. The time from the onset of AP to admission and first surgical intervention was  $23.05 \pm 35.42$  days and  $34.43 \pm 34.95$  days, respectively. The total hospital stay was  $31.54 \pm 25.03$  days. Sixty-two (5.27%) patients after surgical intervention died in the hospital.

The baselines between died and survived patients after surgical intervention were compared. There was no difference between the two groups with respect to age and gender. High-density lipoprotein on admission of survived patients was a little higher than that of died patients. Died patients were 2.45 times and 2.44 times than survived patients for amylase on admission and maximum preoperative creatinine, respectively. Their white blood cell count on admission looked similar. The proportion of infection and organ failure in the death group was higher than that in the surviving group except for respiratory failure without statistical difference. The time from the onset of AP to admission and surgical intervention of died patients was shorter than that of survived patients, while total hospital stay was longer without statistical significance (see Table 1).

**3.2. Modelling Surgical Time and Mortality Adjusted for Infection.** Firstly, we modelled surgical time and mortality adjusted for infection, as well as other covariates. The formula is as follows:  $\text{logit}(Y_i) = \alpha + s(x_{1i}, \beta_1) + s(x_{2i}, \beta_2) + \beta_3 x_{3i} + \beta_4 x_{4i} + s(x_{5i}, \beta_5) + s(x_{6i}, \beta_6) + s(x_{7i}, \beta_7)$ , where  $x_1$  is the time from the onset of AP to surgical intervention,  $x_2$  is age,  $x_3$  is the gender,  $x_4$  is infection or not,  $x_5$  is the high-density lipoprotein on admission,  $x_6$  is the amylase on admission, and  $x_7$  is the white blood cell count on admission ( $n = 708$ ,  $R^2 = 18.2\%$ ). Amylase, high-density lipoprotein on admission, and surgical time had statistical association with death adjusted for age, gender, infection, and white blood cell count on admission (see Table 2).

We further analyzed the independent relationships between risk factors and risk of death. Figure 2 shows that there was a roughly L-shaped relationship between the time from the onset of AP to surgical intervention and risk of death, which indicates that premature surgery has a higher risk of death than postponed surgery. The older, the smaller the high-density lipoprotein or the higher the amylase on admission and the higher the risk of death. The risk of death in white blood cell count on admission was first rising and then falling. The shaded area represents the 95% confidence interval.

**3.3. Modelling Surgical Time and Mortality Adjusted for Infection and Organ Failure.** Secondly, we modelled surgical time and mortality adjusted for infection and organ failure, as well as other covariates. The formula is as follows:  $\text{logit}(Y_i) = \alpha + s(x_{1i}, \beta_1) + s(x_{2i}, \beta_2) + \beta_3 x_{3i} + \beta_4 x_{4i} + s(x_{5i}, \beta_5) + s(x_{6i}, \beta_6) + s(x_{7i}, \beta_7) + s(x_{8i}, \beta_8) + \beta_9 x_{9i} + \beta_{10} x_{10i} + \beta_{11} x_{11i}$ , where  $x_1$  is the time from the onset of AP to surgical intervention,  $x_2$  is the age,  $x_3$  is the gender,  $x_4$  is infection or not,  $x_5$  is the high-density lipoprotein on admission,  $x_6$  is the amylase on admission,  $x_7$  is the white blood cell count on admission,  $x_8$  is the maximum preoperative creatinine,  $x_9$  is respiratory failure or not,  $x_{10}$  is

TABLE 1: Baselines between died and survived patients after surgical intervention.

Characteristics	Died ( $n = 62$ )	Survived ( $n = 1,114$ )	$P$
Age (year, mean (SD))	48.21 (13.32)	45.43 (12.67)	0.094
Male, $n$ (%)	39 (62.90)	741 (66.52)	0.654
Lab indicator			
High-density lipoprotein on admission <sup>a</sup> (mmol/L, mean (SD))	0.43 (0.31)	0.60 (0.37)	0.001*
Amylase on admission <sup>a</sup> (U/L, mean (SD))	635.79 (647.36)	259.47 (537.16)	<0.001*
White blood cell count on admission <sup>a</sup> ( $10^9/L$ , mean (SD))	11.92 (5.25)	10.82 (6.51)	0.208
Maximum preoperative creatinine <sup>a</sup> ( $\mu\text{mol/L}$ , mean (SD))	217.34 (190.20)	89.10 (103.13)	<0.001*
Infection, $n$ (%)	37 (59.68)	426 (38.24)	0.001*
Organ failure before surgical intervention			
Respiratory failure, $n$ (%)	2 (3.23)	34 (3.05)	1.000
Circulatory failure, $n$ (%)	49 (79.03)	473 (42.46)	<0.001*
Kidney failure, $n$ (%)	37 (59.68)	134 (12.03)	<0.001*
Time from the onset to admission (day, mean (SD))	11.55 (14.96)	23.69 (36.12)	0.009*
Time from the onset to surgical intervention (day, mean (SD))	23.03 (16.33)	35.07 (35.60)	0.008*
Total hospital stay (day, mean (SD))	32.21 (27.54)	31.50 (24.90)	0.829

SD: standard deviation;  $n$  (%): number and percentage; \* indicates statistical significance; <sup>a</sup>Different missing rates 2.7%, 8.3%, 4.2%, and 2.7% for high-density lipoprotein, amylase, white blood cell count on admission, and maximum preoperative creatinine, respectively.

TABLE 2: Model results between surgical time and mortality adjusted for infection.

Covariates	$\beta$	SD	$Z$ or $\chi^2$	$P$
Intercept	-4.117	0.502	-8.200	<0.001*
$s$ (time from the onset to surgical intervention)	-	-	4.282	0.042*
$s$ (age)	-	-	0.836	0.563
Male	-0.300	0.453	-0.662	0.508
Infection	0.597	0.456	1.308	0.191
$s$ (high-density lipoprotein)	—	—	7.037	0.022*
$s$ (amylase)	-	-	20.197	<0.001*
$s$ (white blood cell count)	-	-	0.952	0.575

‘-’ no traditional slope concept in this study; \* indicates statistical significance.

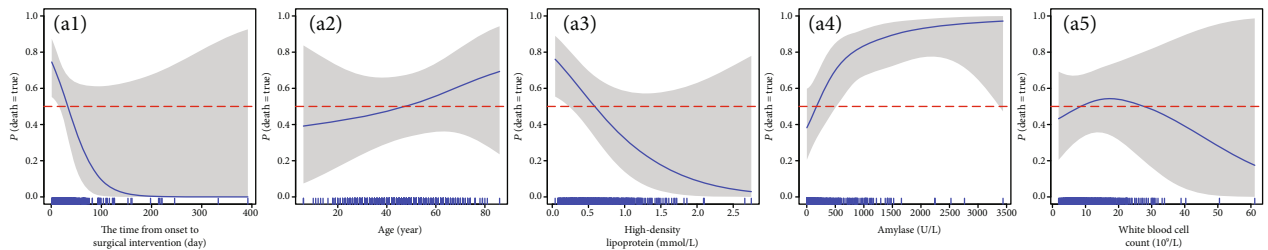


FIGURE 2: The relationship between risk of death and five risk factors.

circulatory failure or not, and  $x_{11}$  is kidney failure or not ( $n = 708$ ,  $R^2 = 31.5\%$ ). Amylase on admission had a statistical association with death adjusted for surgical time, age, gender, infection, organ failure, and other lab indicators (see Table 3).

The independent relationships between risk factors and risk of death were also analyzed. Figure 3 shows that after the inclusion of more variables, the relationship between surgical time, age, high-density lipoprotein, amylase, and white blood cell count and risk of death remained similar.

TABLE 3: Model results between surgical time and mortality adjusted for infection and organ failure.

Covariates	$\beta$	SD	Z or $\chi^2$	P
Intercept	-4.586	0.608	-7.541	<0.001*
s (time from the onset to surgical intervention)	-	-	1.489	0.235
s (age)	-	-	0.919	0.459
Male	-0.895	0.516	-1.734	0.083
Infection	0.531	0.505	1.051	0.293
s (high-density lipoprotein)	-	-	1.739	0.268
s (amylase)	-	-	12.749	0.005*
s (white blood cell count)	-	-	0.697	0.665
s (creatinine)	-	-	2.845	0.408
Respiratory failure	-1.794	1.409	-1.273	0.203
Circulatory failure	0.858	0.498	1.722	0.085
Kidney failure	1.306	0.760	1.718	0.086

'-' no traditional slope concept in this study; \* indicates statistical significance.

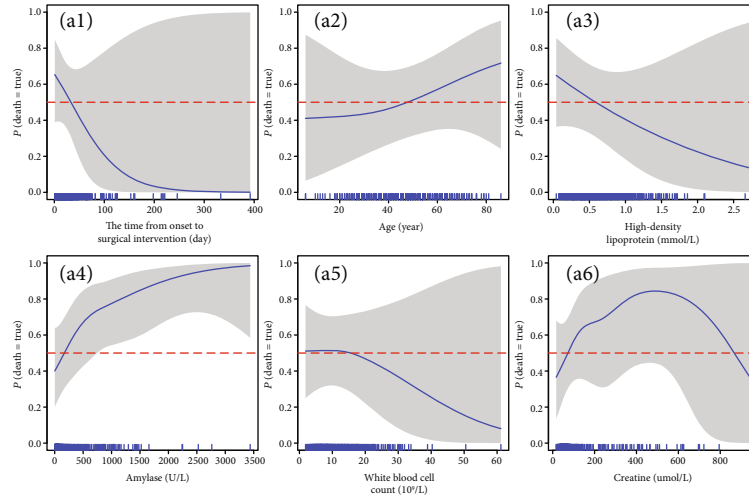


FIGURE 3: The relationship between risk of death and six risk factors.

The risk of death was high in a specific range and low in both ends of creatinine.

**3.4. Modelling Surgical Time and Mortality.** Finally, we developed a model adjusted for age, gender, and high-density lipoprotein formulated as  $\text{logit}(Y_i) = \alpha + s(x_{1i}, \beta_1) + s(x_{2i}, \beta_2) + \beta_3 x_{3i} + s(x_{4i}, \beta_4)$ , where  $x_1$  is the time from the onset of AP to surgical intervention,  $x_2$  is the age,  $x_3$  is the gender, and  $x_4$  is the high-density lipoprotein on admission, to find the relationship between surgical time and risk of death. Age and gender, as well as high-density lipoprotein, were used to adjust for basic characteristics and severity of AP, respectively. Based on the premise, the relationship between surgical time and risk of death in the infected and noninfected groups was also studied. In this section, we applied generalized additive model based on different samples: all patients ( $n = 1,144$ ,  $R^2 = 7.92\%$ ), infected patients ( $n = 463$ ,  $R^2 = 5.96\%$ ), and noninfected patients ( $n = 681$ ,  $R^2 = 13.40\%$ ). There was a

statistical correlation between surgical time and mortality in the three groups.

Figure 4 shows that the relationship between surgical time and death was similar among all, infected, and noninfected patients. Because the risk of death was very low after 100 days of surgical time, we only figured out the surgical time within 100 days. The relationship between surgical time and death was the same in the infected and noninfected patient groups. Surgical time 32.60, 32.84, and 36.55 days in all, infected, and noninfected patients, respectively, had 50% risk of death. The risk of death would be more than 50% if the surgical time was less than the thresholds.

#### 4. Discussion

This study investigated the relationship between surgical timing and death in necrotizing pancreatitis based on a large sample of EMR. The inpatients of AP with the specific surgery (pancreatic drainage, pancreatic debridement, or

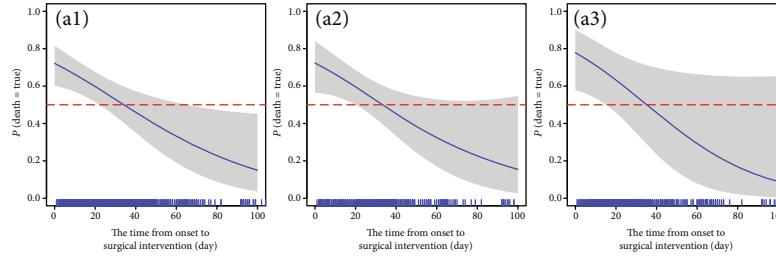


FIGURE 4: The relationship between risk of death and surgical time ((a1) is for all patients, (a2) is for infected patients, and (a3) is for noninfected patients).

pancreatectomy) were modelled, who were almost patients with necrotizing pancreatitis. According to our best knowledge, there is no quantitative study between the timing of surgical intervention (from the onset of AP to first surgical intervention) and risk of death in necrotizing pancreatitis. This study is the first case. There is an L-shaped relationship between surgical time and risk of death in necrotizing pancreatitis, showing that premature surgery carries a higher risk of death among patients with necrotizing pancreatitis. This kind of relationship is still robust after sensitivity analyses.

In descriptive analyses, the time from the onset of AP to surgical intervention, time from the onset of AP to admission, high-density lipoprotein on admission, amylase on admission, maximum preoperative creatinine, infection, circulatory failure, and kidney failure had a statistical difference with respect to death. These variables were further put in the model at the same time to check if there was a real impact on death. In the first model incorporating infection and other covariates, results showed that the lower high-density lipoprotein on admission, the higher the risk of death, which is consistent with previous study [15]. And for the second model inclusion of infection and organ failure as well as other covariates, the relationship between the two was similar, but it was not statistically significant. Amylase was statistically significant in the inclusion of infection or organ failure. The higher the amylase, the higher the risk of death, and risk of death exceeded 50% when amylase was over 175.54 mmol/L. The risk of death for white blood cell count was first rising slowly and then decreasing quickly. One of the most possible reasons is that the doctor will give an antibiotic treatment to control the white blood cell count in a normal range and reduce the probability of infection when the white blood cell count exceeds  $10 \times 10^9/L$ . Therefore, the risk of death would decline when the white blood cell count exceeds  $10 \times 10^9/L$ . In both models, infection was not statistically significant, which has similar results with Guo et al. [11]. Our proposed model can deal with a collinear independent variable. Respiratory failure, circulatory failure, kidney failure, and creatinine were not statistically significant after including in the second model, consistent with the findings of the Dutch Pancreatitis Study Group [13]. However, we found that risk of death was low when creatinine was too low or too high, and risk of death was higher than 50% with creatinine ranging from 73.55 to 818.06  $\mu\text{mol/L}$ . For the relationship between age and death, although there was no statistical difference in the first and second models, it presented increased risk of death with increase in age.

Some covariates may not have statistical differences, but as can be seen from previous figures, these variables have a regularity with risk of death, and our model gave a threshold of 50% risk of death, which is worthy of attention of surgeons. After adjusting for infection, of surgical time and death that was still statistically significant. But after adjusting for organ failure, there was no statistical significance. No statistical difference does not mean that there is no real association between the two. Statistical difference is related to many factors such as the choice of independent variables and sample size. Therefore, in order to find out the relationship between surgical time and risk of death, we finally modelled surgical time adjusted for age, gender and high-density lipoprotein on admission since demographic factors can also be utilized as predictors of inpatients mortality in AP [19]. It was found that when surgical time was either modelled alone or adjusted for infection or organ failure, an L-shaped relationship was presented. Surgical time was within 32.60 days, the risk of death was greater than 50%. Not only that, but this study also obtained the mortality risk corresponding to the timing of surgical intervention at each time point. Although the relationship between surgical time and death was similar in the infected and noninfected groups, surgical time of the infected group (32.84 days) was earlier than that of the noninfected group (36.55 days) at 50% risk of death, and risk of death from early surgery for the noninfected group was 77%, which was a little higher than that (72%) of the infected group.

Although amylase is one of the criteria for the diagnosis of acute pancreatitis, the relationship between amylase and the severity of acute pancreatitis is rarely reported. However, there are many reasons for the patients who have abnormal levels of amylase in their blood, including sudden inflammation of the pancreas, long-term inflammation of the pancreas, fluid-filled sac around the pancreas, pancreatic cancer, inflammation of the gallbladder, and kidney problems. The results of this study showed that the higher the amylase, the higher the risk of death. The reason for this result may be that there are other diseases that also cause high amylase, except acute pancreatitis. Under the combined effect of various diseases, the risk of death is increased. If considering the effects separately, the quantitative relationships between different surgical time and other covariates at different levels and risk of death can be a good reference for surgeons. The results of this work are based on EMR. Other hospitals can use this research strategy to obtain preliminary results and then conduct prospective

design. Therefore, this study provides an important prerequisite for a prospective study.

However, there are still some limitations in this study. The data is retrospectively extracted from EMR, and the performance of the model is strongly correlated with the quality of the data. The death cases were recorded during hospitalizations, and the cause of death was not available based on EMR. Due to the Chinese cultural characteristics, some patients who do not want to die in the hospital will be discharged early and those deaths will not be recorded in the EMR. Therefore, in this study, mortality was underestimated, and its relationship with surgical time was also underestimated. On the other hand, since the generalized additive model cannot analyze the interaction between variables, there may be interactions between variables. It is the limitation of the model itself.

## 5. Conclusions

In conclusion, by applying the generalized additive model, we obtained the relationship between surgical time (from the onset of AP to first surgical intervention) and risk of death in the case of controlling demographic characteristics, infection, organ failure, and important lab indicators in necrotizing pancreatitis. There is an L-shaped relationship between timing of surgical intervention and risk of death in necrotizing pancreatitis, providing an important reference for surgeons when making surgical decisions.

## Data Availability

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

## Conflicts of Interest

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

## Authors' Contributions

Lan Lan and Jiawei Luo contributed equally.

## Acknowledgments

This work was supported by the Sichuan Science and Technology Program (grant 2019YFS0147 and grant 2016SZ0072), 1-3-5 project for disciplines of excellence, West China Hospital, Sichuan University (grant ZYJC18010), 1-3-5 project for disciplines of excellence—Clinical Research Incubation Project, West China Hospital, Sichuan University (grant 2019HXXFH022), and Post-Doctor Research Project, West China Hospital, Sichuan University (grant 2019HXBH039).

## References

- [1] W. G. I. A. Guidelines, "IAP/APA evidence-based guidelines for the management of acute pancreatitis," *Pancreatology*, vol. 13, no. 4, pp. e1–e15, 2013.
- [2] S. Tenner, J. Baillie, J. DeWitt, S. S. Vege, and American College of Gastroenterology, "American College of Gastroenterology Guideline: Management of Acute Pancreatitis," *American Journal of Gastroenterology*, vol. 108, no. 9, pp. 1400–1415, 2013.
- [3] UK Working Party on Acute Pancreatitis, "UK guidelines for the management of acute pancreatitis," *Gut*, vol. 54, Supplement 3, pp. iii1–iii9, 2005.
- [4] R. Pezzilli, A. Zerbi, D. Campa et al., "Consensus guidelines on severe acute pancreatitis," *Digestive and Liver Disease*, vol. 47, no. 7, pp. 532–543, 2015.
- [5] A. Leppäniemi, M. Tolonen, A. Tarasconi et al., "2019 WSES guidelines for the management of severe acute pancreatitis," *World Journal of Emergency Surgery*, vol. 14, no. 1, p. 27, 2019.
- [6] M. Yokoe, T. Takada, T. Mayumi et al., "Japanese guidelines for the management of acute pancreatitis: Japanese Guidelines 2015," *Journal of Hepato-Biliary-Pancreatic Sciences*, vol. 22, no. 6, pp. 405–432, 2015.
- [7] M. G. Besselink, T. J. Verwer, E. J. Schoenmaeckers et al., "Timing of Surgical Intervention in Necrotizing Pancreatitis," *Archives of Surgery*, vol. 142, no. 12, pp. 1194–1201, 2007.
- [8] Q. Guo, A. Li, Q. Xia et al., "Timing of Intervention in Necrotizing Pancreatitis," *Journal of Gastrointestinal Surgery*, vol. 18, no. 10, pp. 1770–1776, 2014.
- [9] J. van Grinsven, for the Dutch Pancreatitis Study Group, H. C. van Santvoort et al., "Timing of catheter drainage in infected necrotizing pancreatitis," *Nature Reviews Gastroenterology & Hepatology*, vol. 13, p. 306, 2016.
- [10] I. P. Gomas, C. M. Halloran, P. Ghaneh et al., "Outcomes From Minimal Access Retroperitoneal and Open Pancreatic Necrosectomy in 394 Patients With Necrotizing Pancreatitis," *Annals of Surgery*, vol. 263, no. 5, pp. 992–1001, 2016.
- [11] Q. Guo, A. Li, Q. Xia et al., "The role of failure and infection in Pancreatitis," *Annals of Surgery*, vol. 259, no. 6, pp. 1201–1207, 2014.
- [12] N. Shi, T. Liu, D. de la Iglesia-Garcia et al., "Duration of organ failure impacts mortality in acute pancreatitis," *Gut*, vol. 69, no. 3, pp. 604–605, 2020.
- [13] N. J. Schepers, O. J. Bakker, M. G. Besselink et al., "Impact of characteristics of organ failure and infected necrosis on mortality in necrotizing pancreatitis," *Gut*, vol. 68, no. 6, pp. 1044–1051, 2019.
- [14] P. A. Banks, T. L. Bollen, C. Dervenis et al., "Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus," *Gut*, vol. 62, no. 1, pp. 102–111, 2012.
- [15] C.-L. Zhou, C.-H. Zhang, X.-Y. Zhao et al., "Early prediction of persistent organ failure by serum apolipoprotein A-I and high-density lipoprotein cholesterol in patients with acute pancreatitis," *Clinica Chimica Acta*, vol. 476, pp. 139–145, 2018.
- [16] E. L. George and A. Panos, "Does a high WBC count signal infection?," *Nursing*, vol. 35, no. 1, 2005.
- [17] E. von Elm, D. G. Altman, M. Egger et al., "The Strengthening of Reporting of Observational Studies in Epidemiology



(STROBE) Statement: Guidelines for reporting observational studies,” *International Journal of Surgery*, vol. 12, no. 12, pp. 1495–1499, 2014.

- [18] S. N. Wood, “Fast stable direct fitting and smoothness selection for generalized additive models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 70, no. 3, pp. 495–518, 2008.
- [19] D. Mossad, M. Mustafa, R. J. Markert, and S. Agrawal, “Mo1424 Predictors of In-Hospital Mortality in Acute Pancreatitis,” *Gastroenterology*, vol. 150, no. 4, Supplement 1, pp. S708–S709, 2016.

## Research Article

# Fast Enhanced Exemplar-Based Clustering for Incomplete EEG Signals

Anqi Bi, Wenhao Ying, and Lu Zhao 

*School of Computer Science and Engineering, Changshu Institute of Technology, Changshu, Jiangsu, China*

Correspondence should be addressed to Lu Zhao; [constance@csit.edu.cn](mailto:constance@csit.edu.cn)

Received 16 January 2020; Accepted 27 February 2020; Published 8 May 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Anqi Bi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The diagnosis and treatment of epilepsy is a significant direction for both machine learning and brain science. This paper newly proposes a fast enhanced exemplar-based clustering (FEEC) method for incomplete EEG signal. The algorithm first compresses the potential exemplar list and reduces the pairwise similarity matrix. By processing the most complete data in the first stage, FEEC then extends the few incomplete data into the exemplar list. A new compressed similarity matrix will be constructed and the scale of this matrix is greatly reduced. Finally, FEEC optimizes the new target function by the enhanced  $\alpha$ -expansion move method. On the other hand, due to the pairwise relationship, FEEC also improves the generalization of this algorithm. In contrast to other exemplar-based models, the performance of the proposed clustering algorithm is comprehensively verified by the experiments on two datasets.

## 1. Introduction

Epilepsy is a common disease of nervous system, which is characterized by sudden brain dysfunction. Although there are many other neuroimaging modalities for the recognition of brain activity, EEG signals have a high temporal resolution which is up to the millisecond level, and its acquisition equipment is inexpensive, portable, and noninvasive. Nowadays, most diagnoses of epilepsy are based on clinical experience and the analysis of electroencephalogram (EEG) signals. Compared with manual diagnostic method, machine learning methods are less time-consuming and more consistent [1–6]. Specifically, many machine learning methods such as support vector learning [7, 8], Takagi–Sugeno–Kang (TSK) fuzzy system [9, 10], and Naïve Bayes [11] have been applied.

As we know that brain activity is a nonlinear, unstable complex network system, EEG signals we usually get are complicated. That is to say, some EEG signals are complete while others may miss some features, namely, incomplete. Therefore, recognition of epilepsy based on machine learning models will be more promising compared with clinical diagnosis depending on experience. Moreover, EEG

signals have the characteristics of high dimension and stochasticity which limit the performance of most existing clustering models, such as k-means [11] and fuzzy  $c$  mean (fcm) [12]. K-means and fcm clustering models need to preset the number of clusters in advance. More specifically, the performance of the k-means model relies on the initialization of data, while the fcm model requires high interpretability. Thus, we focus on the exemplar-based clustering model [13] which is proposed by Frey in this paper. The exemplar-based clustering model has the advantages of automatically obtaining the cluster number, high efficiency, and not relying on the initialization of data.

In conclusion, we consider the scenario of EEG signals consisting of most complete data and few incomplete data in this paper, as shown in Figure 1. Based on the previous work about the recognition of epileptic signals, we propose a novel fast enhanced exemplar-based clustering (FEEC) model for incomplete EEG signals. As shown in Figure 1, different from existing exemplar-based clustering models, FEEC compresses the exemplar list and reduces the pairwise similarity matrix, and then FEEC optimizes the target model by the enhanced  $\alpha$ -expansion move framework. Moreover, the contributions of this paper can be highlighted as follows:

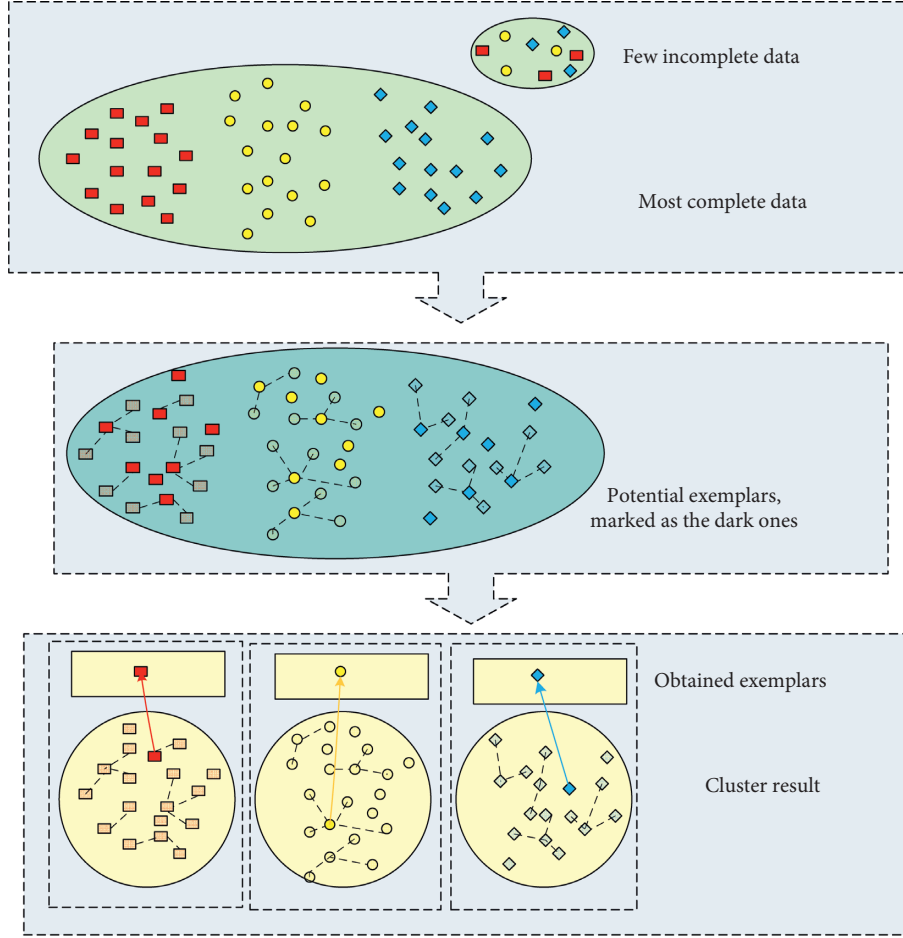


FIGURE 1: Clustering procedure of FEEC algorithm.

- (1) We extend the existing exemplar-based clustering algorithm into a fast version by compressing the potential exemplar lists in this study. FEEC compresses the number of potential exemplars by processing the most complete data in the first stage and extends the few incomplete data into the exemplar list. So, the complexity of FEEC is reduced as well.
- (2) Along with most existing exemplar-based clustering models, FEEC is built on the pairwise similarity matrix of data. Thus, after compression, FEEC would construct a new reduced similarity matrix, and the generalization of this algorithm is improved.
- (3) Moreover, this paper also considers the fact that the graph cuts [14] based optimization performs better than those loopy belief propagation (LBP) [15] based structure. So, the proposed FEEC algorithm optimizes the target model by the enhanced  $\alpha$ -expansion move framework [16, 17].
- (4) Experimental results of both synthetic and real-world datasets indicate the promising efficiency of the proposed FEEC algorithm.

The rest of this paper is listed as follows. In Section 2, we introduce some static exemplar-based clustering models. Section 3 discusses the proposed FEEC algorithm step by

step. In Section 4, we analyze the experimental results and the comparison of FEEC and other existing methods. Section 5 concludes this whole paper.

## 2. Background

Since EEG signal feature extraction methods and exemplar-based clustering models are two important supporting theories for the FEEC model in this study, we will briefly introduce several feature extraction methods and exemplar-based clustering models in this section.

**2.1. Feature Extraction Methods.** Original EEG signals have the characteristics of high dimensionality, stochasticity, and nonlinearity. It would be computationally very expensive to extract features from raw EEG signals; nowadays, many feature extraction methods have been proposed to handle this problem. In sum, there are three categories, i.e., time-domain features, frequency-domain features, and time-frequency features.

More specifically, in time-domain analysis, statistics component features of the raw EEG signals will be analyzed [18]. In frequency-domain analysis, power spectrum analysis and short-time Fourier transform (STFT) [19, 20] are commonly used. In time-frequency analysis, time and

frequency domain are simultaneously extracted from non-stationary EEG signals. Wavelet and other improved versions [21, 22] are widely used in EEG signal processing. We utilize KPCA to extract feature in this paper.

**2.2. Exemplar-Based Clustering Models.** Exemplar-based clustering models select cluster centers, namely, exemplars, from existing actual data. We focus on exemplar-based clustering models in this paper and briefly introduce affinity propagation (AP) [13] and enhanced  $\alpha$ -expansion move (EEM) [17] in this section. And several extended versions for different scenarios are shown in Table 1. The target function defined by exemplar-based clustering model equals to the minimization problem of energy function of Markov random field(MRF). Two existing optimization strategies have been utilized and evolved into AP and EEM frameworks accordingly. Moreover, loopy belief propagation (LBP) [23] is used in AP, while graph cuts technique [15] is used in EEM, respectively.

**2.2.1. Affinity Propagation.** AP is based on message passing among data points, and its target function is defined as follows:

$$\max_E \sum_{p=1}^N \mathbf{S}(i, k) - \sum_{p=1}^N \delta_p(E), \quad (1)$$

where

$$\delta_p(E) = \begin{cases} \infty, & \text{if } E(x_p) \neq p, \text{ but } \exists x_q: E(x_q) = p, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where  $\mathbf{X} = \{x_1, x_2, \dots, x_N\} \in \mathbb{R}^{N \times D}$  is an input dataset and  $N$  is the total number of  $D$ -dimensional data points.  $\mathbf{SE}$  is the output of this framework, and the element  $E(x_p)$  is referred to the exemplar for each  $x_p$ .

According to AP, each point receives availability message  $\mathbf{A}(i, k)$  and sends responsibility  $\mathbf{R}(i, k)$  message simultaneously, which are defined as follows:

$$\mathbf{R}(i, k) = \mathbf{S}(i, k) - \max_{j, s.t. j \neq k} \{\mathbf{A}(i, j) + \mathbf{S}(i, j)\}, \quad (3)$$

$$\mathbf{A}(i, k) = \begin{cases} \min\{0, \mathbf{R}(k, k) + \sum_{j, s.t. j \neq \{i, k\}} \{\max(0, \mathbf{R}(j, k))\}\}, & i \neq k, \\ \sum_{j, s.t. j \neq \{i, k\}} \{\max(0, \mathbf{R}(j, k))\}, & i = k, \end{cases} \quad (4)$$

where  $\mathbf{S}$  is the similarity matrix of data points and is defined as  $\mathbf{S}(i, j) = -\|\mathbf{x}_i - \mathbf{x}_j\|^2$ . Meanwhile,  $\mathbf{S}(k, k) = p$  where  $p$  is named as preferences in this framework. Moreover, its value should be independent and can be set to a constant.

AP does not require presetting the number of the cluster and the performance is stable. Considering these advantages, many extended versions of AP have been proposed [24, 25]. Specifically, AP defines fading factor to adjust the iteration speed; adAP [24] is proposed to determine this fading factor adaptively. Moreover, several extended versions of AP methods have been proposed to deal with large data and link constraints. For instance, IAPKM, IAPNA, and IAPC [26, 27] employ incremental strategy and semisupervised AP and SSAP [28] concentrate on instance-level constraints. A two-stage fast version of AP (FAP) [29] is also proposed to improve the efficiency. However, although AP has been obtaining its success

in various applications, when we attempt to directly apply AP to incomplete EEG signals, the performance is unsatisfactory.

**2.2.2. Enhanced  $\alpha$ -Expansion Move.** In 2014, Zheng and Chen [17] utilized enhanced  $\alpha$ -expansion move framework to optimize the object function of exemplar-based clustering models and accordingly proposed the EEM clustering model. In line with the mathematical symbols above, the target function of EEM is defined as follows:

$$\max_E \sum_{p=1}^N s(x_p, x_{E(x_p)}) - \sum_{p=1}^N \sum_{q > p}^N \theta_{p,q}(E(x_p), E(x_q)), \quad (5)$$

where

$$\theta_{p,q}(E(x_p), E(x_q)) = \begin{cases} M, & E(x_p) = q, E(x_q) \neq q, \text{ or } E(x_q) = p, E(x_p) \neq p, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

$$\theta_{p,q}(E(x_p), E(x_q)) + \theta_{p,q}(\alpha, \alpha) \leq \theta_{p,q}(E(x_p), \alpha) + \theta_{p,q}(\alpha, E(x_q)). \quad (7)$$

In terms of [17],  $\alpha$ -expansion move algorithm has been proved to be effective in the optimization of the target

function equation (5). Specifically, when  $\forall p, q, E(x_p), E(x_q), \alpha \in \{1, 2, \dots, N\}$ , equation (7) is verified.

TABLE 1: Descriptions of several state-of-the-art exemplar-based algorithms.

	Optimization	Extended version	Descriptions
AP	Message-passing	adAP	Determine fading factor adaptively
		IAPKM, IAPNA, IAPC	Incremental version for data stream
		SAP, SSAP	Deal with instance-level constraints
		FAP	Two-stage fast version
EEM	Graph cuts	DSC	Dynamic data stream cluster
		IEEM	Deal with link constraints
		FEEC (newly proposed)	Two-stage fast version

Furthermore, according to graph theory, in the fast  $\alpha$ -expansion move algorithm, the expansion range is limited in a one exemplar. To break this limit, the EEM model enlarges the range to the whole exemplar set  $E$  when optimizing and defines a second exemplar  $S(i)$  for each point  $\mathbf{x}_i$  as follows:

$$S(i) = \operatorname{argmax}_{s \in (E/l)} d(\mathbf{x}_i, \mathbf{x}_s), \quad \forall \mathbf{x}_i \in \mathbf{X}_l, \quad (8)$$

where  $\mathbf{X}_l = \{\mathbf{x}_i | E(\mathbf{x}_i) = l\}$  is the dataset among which the exemplar is  $l$  and  $s \in (E/l)$  represents other exemplars in  $E$  except for  $l$ .

The EEM clustering model is a state-of-the-art achievement of exemplar-based clustering model and has been proved to be efficient and effective for numerous scenarios [16, 17, 30]. IEEM [30] is proposed to deal with link constraints by embedding a bound term in the target function. For dynamic data stream, Bi and Wang [16] also proposed an incremental EEM version DSC which processes data chunk by chunk. However, for incomplete EEG signals, these methods would not recognize epilepsy well.

### 3. Fast Enhanced Exemplar-Based Clustering Model

In this section, the proposed FEEC model will be stated and theoretically analyzed in detail. We first compress the exemplar list and reduce the pairwise similarity matrix, and then the target model is optimized by the enhanced  $\alpha$ -expansion move framework.

**3.1. Framework.** As mentioned in the introduction section, we focus on the incomplete EEG signals which consist of most complete data and few incomplete data. To improve the efficiency of the EEM clustering model for these incomplete EEG signals, the proposed FEEC framework includes two stages, namely, compression stage and optimization stage. As shown in Figure 2, the compression stage compresses the potential exemplar list and the optimization stage determines the optimal exemplars from the potential exemplar list. Accordingly, the target function can be defined as follows:

$$\max_E \sum_{i=1}^N s(\mathbf{x}_i, \mathbf{x}_{E(\mathbf{x}_i)}) - \sum_{i=1}^N \eta_i(E), \quad (9)$$

where  $\mathbf{X} = [\mathbf{X}_c, \mathbf{X}_l]$  is the input dataset consisting of most complete data  $\mathbf{X}_c = \{\mathbf{x}_{c,1}, \mathbf{x}_{c,2}, \dots, \mathbf{x}_{c,N_c}\}$  and few incomplete data  $\mathbf{X}_l = \{\mathbf{x}_{l,1}, \mathbf{x}_{l,2}, \dots, \mathbf{x}_{l,N_l}\}$ . The total number of data is

defined as  $N = N_c + N_l$ , where  $N_c$  and  $N_l$  are the number of complete and incomplete data, respectively. Remember that we only consider the scenario that  $N_c \gg N_l$  in this study. The second term in equation (9) guarantees the validity of the exemplar list; its definition is similar to that of  $\delta$  in equation (2). In the end,  $E = \{E(\mathbf{x}_1), E(\mathbf{x}_2), \dots, E(\mathbf{x}_N)\}$  represents the exemplar set in question.

In the compression stage, the number of potential exemplar list will be reduced by exemplar-based selection algorithm, namely, EEM method in this study. To be specific, we apply the EEM model on the most complete data to obtain the potential exemplars for these data. FEEC also pulls the few incomplete data into this potential exemplar list and then constructs compressed similarity matrix. Therefore, after compression, only the pairwise similarities between data and potential exemplars would be preserved. Considering that the FEEC method is built on the pairwise similarity matrix, the following clustering procedure would be applied on this compressed similarity matrix. Furthermore, the scale of similarity matrix is reduced from  $N^2$  to  $N_c$ , where  $N$  and  $c$  are the number of data and potential exemplars, respectively.

In the optimization stage, only the similarity relationship between data and potential exemplars is considered. The new target function after compression is similar to that of other exemplar-based clustering model, like equations (1) and (5), so we take graph cuts and LBP into account. Nevertheless, graph cuts based optimization framework outperforms LBP structure [31]. So, the proposed FEEC utilizes the  $\alpha$ -expansion move method to optimize the new target function. Moreover, along with EEM, FEEC also expands the expansion move space from a single data to the second optimal exemplar.

**3.2. Compression Stage.** In the compression stage, the target function of complete data can be defined as follows:

$$\min_{E_c} \sum_{i=1}^{N_c} d(\mathbf{x}_{c,i}, \mathbf{x}_{E_c(\mathbf{x}_{c,i})}) + \sum_{i=1}^{N_c} \sum_{j>i}^{N_c} \theta_{i,j}(E_c(\mathbf{x}_{c,i}), E_c(\mathbf{x}_{c,j})), \quad (10)$$

where  $\mathbf{X}_c = \{\mathbf{x}_{c,1}, \mathbf{x}_{c,2}, \dots, \mathbf{x}_{c,N_c}\} \in \mathbb{R}^{N_c \times D}$  is the complete  $D$ -dimensional data and  $N_c$  is the number of these data.  $E_c$  is the potential exemplar list for most complete data, and the element among  $E(\mathbf{x}_{c,i})$  is referred to the potential exemplar for each  $\mathbf{x}_{c,i}$ . The optimization framework of other exemplar-based clustering models, like EEM, can be utilized to solve



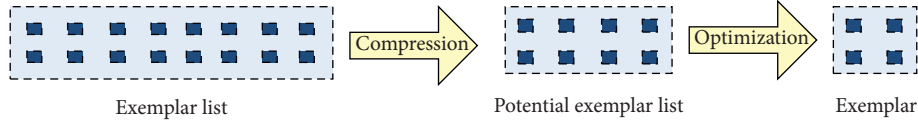


FIGURE 2: Framework of FEEC algorithm: compression stage and optimization stage.

equation (10). In this paper, we select the graph cuts algorithm instead of message-passing algorithm to compress the potential exemplar list. Thus, the potential exemplar list for complete data  $E_c$  can be determined, and the number of potential exemplars is defined as  $c_c$ .

The potential exemplar list after compression stage would be

$$E_{\text{new}} = [E_c, E_l], \quad (11)$$

where  $E_l$  is the exemplar set for the few incomplete data, which is the incomplete data itself actually. That is to say,  $E_l(\mathbf{x}_{l,i}) = i$ .

In this stage, we reduced the number of potential exemplars from  $N_c$  to  $c_c$ . In terms of the analysis in [13, 17, 30], the time complexity of this stage will be  $O(N_c^2)$ . Compared with the time complexity  $O(N^2)$ , if we apply exemplar-based clustering model directly considering the fact that  $N < N_c$ , the time complexity of this compression algorithm would be acceptable.

Therefore, on the basis of the new exemplar list after compression, we can construct the new similarity matrix  $\mathbf{S}_{\text{new}} \in \mathbb{R}^{N \times c}$ ; the element relates to the distance, namely,  $S_{\text{new}}(i, j) = -\|\mathbf{x}_i - \mathbf{x}_{E_{\text{new}}(j)}\|^2$ . The scale of the similarity matrix reduces from  $N^2$  to  $Nc$ , where  $c = c_c + N_l$  represents the number of potential exemplars.

**3.3. Optimization Stage.** After compression, we define the new target function as follows:

$$\max_E \sum_{i=1}^N \sum_{j=1}^c S_{\text{new}}(i, j) - \sum_{i=1}^N \eta_i(E), \quad (12)$$

where  $\mathbf{S}_{\text{new}}$  is the new similarity matrix constructed after compression.

In this section, we construct an optimization framework for equation (12). The second term of equation (12) is set to guarantee the validity of the exemplar list; in order to utilize the graph cuts based method, this term should be pairwise [17]. So,  $\eta_i(E)$  is modified as  $\eta_{i,j}(E)$ . Furthermore, similar to equation (5), we define  $\eta_{i,j}(E)$  as follows:

$$\eta_{i,j}(E) = \begin{cases} M, & E(\mathbf{x}_i) = j, E(\mathbf{x}_j) \neq j, \text{ or } E(\mathbf{x}_j) = i, \quad E(\mathbf{x}_i) \neq i, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

It has been proved that with the definition of  $\eta_{i,j}(E)$ , equation (12) can be optimized by the enhanced  $\alpha$ -expansion method [30]. To improve the efficiency of framework, this method enlarges the expansion move to the second optimal exemplar.

Before optimization, we explain several symbols involved. First, we define  $\mathbf{X}_e$  as those data with exemplar  $\mathbf{x}_e$  and  $\mathbf{x}_\alpha$  as the current potential exemplar. Then, the enhanced  $\alpha$ -expansion move method considers the second optimal exemplar, which is defined as

$$S(\mathbf{x}_i) = \operatorname{argmax}_{s \in (E/\alpha)} S_{\text{new}}(\mathbf{x}_i, \mathbf{x}_s), \quad \forall \mathbf{x}_i \in \mathbf{X}_\alpha, \quad (14)$$

where  $(E/\alpha)$  is the potential exemplar list except for  $\alpha$ .

Apparently, this optimization method should consider two cases, namely,  $\mathbf{x}_e$  is among exemplar list or not, as shown in Figures 3 and 4. To be specific, Figure 3 illustrates the case when  $\mathbf{x}_\alpha$  is an exemplar, while Figure 4 shows the case when  $\mathbf{x}_\alpha$  is not an exemplar. Remember that only when  $\mathbf{x}_\alpha$  is a potential exemplar,  $S(\mathbf{x}_i)$  works. We utilize the concepts of “energy reduction” because this method was first used to optimize the Markov random field (MRF) energy function.

In the situation shown in Figure 3, either  $\forall \mathbf{x}_i \in \mathbf{X}_\alpha$  changes its exemplar to  $S(\mathbf{x}_i)$  or nothing is changed. Therefore, the energy reduction  $R1$  would be defined as

$$R1 = \max(0, R1_\alpha), \quad (15)$$

where  $R1_\alpha$  is the energy reduction when  $\forall \mathbf{x}_i \in \mathbf{X}_\alpha$  changes its exemplar to  $S(\mathbf{x}_i)$  and is defined as

$$R1_\alpha = \sum_{\mathbf{x}_i \in \mathbf{X}_\alpha} (S_{\text{new}}(\mathbf{x}_i, S(\mathbf{x}_i)) - S_{\text{new}}(\mathbf{x}_i, \mathbf{x}_\alpha)). \quad (16)$$

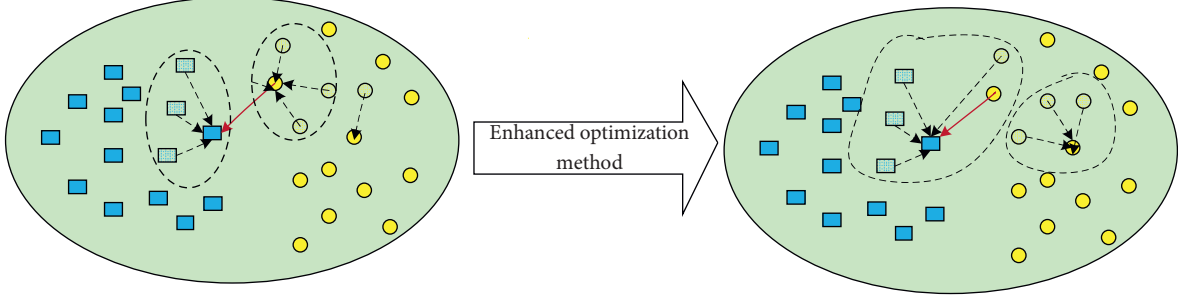
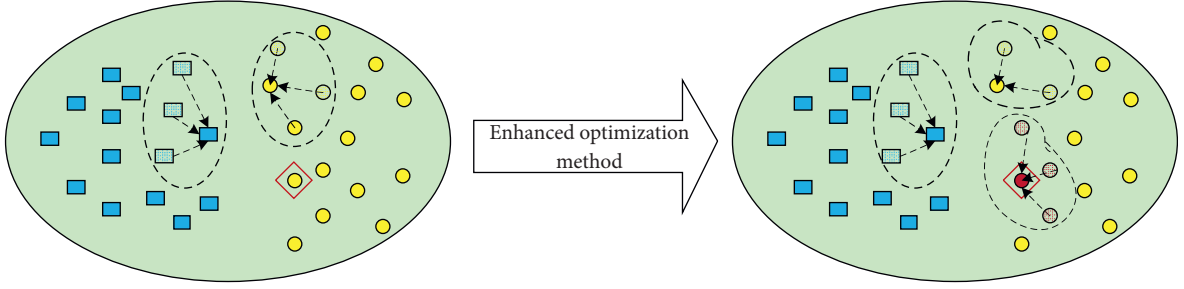
On the other hand, as shown in Figure 4, a new exemplar  $\mathbf{x}_\alpha$  should be considered. Whether to accept the new exemplar is decided by the energy reduction  $R2$ , which will be discussed next. First, we assume a new exemplar  $\mathbf{x}_\alpha$  is accepted. In fact, the following procedure is similar to that shown in Figure 3. Specially, the remaining data would change its exemplar to either  $\mathbf{x}_\alpha$  or  $S(\mathbf{x}_i)$ . For data in cluster  $e \in E$ , theoretical analysis proves that only when the exemplar  $\mathbf{x}_e$  changes its exemplar,  $\forall \mathbf{x}_i \in \mathbf{X}_e$  would change its exemplar as  $S(\mathbf{x}_i)$ . In this case, the energy reduction is defined as follows:

$$R2_e = \sum_{\mathbf{x}_i \in \mathbf{X}_e} (S_{\text{new}}(\mathbf{x}_i, S(\mathbf{x}_i)) - S_{\text{new}}(\mathbf{x}_i, \mathbf{x}_e)). \quad (17)$$

Otherwise, some data in cluster  $e$  may change their exemplar as  $\mathbf{x}_\alpha$ ; we define these data as  $\mathbf{X}_{e,\alpha}^e$ , and the corresponding energy reduction  $R2_e$  is defined in the following equation:

$$R2_\alpha = \sum_{\mathbf{x}_i \in \mathbf{X}_{e,\alpha}^e} (S_{\text{new}}(\mathbf{x}_i, \mathbf{x}_\alpha) - S_{\text{new}}(\mathbf{x}_i, \mathbf{x}_e)). \quad (18)$$

Then, the energy reduction  $R2$  is defined as follows:

FIGURE 3: Case (i):  $\mathbf{x}_\alpha$  is an exemplar.FIGURE 4: Case (ii):  $\mathbf{x}_\alpha$  is not an exemplar.

$$R2 = S_{\text{new}}(\mathbf{x}_\alpha, \mathbf{x}_\alpha) - S_{\text{new}}(\mathbf{x}_\alpha, \mathbf{x}_{E_{\text{new}}(\alpha)}) + \max_{e \in E} \{R2_e, R2_\alpha\}. \quad (19)$$

In sum, the new target function equation (12) is optimized, and the optimal exemplar list for the EEG signals is generated.

**3.4. Time Complexity and Description.** The similarity relationship can be measured by Euclidean distance between data, defined as  $d(\mathbf{x}_i, \mathbf{x}_j)$  in this study. The proposed algorithm FEEC consists of two stages, namely, compression stage and optimization stage. After compression, the scale of similarity matrix reduces from  $N^2$  to  $Nc$ , so the optimization stage has the time complexity of  $O(c^2)$ . Therefore, the complexity of FEEC is considerably promising.

Based on the theoretical analysis above, the proposed FEEC for incomplete data can be summarized as Algorithm 1.

## 4. Experimental Study

To comprehensively evaluate the proposed algorithm FEEC, we have conducted several experiments based on both synthetic and real datasets. We also compare our new model with basic exemplar-based clustering model, namely, AP and EEM; to show these experimental results, we choose four performance indices in this section. In our experiments, all the algorithms were implemented using 2010a Matlab on a PC with 64 bit Microsoft Windows 10, an Intel(R) Core(TM) i7-4712MQ, and 8 GB memory.

**4.1. Data Preparation.** We choose Aggregation [32], as shown in Figure 5, and the Bonn EEG signal datasets in this section. The Bonn dataset [9, 10] is from the University of Bonn, Germany (<http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html>). The EEG dataset contains five groups (A to E and each group contains 100 single channel EEG segments of 23.6s duration. The sampling rate of all the datasets was 173.6 Hz. Figure 6 shows five healthy and epileptic EEG signals, and Table 2 lists detailed descriptions of these signals. Table 3 shows a brief description of these datasets. To construct the incomplete data scenario, we randomly choose 80% data as complete data and the remaining 20% as the incomplete data. We utilize KPCA to extract features from EEG signals in this section.

**4.2. Performance Indices.** Here, we give the definitions of the three adopted performance indices ENERGY, NMI, and accuracy. Along with the description in [12, 16, 30, 33, 34], we call the result outputted by these involved models as cluster and the true labels as class.

**4.2.1. ENERGY.** Since all the mentioned clustering algorithms are optimized, respectively, by the energy functions of the same type, we can compare them in terms of their energy values, defined as follows:

$$\text{ENERGY} = \sum_k \sum_i d(\mathbf{x}_k, \mathbf{x}_{k,i}), \quad (20)$$

where  $\mathbf{x}_k$  denotes the  $k$ th exemplar,  $\mathbf{x}_{k,i}$  is the  $i$ th data point in  $k$ th cluster, and  $d(\mathbf{x}_k, \mathbf{x}_{k,i})$  is the Euclidean distance between  $\mathbf{x}_k$  and  $\mathbf{x}_{k,i}$  which can be seen as a measurement of energy.

**Input:** Given incomplete data  $\mathbf{X} = [\mathbf{X}_c, \mathbf{X}_l] \in \mathbb{R}^{N \times D}$ ,  $\mathbf{X}_c$   
**Output:** Valid exemplar set  $E$ .

- (1) **Compression:** Apply basic exemplar-based clustering algorithm to the complete data  $\mathbf{X}_c$ ;
- (2) Get the potential exemplar set  $E_{\text{new}}$  by equation (11) and construct the new similarity matrix  $\mathbf{S}_{\text{new}} \in \mathbb{R}^{N \times c}$ ;
- (3) Generate the expansion order  $o$  on the potential exemplar list  $E_{\text{new}}$
- (4) Let  $t = 1$ ;
- (5) **for**  $e \in o$  **do**
- (6)   **if**  $e \in E$  **then**
- (7)     compute  $R1_\alpha, R1$  by equations (15) and (16);
- (8)     **if**  $R1_\alpha > 0$  **then**
- (9)       for  $\forall x_i \in \mathbf{X}_e$ , set  $E(x_i) = S(i)$  ;
- (10)    **end**
- (11)   **else**
- (12)     compute  $R2_e, R2_\alpha, R2$  by equations (17)–(19)
- (13)     **if**  $R2_\alpha > R2_e$  **then**
- (14)       for  $\forall x_i \in \mathbf{X}_{e,\alpha}^{le}$ , set  $E(x_i) = \alpha$
- (15)     **else**
- (16)       for  $\forall x_i \in \mathbf{X}_e$ , set  $E(x_i) = S(i)$
- (17)     **else**
- (18)       **if**  $R2 > 0$  **then**
- (19)         Accept the new exemplar  $\alpha$
- (20)     **end**
- (21)   **end**
- (22)    $t = t + 1$
- (23) **end**
- (24) Until converge.

ALGORITHM 1: Fast enhanced exemplar-based clustering algorithm.

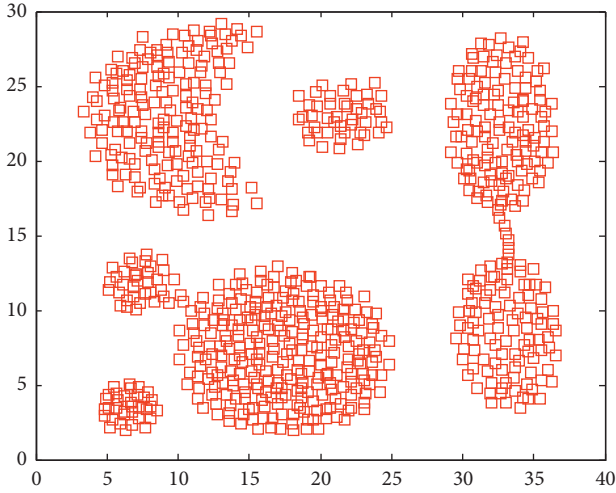


FIGURE 5: Description of Aggregation data.

**4.2.2. NMI.** NMI has been widely used to evaluate the clustering quality as well, and its value can be calculated by the following equation:

$$\text{NMI} = \frac{\sum_{i=1}^{|C|} \sum_{j=1}^{|C|} N_{i,j} \ln(N N_{i,j} / N_i N_j)}{\sqrt{(\sum_{i=1}^{|C|} N_i \ln(N_i / N)) (\sum_{j=1}^{|C|} N_j \ln(N_j / N))}} \quad (21)$$

where  $N_{i,j}$  is how clusters fit the classes,  $N_i$  is the number of data points in  $i$ th cluster,  $N_j$  is the number of data in  $j$ th class, and  $N$  is the total number of data points.

**4.2.3. Accuracy.** Accuracy is a more direct measure to reflect the effectiveness of clustering algorithms, which is defined as

$$\text{accuracy} = \frac{\sum_{i=1}^N \delta(c_i, \text{map}(\hat{c}_i))}{N}, \quad (22)$$

where  $c_i$  is the real label of data points and  $\hat{c}_i$  is the obtained clustering label.  $\delta(i, j) = 1$ , if  $i = j$ ;  $\delta(i, j) = 0$ , otherwise. Function  $\text{map}(\cdot)$  maps each obtained cluster to real class, and the optimized mapping function can be found in Hungarian algorithm.

The values of NMI and Acc range from 0 to 1, and the more it is close to 1, the more effective the clustering algorithm is. What is worth to mention is that we put % in the following relevant tables to show better precision. As to the performance index ENERGY, the smaller the value is, the better the clustering algorithm is.

**4.3. Experimental Results and Discussion.** The parameters involved FAP, AP, and EEM are in line with [13, 17, 29]. The preference  $s(i, i)$  is set to be the median value of similarities between data. We run each algorithm over 10 runs under same parameters; the average results are shown in Table 4. Moreover, the detailed comparison in terms of the above 3 terms, NMI, accuracy, and ENERGY, are shown in Figures 7–12 and Table 4, respectively.

By analyzing Figures 7–11 and Table 4 in detail, we can conclude the following:

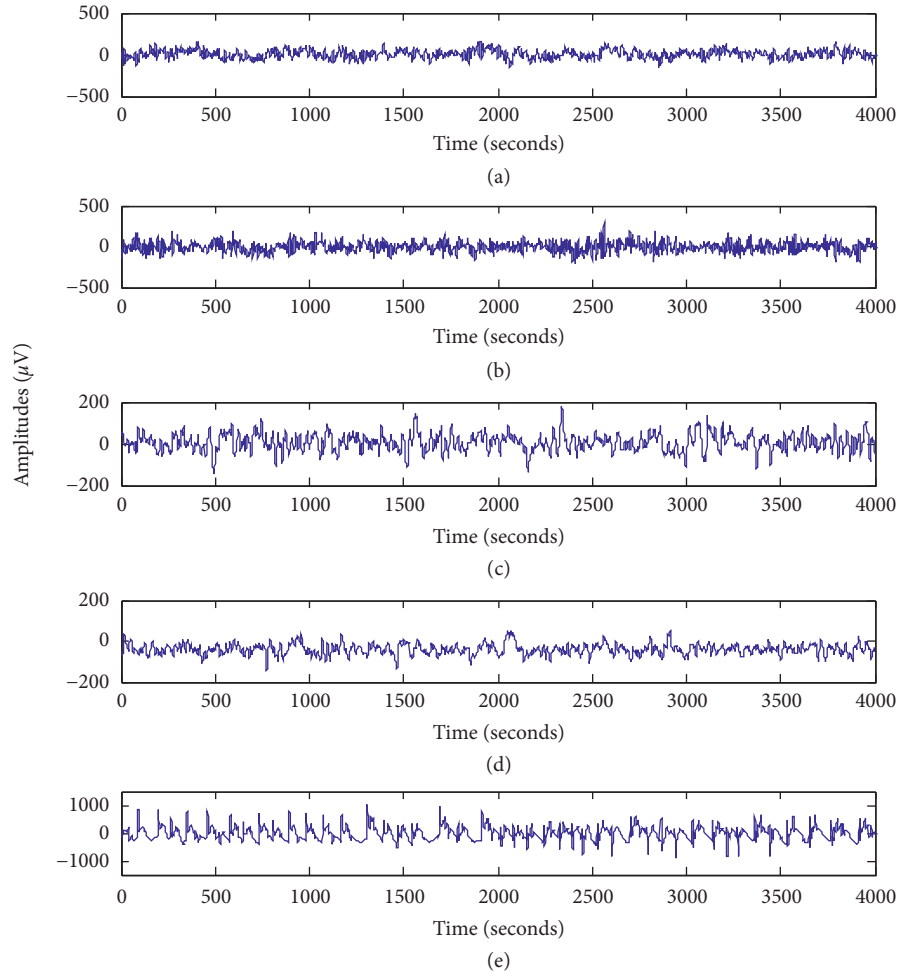


FIGURE 6: Typical EEG signals in groups A-E.

TABLE 2: Description of healthy and epileptic EEG data.

Subjects	Groups	Size of data	Descriptions
Healthy	A	100	Signals captured from volunteers with eyes open
	B	100	Signals captured from volunteers with eyes closed
Epileptic	C	100	Signals captured from volunteers during seizure silence intervals
	D	100	Signals captured from volunteers during seizure silence intervals
	E	100	Signals captured from volunteers during seizure activity

TABLE 3: Brief description of Aggregation and Bonn Epileptic EEG signals.

Datasets	Number of objects	Number of attributes	Number of clusters
Aggregation	788	2	7
Bonn	500	6	5

- (1) The proposed algorithm FEEC can cluster data with 80% complete data and 20% incomplete data, and in most cases, the performance is very convincing. Specifically, for both Aggregation and epileptic EEG signals,

FEEC performs best, in terms of NMI, accuracy, and ENERGY.

- (2) As to the computational time, compared with EEM, FEEC takes less time. Thus, with the assistance of

TABLE 4: Average experimental results over 10 runs on Aggregation and Bonn epileptic EEG signals.

Dataset	Algorithms	NMI	Accuracy	ENERGY	Time
Aggregation	FEEC	0.9636	0.9530	3522.38	9.54
	FAP	0.9017	0.9130	3609.79	9.65
	AP	0.8455	0.7016	4122.64	28.59
	EEM	0.8765	0.7208	3838.29	16.804
Bonn epileptic EEG signals	FEEC	0.9038	0.9786	1602.87	3.75
	FAP	0.8794	0.9445	1744.22	3.65
	AP	0.3324	0.4387	3018.12	3.97
	EEM	0.3728	0.4991	2050.51	6.01

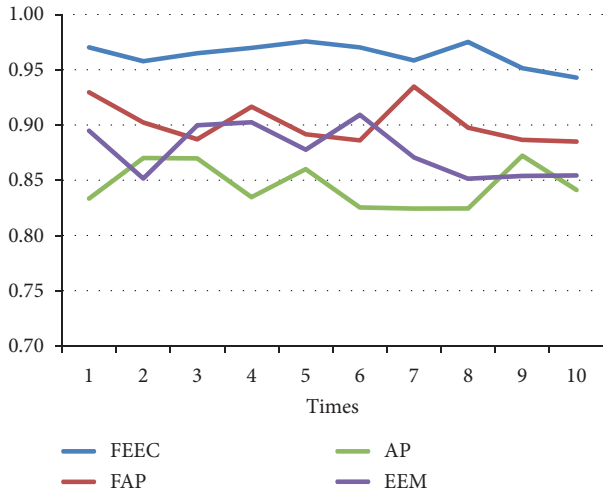


FIGURE 7: Comparison of NMI on Aggregation dataset.

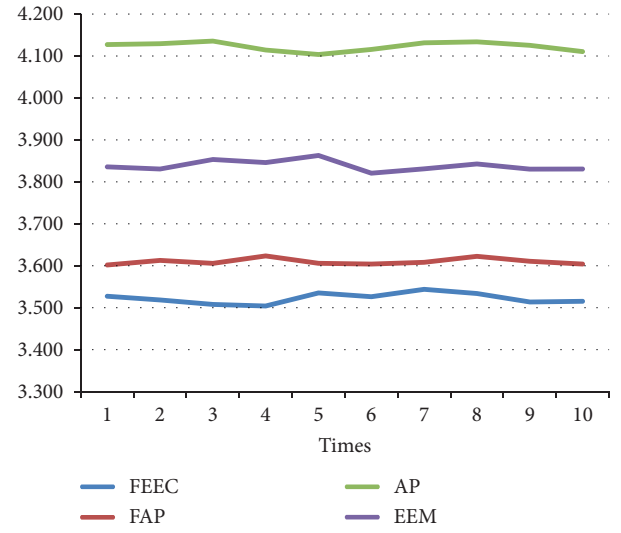


FIGURE 9: Comparison of accuracy on Aggregation.

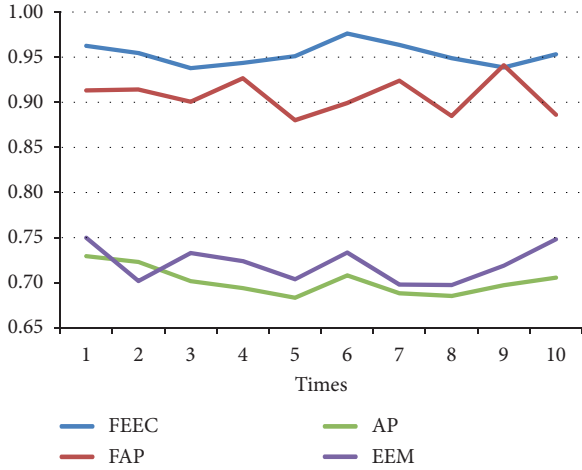


FIGURE 8: Comparison of accuracy on Aggregation dataset.

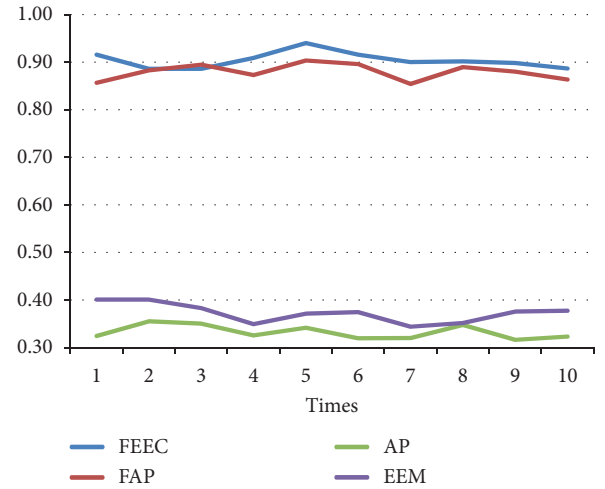


FIGURE 10: Comparison of NMI on Bonn epileptic EEG signals.

compression stage, the time complexity of FEEC has been reduced, and the efficiency is improved as well. And FEEC has an equivalent computational time with FAP. Comparing other criteria of FEEC and FAP, it is worthwhile to spend more time.

- (3) The proposed FEEC needs no more parameters except for preferences, while the performance of FAP

relies much on  $k$ , which determines the number of nearest exemplars. Accordingly, in terms of the involved datasets in this section, FEEC would achieve satisfactory clustering results.



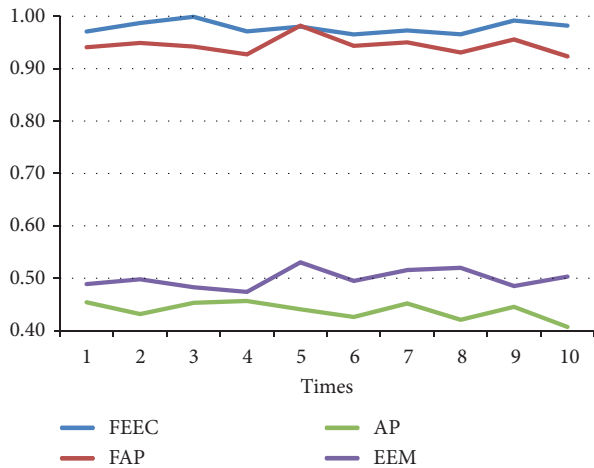


FIGURE 11: Comparison of accuracy on Bonn epileptic EEG signals.

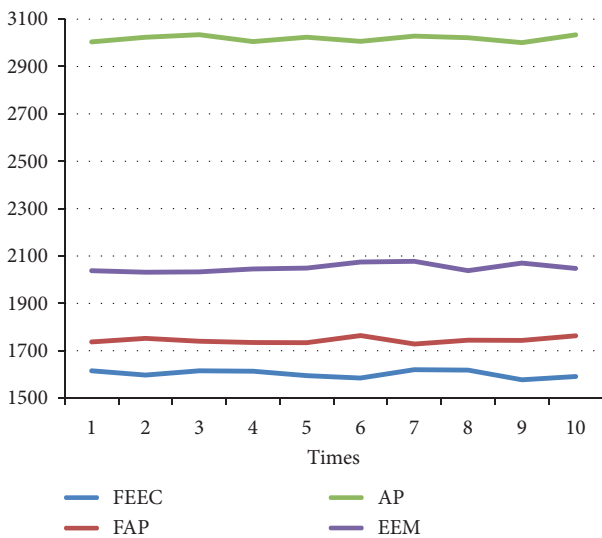


FIGURE 12: Comparison of ENERGY on Bonn epileptic EEG signals.

## 5. Conclusions

The diagnosis and treatment of epilepsy is always a significant direction for both machine learning and brain science. This paper newly proposes a fast exemplar-based clustering method for incomplete EEG signal. The FEEC method includes two stages, namely, compression and optimization. The performance of the proposed clustering algorithm is comprehensively verified by the experiments on two datasets.

Although most recognition methods of epilepsy are based on EEG signals at present, researchers also have to study on other neuroimaging modalities, such as cortical electroencephalography (ECoG), functional infrared optical imaging (fNIR), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and magnetoencephalography (MEG). Considering the fact that the brain activity is a nonlinear, networked, and unstable

complex system, we would focus on the multimodal clustering model for these neuroimaging modality signals in future.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported in part by the 2018 Natural Science Foundation of Jiangsu Higher Education Institutions under grant no. 18KJB5200001, the Natural Science Foundation of Jiangsu Province under grant no. BK20161268, and the Humanities and Social Sciences Foundation of the Ministry of Education under grant no. 18YJCZH229.

## References

- [1] H. Gao, H. Miao, L. Liu, J. Kai, and K. Zhao, "Automated quantitative verification for service-based system design: a visualization transform tool perspective," *International Journal of Software Engineering and Knowledge Engineering*, vol. 28, no. 10, pp. 1369–1397, 2018.
- [2] H. Gao, W. Huang, and X. Yang, "Applying probabilistic model checking to path planning in an intelligent transportation system using mobility trajectories and their statistical data," *Intelligent Automation and Soft Computing*, vol. 25, no. 3, Article ID 5478C559, 2019.
- [3] H. Gao, W. Huang, X. Yang, Y. Duan, and Y. Yin, "Toward service selection for workflow reconfiguration: an interface-based computing solution," *Future Generation Computer Systems*, vol. 87, pp. 298–311, 2018.
- [4] H. Gao, W. Huang, Y. Duan et al., "Research on cost-driven services composition in an uncertain environment," *Journal of Internet Technology*, vol. 20, no. 3, pp. 755–769, 2019.
- [5] L. Wang, K. C. L. Wong, H. Zhang, H. Liu, and P. Shi, "Noninvasive computational imaging of cardiac electrophysiology for 3-D infarct," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 4, pp. 1033–1043, 2011.
- [6] Y. Xia, H. Zhang, L. Xu et al., "An automatic cardiac arrhythmia classification system with wearable electrocardiogram," *IEEE Access*, vol. 6, pp. 16529–16538, 2018.
- [7] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [8] Y. Peng and B. L. Lu, "Immune clonal algorithm based feature selection for epileptic EEG signal classification," in *Proceedings of the 2012 11th International Conference on Information Science, Signal Processing and Their Applications (ISSPA)*, Montreal, Quebec, Canada, July 2012.
- [9] Y. Jiang, D. Wu, Z. Deng et al., "Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2270–2284, 2017.
- [10] Z. Jiang, F.-L. Chung, and S. Wang, "Recognition of multiclass epileptic EEG signals based on knowledge and label space

- inductive transfer," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 630–642, 2019.
- [11] Z. Iscan, Z. Dokur, and T. Demiralp, "Classification of electroencephalogram signals with combined time and frequency features," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10499–10505, 2011.
  - [12] Y. Jiang, F.-L. Chung, S. Wang, Z. Deng, J. Wang, and P. Qian, "Collaborative fuzzy clustering from multiple weighted views," *IEEE Transactions on Cybernetics*, vol. 45, no. 4, pp. 688–701, 2015.
  - [13] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2017.
  - [14] V. Kolmogorov and C. Rother, "Comparison of energy minimization algorithms for highly connected graphs," in *Proceedings of the 9th European Conference on Computer Vision (ECCV)*, vol. 3952, Graz, Austria, May 2006.
  - [15] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pp. 900–906, Nice, France, October 2003.
  - [16] A. Bi and S. Wang, "Incremental enhanced  $\alpha$ -expansion move for large data: a probability regularization perspective," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 5, pp. 1615–1631, 2016.
  - [17] Y. Zheng and P. Chen, "Clustering based on enhanced  $\alpha$ -expansion move," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2206–2216, 2013.
  - [18] B. R. Greene, S. Faul, W. Marnane, G. Lightbody, I. Korotchikova, and G. Boylan, "A comparison of quantitative EEG features for neonatal seizure detection," *Clinical Neurophysiology*, vol. 119, no. 6, pp. 1248–1261, 2008.
  - [19] C. W. N. F. C. W. Fadzal, W. Mansor, L. Y. Khuan, and A. Zabidi, "Short-time fourier transform analysis of EEG signal from writing," in *Proceedings of the 2012 IEEE 8th International Colloquium on Signal Processing and its Applications*, Malacca, Malaysia, March 2012.
  - [20] E. A. Vivaldi and A. Bassi, "Frequency domain analysis of sleep EEG for visualization and automated state detection," in *Proceedings of the 2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, New York, NY, USA, September 2006.
  - [21] D. Hu, W. Li, and X. Chen, "Feature extraction of motor imagery eeg signals based on wavelet packet decomposition," in *Proceedings of the 2011 IEEE/ICME International Conference on Complex Medical Engineering*, pp. 694–697, Harbin, China, May 2011.
  - [22] Z. Zhang, H. Kawabata, and Z. Q. Liu, "EEG analysis using fast wavelet transform," in *Proceedings of the SMC 2000 Conference Proceedings: 2000 IEEE International Conference on Systems, Man and Cybernetics*, Nashville, TN, USA, October 2000.
  - [23] K. Murphy, Y. Weiss, and M. Jordan, "Loopy belief propagation for approximate inference: an empirical study," in *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, Bellevue, WA, USA, August 2013.
  - [24] K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, "Adaptive affinity propagation clustering," 2008, <https://arxiv.org/abs/0805.1096>.
  - [25] J. Vlasblom and S. J. Wodak, "Markov clustering versus affinity propagation for the partitioning of protein interaction graphs," vol. 10, no. 1, pp. 99–100, 2009.
  - [26] X. Mishra, R. Guan, L. Wang, Z. Pei, and Y. Liang, "An incremental affinity propagation algorithm and its applications for text clustering," in *Proceedings of the 2009 International Joint Conference on Neural Networks*, pp. 2914–2919, Atlanta, GA, USA, June 2009.
  - [27] L. Sun and C. Guo, "Incremental affinity propagation clustering based on message passing," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 11, pp. 2731–2744, 2014.
  - [28] I. Givoni and B. Frey, "Semi-supervised affinity propagation with instance-level constraints," *Journal of Machine Learning Research—Proceedings Track*, vol. 5, pp. 161–168, 2009.
  - [29] L. Sun, C. Guo, C. Liu, and H. Xiong, "Fast affinity propagation clustering based on incomplete similarity matrix," *Knowledge and Information Systems*, vol. 51, no. 3, pp. 941–963, 2017.
  - [30] A. Bi, F. Chung, S. Wang, Y. Jiang, and C. Huang, "Bayesian enhanced  $\alpha$ -expansion move clustering with loose link constraints," *Neurocomputing*, vol. 194, pp. 288–300, 2016.
  - [31] S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O'Callaghan, "Clustering data streams: theory and practice," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 3, pp. 515–528, 2003.
  - [32] A. Gionis, H. Mannila, and P. Tsaparas, "Clustering aggregation," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, p. 4, 2007.
  - [33] Y. Jiang, Z. Deng, F.-L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
  - [34] Y. Jiang, F.-L. Chung, H. Ishibuchi, Z. Deng, and S. Wang, "Multitask TSK fuzzy system modeling by mining intertask common hidden structure," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 548–561, 2015.

## Research Article

# Cross-Subject Seizure Detection in EEGs Using Deep Transfer Learning

Baocan Zhang<sup>1</sup>, Wennan Wang<sup>2</sup>, Yutian Xiao<sup>3</sup>, Shixiao Xiao<sup>1</sup>, Shuaichen Chen<sup>3</sup>,  
Sirui Chen<sup>4</sup>, Gaowei Xu<sup>4</sup> and Wenliang Che<sup>5</sup>

<sup>1</sup>Chengyi University College, Jimei University, Xiamen 361021, China

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

<sup>3</sup>School of Informatics, Xiamen University, Xiamen 361001, China

<sup>4</sup>School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

<sup>5</sup>Department of Cardiology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 201804, China

Correspondence should be addressed to Shixiao Xiao; [xiaoshixiao@jmu.edu.cn](mailto:xiaoshixiao@jmu.edu.cn), Gaowei Xu; [gaoweixu@tongji.edu.cn](mailto:gaoweixu@tongji.edu.cn), and Wenliang Che; [chewenliang@tongji.edu.cn](mailto:chewenliang@tongji.edu.cn)

Received 28 January 2020; Revised 6 March 2020; Accepted 26 March 2020; Published 8 May 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Baocan 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.

Electroencephalography (EEG) plays an import role in monitoring the brain activities of patients with epilepsy and has been extensively used to diagnose epilepsy. Clinically reading tens or even hundreds of hours of EEG recordings is very time consuming. Therefore, automatic detection of seizure is of great importance. But the huge diversity of EEG signals belonging to different patients makes the task of seizure detection much challenging, for both human experts and automation methods. We propose three deep transfer convolutional neural networks (CNN) for automatic cross-subject seizure detection, based on VGG16, VGG19, and ResNet50, respectively. The original dataset is the CHB-MIT scalp EEG dataset. We use short time Fourier transform to generate time-frequency spectrum images as the input dataset, while positive samples are augmented due to the infrequent nature of seizure. The model parameters pretrained on ImageNet are transferred to our models. And the fine-tuned top layers, with an output layer of two neurons for binary classification (seizure or nonseizure), are trained from scratch. Then, the input dataset are randomly shuffled and divided into three partitions for training, validating, and testing the deep transfer CNNs, respectively. The average accuracies achieved by the deep transfer CNNs based on VGG16, VGG19, and ResNet50 are 97.75%, 98.26%, and 96.17% correspondingly. On those results of experiments, our method could prove to be an effective method for cross-subject seizure detection.

## 1. Introduction

Epilepsy, a disorder of normal brain function characterized by the existence of abnormal synchronous discharges in the cerebral cortex, impacts approximately 2% of the world's population and is likely to jeopardize their health and life. The epilepsy diagnose is always down by analyzing electroencephalogram (EEG), which includes scalp EEG and intracranial EEG. Scalp EEG signals have been widely studied because they are relatively cheap and easy to gain. Commonly, 19 recording electrodes and a system reference are placed on the scalp area according to specifications by the International 10-20 system. For the seizure detection task, it is required to

analyze the EEG signal thoroughly towards the decision of the existence of epileptic seizure or not, and this is an extensive clinical experience required and time-consuming job [1]. Due to the relatively infrequent nature of epileptic seizures, long-term EEG recordings are necessary, to make the situation of visually reading EEG signal by human experts worse. Therefore, computer-based technology for automatic seizure detection is urgently needed and is also a key method to save time and effort.

In previous studies, numerous detection algorithms have been proposed [2–4]. The features extracted from EEG signals come from time domain, frequency domain, time-frequency analysis [5, 6], wavelet analysis [7, 8], and so on.

After the features are extracted, classifiers such as SVM and other machine learning algorithms are frequently used in the classification phase [9]. For example, Subasi et al. created a hybrid model to optimize the SVM parameters, showing that the proposed hybrid SVM model is a competent method to detect epileptic seizures using EEG [10].

In recent years, deep learning (DL) has been proven to be very successful in image classification, object detection, and segmenting, exhibiting near-human abilities to perform many tasks. DL extracts the global synchronization features automatically and does not need any a priori knowledge. Huang et al. proposed a coupled neural network for brain medical image [11, 12] and a deep residual segmentation network for analysis of IVOCT images [13]. Also, a number of recent studies demonstrated the efficacy of deep learning in the classification of EEG signals and seizure detection [14]. Convolutional neural network (CNN), as one of the most widely used deep learning models, is always used. For example, Wang et al. proposed a 14-layer CNN for multiple sclerosis identification [15]. For the seizure detection task, there are two ways of using the original EEG signals as the input image of CNN. On one hand, segments of raw EEG signals with different lengths serve as input directly. Emami et al. divided EEG signals into short segments on a given time window and converted them into plot images; each of which was classified by VGG16 as “seizure” or “nonseizure” [16]. Their experiments resulted in that the median true positive rate of CNN labeling was 74%. On the other hand, time and frequency domain signals extracted from raw EEG signals serve as input image of CNN. Zhou et al. designed a CNN with no more than three layers to detect seizure, with time-frequency spectrum image as input and achieved average accuracy 93% [17]. They also compared the performance of time domain with that of time-frequency domain, which resulted that frequency domain signals have greater potential than time domain signals for CNN applications. Although, deep learning especially CNN has made remarkable progress in the field of EEG classification and seizure detection, the performance of seizure detection still requires improvement. The challenge comes from two aspects. Firstly, training a deep learning model such as VGG16 needs a large amount of labeled data. However, most of the EEG signals data are unlabeled. In this study, the public-labeled scalp EEG dataset from the Children’s Hospital Boston-Massachusetts Institute of Technology (CHB-MIT, see <http://physionet.org/physiobank/dataset/chbmit> [18]) will be used as the raw signals. Due to the relatively infrequent nature of epileptic seizures, the raw signals will be augmented to avoid extremely unbalanced dataset for training. Secondly, EEG signals are person-specific. On the other hand, Orosco proposed a patient nonspecific strategy for seizure detection based on stationary wavelet transform of EEG signals and reported the mean sensitivity of 87.5% and specificity of 99.9% [19]. Hang et al. proposed a novel deep domain adaption network for cross-subject EEG signal recognition based on CNN and used the maximum mean discrepancy to minimize the distribution discrepancy between source and target subjects [20]. Akyol presented a stacking ensemble based deep neural network model for seizure detection. Experiments were carried out on the EEG dataset from

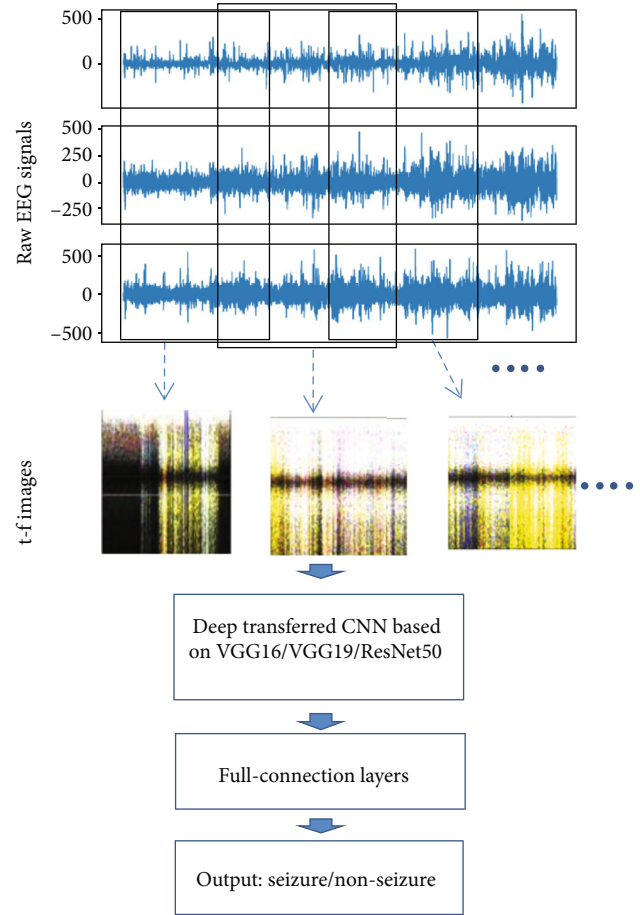


FIGURE 1: The flow of seizure detection.

Bonn University and came to the result that the average accuracy is 97.17% along with average sensitivity of 93.11% [21]. Zhang et al. proposed an explainable epileptic seizure detection model to the pure seizure-specific representation for EEG signal through adversarial training, in order to overcome the discrepancy of different subjects [22].

CNN models like VGG16 have millions of parameters to be trained, not to mention deeper network like googLeNet. Transfer learning has emerged to tackle this problem, especially in real-world applications. Transfer learning is always down by a pretrained model, which is trained on the benchmark dataset like ImageNet. The pretrained model can extract universal low-level features of images and can tremendously improve the efficiency of using CNN. However, the pretrained model should be fine-tuned in order to match the target dataset and its goal. For example, Shao et al. created a deep transfer CNN for fault diagnosis, in which a pretrained CNN model is used to accelerate the training process [23].

In this paper, three transfer CNN models, based on VGG16, VGG19, and ResNet50, respectively, are proposed for seizure detection. The flow of seizure detection is shown in Figure 1. Let us take VGG16 as an example. The target CNN network consists of a pretrained VGG16 model with nontop layers frozen and fine-tuned top layers. The pretrained model uses parameters based on ImageNet. The



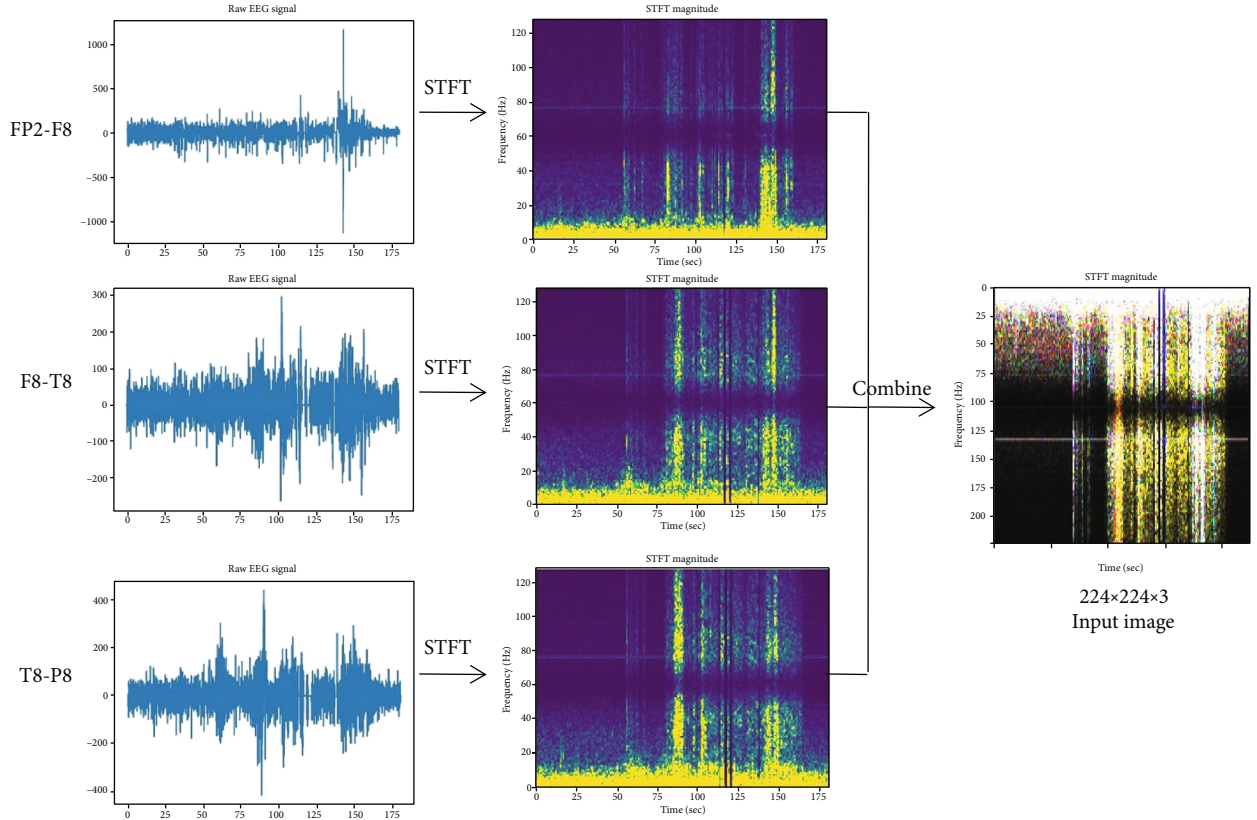


FIGURE 2: The conversion process from raw EEG signals to time-frequency image. In CHB-MIT EEG database, segments of raw signal from subject No. 13 are converted to t-f images by using STFT. Then, three t-f images are treated as three channels to form an input image.

fine-tuned top layers include an output layer with two categories: seizure and nonseizure.

For experiments, the public CHB-MIT EEG dataset is used. Raw EEG signals from FP2-F8, F8-T8, and T8-P8 electrodes are converted to time-frequency spectrum image using short-time Fourier transform (STFT) [24] and then fused as one image, inspired by [25]. The fused images from different persons all putted together are the target dataset. Then, the target dataset serves as the input of the deep transfer models, to perform cross-subject seizure detection.

The remainder of this paper is organized as follows. Section 2 introduces the used dataset from CHB-MIT. Section 3 describes in detail the conversion process of EEG signal to time-frequency spectrum image and introduces all three deep transfer models. Section 4 conducts experiments and gives a comparison of three models. Section 5 discusses the results. Section 6 concludes the paper with a summary.

## 2. Dataset Description

The dataset used in this paper is an open-source EEG database from the MIT PhysioNet, collected at the Children's Hospital Boston (CHB-MIT). The dataset consists of recordings from pediatric subjects with intractable seizures using scalp electrodes. Recordings are grouped into 23 cases. Each case contains 9 to 42 hours' continuous recordings from a single subject. All subjects were asked to stop related medical treatments one week before data collection. The sampling

frequency for all subjects was 256 Hz. The start time and end time of epileptic seizure were labeled explicitly based on expert judgments. Most recordings contain 23 EEG channels and multiple seizure occurrences.

The duration of seizure varies for each subject very much. Thus, some of the recordings with relatively long seizure durations are used only. The reason for not including all the recordings is that recordings with a low proportion of seizure duration lead to an unbalanced dataset and would cause over-fitting of the CNN model.

## 3. Methods

In this section, we convert the raw EEG signals to time-frequency (t-f) spectrum images by using STFT and combine three t-f images from different channels as one input image, as shown in Figure 2. Then, deep transfer CNNs are proposed for seizure detection.

**3.1. Data Preparation Based on STFT.** The short-time Fourier transform (STFT) is used to analyze how the frequency content of a nonstationary signal changes over time. The STFT of a signal is calculated by sliding a window over the signal and calculating the discrete Fourier transform of the windowed signal. The window moves along the time axis at a given interval, with overlap or not. Commonly, overlap is used in order to compensate for the signal attenuation at the window edges.



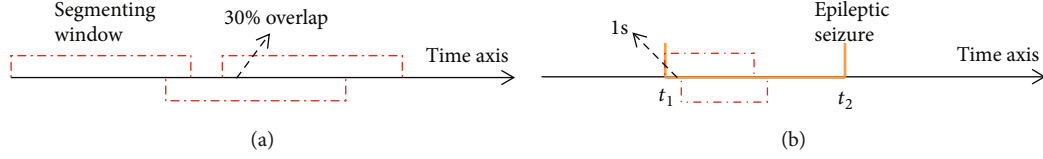


FIGURE 3: (a) Segmenting the signal of EEG along the time axis with 30% overlap and (b) segmenting the signal with one second per step in time interval  $[t_1, t_2]$ , while epileptic seizure happening.

Formally, STFT is defined as follows:

$$\text{STFT}(\tau, \omega) = \int_{-\infty}^{+\infty} x(t)h(t - \tau)e^{-j\omega t} dt, \quad (1)$$

where  $x(t)$  is the original EEG signal,  $h(t)$  is a window function, and  $\tau$  is the window position on time axis.

Although, epileptic seizure is person-specific, they could have something in common. According to [24], signals from FP2-F8, F8-T8, and T8-P8 electrodes are relatively prominent. In this paper, we use signals from those three electrodes. And the t-f images from FP2-F8, F8-T8, and T8-P8 are treated as red, green, and blue channels, respectively, while combining them as one input image.

Due to the infrequent nature of seizures, the number of positive samples should be increased in order to avoid unbalanced dataset by augmenting. In detail, we prepare the dataset by two steps. Step one, as shown in Figure 3(a), for each signal of EEG, we move a window of length 180 seconds along the time axis with 30% overlap. Then for each windowed segment, we use STFT to calculate complex amplitude versus time and frequency, where window used by STFT is 413 signal points long and is moved along the time axis with 50% overlap. This results in a spectrum image whose size is  $207 \times 224$ , where 207 is the number of sample frequency and 224 the number of segment times. At last, the spectrum image is resized to  $224 \times 224$ . When epileptic seizure happens in the windowed segment, the resized spectrum image is labeled positive, otherwise negative. Step two, as shown in Figure 3(b), for each signal of EEG, let us assume that epileptic seizure happens at time interval  $[t_1, t_2]$ . Let the segmenting window start at  $t_1$  and move along the time axis one second per step, and furthermore overlap with  $[t_1, t_2]$  no less than 3 seconds. Then, each segmented window is converted to a spectrum image of size  $224 \times 224$  as in step one. All spectrum images in this step are labeled positive.

In order to avoid an unbalanced dataset, only the EEG signals of subjects No. 05, 08, 11, 12, 13, 14, 15, 23, and 24 with relative long duration of epileptic seizure are used. This results in the target dataset of 8474 images of size  $224 \times 224 \times 3$ , of which about 49.5% are labeled positive.

**3.2. Deep Transfer Model.** A pretrained model is a saved network that was previously trained on a huge dataset, such as ImageNet. Then, we can use transfer learning to customize this model to a given task. Intuitively, if a model is trained on a large and general dataset, this model could extract low-level features and serve as a generic model of

the visual world. Then, we could use this model to extract meaningful features from new samples and add a new classifier on top of it to do specific classification, where only the added layers should be trained from scratch on our dataset. Transfer learning has also been proved effective in applications [26, 27].

In this paper, three deep transfer models are proposed for comparison:

- (1) The deep transfer model based on VGG16 (referred as Model-1): deep neural network VGG16 is a well-known CNN model with 16 layers introduced in 2014 and has achieved amazing performance in various image tasks. VGG16 is characterized by small-sized filters, which is very suitable for our purpose of detecting the difference in frequency between seizure and nonseizure. As can be seen in Figure 4(a), Model-1 consists of a transferred VGG16 and top trainable layers. The transferred VGG16, with the output layer removed from the original model, is pretrained using the ImageNet database. The added top trainable layers consist of two trainable full connection layers and a softmax output layer with two neurons corresponding to seizure or nonseizure.
- (2) The deep transfer model based on VGG19 (referred as Model-2): Model-2 has almost the same structure as Model-1. The only difference between those two transfer models is that one use pretrained VGG16 and the other use pretrained VGG19.
- (3) The deep transfer model based on ResNet50 (referred as Model-3): ResNet50 is one of the famous residual neural networks, which are characterized by utilizing shortcuts to jump over some layers. As shown in Figure 4(b), the first part of Model-3 is a pretrained ResNet50 without top layers. Then after global average pooling, two full connection layers of 2048 neurons are added. The output is a softmax output layer with two neurons.

The loss function used in three models is softmax cross entropy, which is defined as

$$H(r, p) = -\sum_i r_i \cdot \log(p_i) \quad (2)$$

where  $r$  and  $p$  are the labeled and predicted probabilities, respectively.

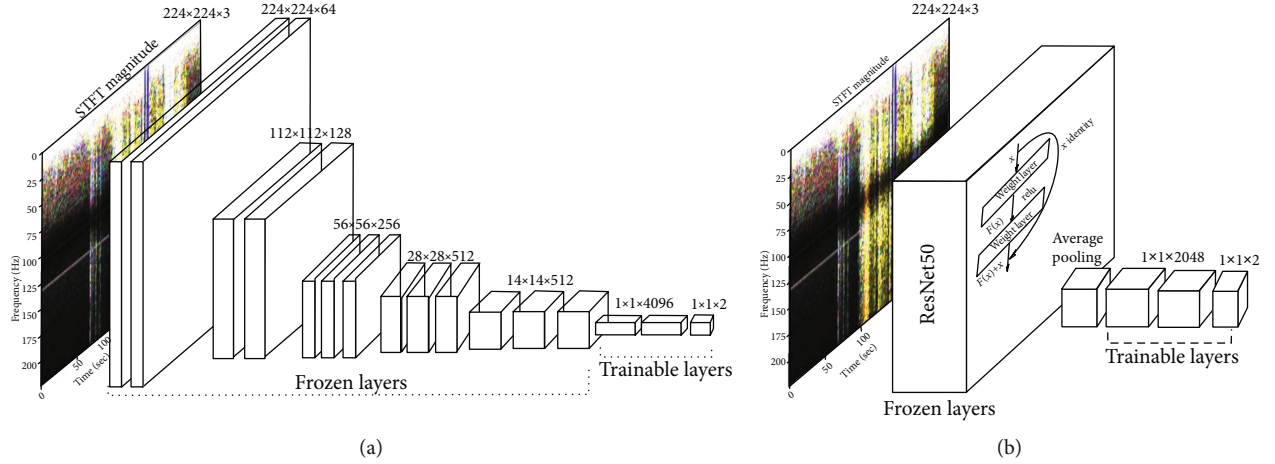


FIGURE 4: (a) The deep transfer model from VGG16, where the output layer is replaced by a new softmax output layer with two neurons, corresponding to seizure or nonseizure. (b) The deep transfer model from ResNet50, where the output layer is replaced by a new softmax output layer with two neurons, corresponding to seizure or nonseizure, and the size of full connection layers are also altered.

## 4. Experiments and Results

Experiments have been carried out to evaluate the performance of the proposed deep transfer models. The test platform is a desktop system with Nvidia RTX 2080Ti and 64GB memory running Ubuntu.

**4.1. Training the Deep Transfer Network.** After being shuffled, the target datasets are divided into training set, validation set, and test set, which occupy 60%, 20%, and 20%, respectively.

Model-1 and Model-2 almost have the same structure, so their training methods are of the same. Let us take Model-1 for example. The parameters of VGG16 pretrained on ImageNet are transferred to the network and would be frozen while training. Other trainable parameters are initialized randomly. The optimizer is SGD with learning rate of 0.001 and a small decay of  $1e-5$ . Then, the network is trained and validated on the training set and validation set, respectively, with batch size of 64. As for Model-3, we transfer the parameters of ResNet50 pretrained also on ImageNet and froze them through training. Other trainable parameters are also initialized randomly. For the training optimizer, the Adam algorithm is used, with a starting learning rate of 0.001, beta\_1 of 0.9, and beta\_2 of 0.999. The learning rate would be reduced by a factor 0.8 when validation loss has stopped descending for 5 epochs. Then, the network is trained and validated on the training set and validation set, respectively, with a batch size of 16. For all three models, the epoch is set to be 500, but the training would be stopped while validation loss is not descending for 20 epochs.

**4.2. Results and Analysis.** The statistical measures for evaluating classification performance contain accuracy (acc), sensitivity (recall), precision, and the Matthews correlation coefficient (mcor). The measure mcor takes into account true and false positives and negatives and is regarded as a bal-

anced measure. The more mcor approaches 1, the better the prediction is. Formally, their definitions are as follows:

$$\text{accuracy} = \frac{TP + TN}{N},$$

$$\text{precision} = \frac{TP}{TP + FP},$$

$$\text{recall} = \frac{TP}{TP + FN},$$

$$\text{mcor} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (3)$$

where TP means the true positive, TN the true negative, FP the false positive, FN the false negative, and  $N$  the total.

As shown in Figure 5, both the loss and accuracy of Model-1 and Model-2 converge after about 170 epochs. The loss and accuracy of Model-3 converge after about 100 epochs. But the metrics, including loss and accuracy, of deep transfer networks based on VGG16 and VGG19 are better than those of the deep transfer network based on ResNet50. Because the test dataset is randomly selected from the target dataset, the training and testing processes are carried out 10 times, and then, the average of all metrics is used. As shown in Table 1, the performance of Model-1 is almost the same as that of Model-2. But both models based on VGG outperform the model based on ResNet50 (Model-3).

For comparison, different sizes of full connection layers of each model are used for seizure detection. All experiments are carried out with the same training, validating and testing datasets. The results are shown in Table 2.

## 5. Discussion

For seizure detection, the methods frequently used include classical ones like SVM and modern ones like convolutional

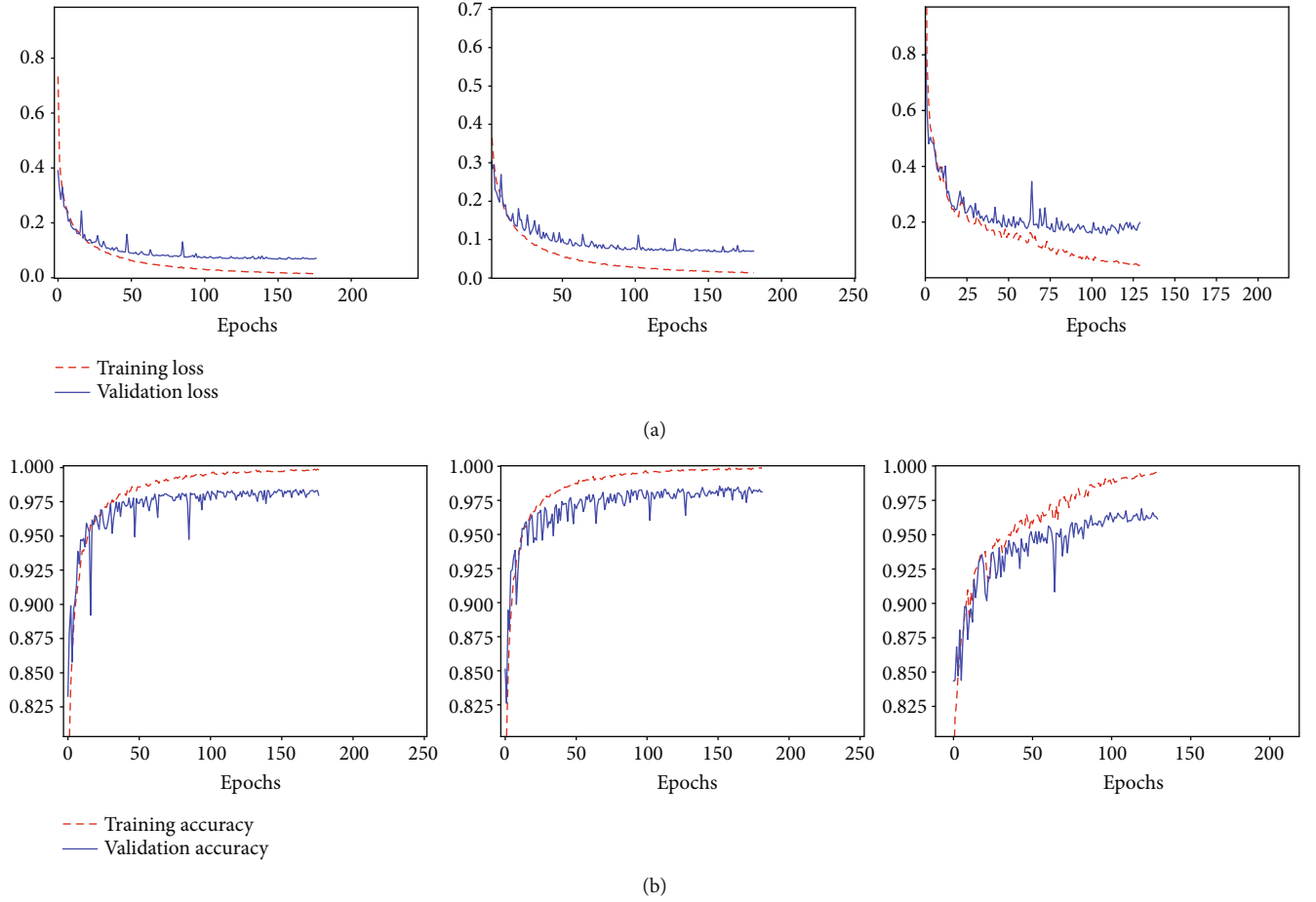


FIGURE 5: From left to right (with full connection size of  $4096 \times 4096$ ,  $4096 \times 4096$ , and  $2048 \times 2048$ , respectively). (a) The loss of transferred mode based on VGG16, VGG19, and ResNet50. (b) The accuracy of transferred mode based on VGG16, VGG19, and ResNet50.

TABLE 1: The test loss, accuracy, and metrics of recall, precision, and mcor (average).

Deep transfer network	Loss	Accuracy	Recall	Precision	Mcor
Model-1 (on VGG16)	0.0659	0.9795	0.9844	0.9747	0.9589
Model-2 (on VGG19)	0.0699	0.9826	0.9801	0.9851	0.9653
Model-3 (on ResNet50)	0.2249	0.9617	0.9865	0.9400	0.9246

neural networks. Conventional approaches rely on time and frequency, where methods based on frequency are more capable and efficient than on time. But the frequency bands should be customized for a particular patient, which makes it very difficult to generalize this method to different patients. The latest CNNs used in this study are highly suited for EEG classification because they can select features adaptively and automatically. In order to take the advantage of frequency, we use the time-frequency images as an input dataset of the CNNs.

The prominent CNN like VGG16 has tens of millions of parameters. If all those parameters are trained from scratch, millions of images would be needed to ensure that the network could select features properly. The demand of so many images could be almost impossible to meet due to the infrequent nature of epileptic seizure. On the other hand, the images from ImageNet and our time-frequency

images would have low-level universal features in common. So, we transfer the parameters pretrained on ImageNet to our models, to extract universal features, and train from scratch the parameters of the full connection layers.

Most existing literatures on seizure detection are patient-specific, which require a priori knowledge of the patient. In this study, the deep transfer models, by using the CHB-MIT EEG dataset as the original signals, are patient-independent. The t-f images from different objects are put together and shuffled, from which the testing dataset is selected randomly. For comparison, three transfer models based on VGG16, VGG19, and ResNet50, respectively, are proposed. Experiments are carried out to evaluate their performance of cross-subject seizure detecting. In detail, the input dataset, generated from the original signals, consists of 8474 t-f images. The ratio of positive to negative samples

TABLE 2: The metrics of each model with different full connection size.

Size of FC	Base model	Loss	Accuracy	Recall	Precision	Mcor
2048 × 2048	Model-1	0.0631	0.9837	0.9837	0.9839	0.9674
	Model-2	0.0698	0.9816	0.9844	0.9789	0.9632
	Model-3	0.1790	0.9578	0.9596	0.9563	0.9157
1024 × 1024	Model-1	0.0638	0.9826	0.9844	0.9810	0.9653
	Model-2	0.0731	0.9805	0.9823	0.9788	0.9611
	Model-3	0.1739	0.9663	0.9806	0.9526	0.9331
4096 × 2048	Model-1	0.0596	0.9833	0.9823	0.9844	0.9667
	Model-2	0.0712	0.9791	0.9851	0.9734	0.9583
	Model-3	0.1820	0.9568	0.9724	0.9430	0.9140
2048 × 1024	Model-1	0.0618	0.9830	0.9851	0.9810	0.9660
	Model-2	0.0690	0.9815	0.9844	0.9789	0.9631
	Model-3	0.2020	0.9567	0.9752	0.9406	0.9142
4096 × 1024	Model-1	0.0637	0.9816	0.9851	0.9782	0.9631
	Model-2	0.0672	0.9808	0.9851	0.9768	0.9617
	Model-3	0.2091	0.9564	0.9752	0.9400	0.9135
1024 × 512	Model-1	0.0673	0.9809	0.9837	0.9782	0.9618
	Model-2	0.0758	0.9777	0.9872	0.9687	0.9555
	Model-3	0.1649	0.9642	0.9851	0.9456	0.9292
1024 × 256	Model-1	0.0624	0.9844	0.9823	0.9865	0.9688
	Model-2	0.0664	0.9808	0.9872	0.9748	0.9618
	Model-3	0.1452	0.9610	0.9653	0.9572	0.9221
512 × 512	Model-1	0.0678	0.9769	0.9851	0.9693	0.9541
	Model-2	0.0669	0.9815	0.9851	0.9782	0.9632
	Model-3	0.1201	0.9773	0.9879	0.9674	0.9548
512 × 256	Model-1	0.0653	0.9794	0.9844	0.9747	0.9589
	Model-2	0.0841	0.9720	0.9880	0.9575	0.9445
	Model-3	0.1198	0.9745	0.9844	0.9653	0.9491

is almost one to one after augmentation. Then, all three transfer models are trained, validated, and tested on our augmented dataset. The average accuracies are 97.95%, 98.26, and 96.17%, respectively.

However, in the original EEG dataset, not all signals are used, due to the short durations of epileptic seizure. Those cases are not rare. So, a method for seizure detection including those objects will be addressed in the future. GAN might be a choice to tackle the problem, with its utilization of generative model.

## 6. Conclusions

This study gives a method to detect seizure in EEGs for cross-subjects, by using deep transfer learning. Three deep transfer models are proposed based on VGG16, VGG19, and ResNet50, respectively. Experiments are performed to evaluate the models over the CHB-MIT EEG dataset, without the need for denoising the EEG signals. Also, on the same dataset, experiments of the three models with full connection layers of different sizes are carried out for comparison. In the future, we plan to extend this method to EEG signals with a relatively short duration of epileptic seizure.

## 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 Youth Teacher Education and Research Funds of Fujian (Grant No. JAT191152).

## References

- [1] S. R. Benbadis, "The tragedy of over-read EEGs and wrong diagnoses of epilepsy," *Expert Review of Neurotherapeutics*, vol. 10, no. 3, pp. 343–346, 2014.
- [2] L. Orosco, A. G. Correa, and E. Laciari, "Review: a survey of performance and techniques for automatic epilepsy detection," *Journal of Medical and Biological Engineering*, vol. 33, no. 6, pp. 526–537, 2013.

- [3] T. N. Alotaiby, S. A. Alshebeili, T. Alshawhi, I. Ahmad, and F. E. Abd el-Samie, "EEG seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, 2014.
- [4] S. N. Baldassano, B. H. Brinkmann, H. Ung et al., "Crowdsourcing seizure detection: algorithm development and validation on human implanted device recordings," *Brain*, vol. 140, no. 6, pp. 1680–1691, 2017.
- [5] K. S. Anusha, M. T. Mathews, and S. D. Puthankattil, "Classification of normal and epileptic EEG signal using time & frequency domain features through artificial neural network," in *2012 international conference on advances in computing and communications*, pp. 98–101, IEEE, 2012.
- [6] Z. K. Gao, Q. Cai, Y. X. Yang, N. Dong, and S. S. Zhang, "Visibility graph from adaptive optimal kernel time-frequency representation for classification of Epileptiform EEG," *International Journal of Neural Systems*, vol. 27, no. 4, 2017.
- [7] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *Journal of Neuroscience Methods*, vol. 123, no. 1, pp. 69–87, 2003.
- [8] H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, pp. 205–211, 2007.
- [9] E. Gysels, P. Renevey, and P. Celka, "SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces," *Signal Processing*, vol. 85, no. 11, pp. 2178–2189, 2005.
- [10] A. Subasi, J. Kevric, and M. A. Canbaz, "Epileptic seizure detection using hybrid machine learning methods," *Neural Computing and Applications*, vol. 31, no. 1, pp. 317–325, 2019.
- [11] C. Huang, X. Shan, Y. Lan et al., "A hybrid active contour segmentation method for Myocardial D-SPECT Images," *IEEE Access*, vol. 6, pp. 39334–39343, 2018.
- [12] C. Huang, C. Wang, J. Tong, L. Zhang, F. Chen, and Y. Hao, "Automatic quantitative analysis of bioresorbable vascular scaffold struts in optical coherence tomography images using region growing," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 1, pp. 98–104, 2018.
- [13] C. Huang, Y. Peng, F. Chen et al., "A deep segmentation network of multi-scale feature fusion based on attention mechanism for IVOCT lumen contour," *IEEE/ACM Transactions on computational biology and bioinformatics*, p. 1, 2020.
- [14] C. Huang, G. Tian, Y. Lan et al., "A new pulse coupled neural network (PCNN) for brain medical image fusion empowered by shuffled frog leaping algorithm," *Frontiers in Neuroscience*, vol. 13, 2019.
- [15] S. H. Wang, C. Tang, J. Sun et al., "Multiple Sclerosis identification by 14-layer convolutional neural network with Batch normalization, Dropout, and stochastic pooling," *Frontiers in Neuroscience*, vol. 12, 2018.
- [16] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi, "Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images," *NeuroImage: Clinical*, vol. 22, 2019.
- [17] M. Zhou, C. Tian, R. Cao et al., "Epileptic seizure detection based on EEG signals and CNN," *Frontiers in Neuroinformatics*, vol. 12, 2018.
- [18] A. L. Goldberger, L. A. N. Amaral, L. Glass et al., "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [19] L. Orosco, A. G. Correa, P. Diez, and E. Laciari, "Patient non-specific algorithm for seizures detection in scalp EEG," *Computers in Biology and Medicine*, vol. 71, pp. 128–134, 2016.
- [20] W. Hang, W. Feng, R. Du et al., "Cross-subject EEG signal recognition using deep domain Adaptation network," *IEEE Access*, vol. 7, pp. 128273–128282, 2019.
- [21] K. Akyol, "Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection," *Expert Systems with Applications*, vol. 148, p. 113239, 2020.
- [22] X. Zhang, L. Yao, M. Dong, Z. Liu, Y. Zhang, and Y. Li, "Adversarial representation learning for robust patient independent epileptic seizure detection," *IEEE Journal of Biomedical and Health Informatics*, p. 1, 2020.
- [23] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, 2019.
- [24] N. J. Stevenson, K. Tapani, L. Lauronen, and S. Vanhatalo, "A dataset of neonatal EEG recordings with seizure annotations," *Scientific Data*, vol. 6, no. 1, 2019.
- [25] C. Huang, Y. Zhou, X. Mao et al., "Fusion of optical coherence tomography and angiography for numerical simulation of hemodynamics in bioresorbable stented coronary artery based on patient-specific model," *Computer Assisted Surgery*, vol. 22, no. sup1, pp. 127–134, 2017.
- [26] H. C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN Architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [27] Y.-D. Zhang, Z. Dong, X. Chen et al., "Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 3613–3632, 2019.



## Research Article

# A Fast Subpixel Registration Algorithm Based on Single-Step DFT Combined with Phase Correlation Constraint in Multimodality Brain Image

Jianguo Li <sup>1</sup> and Quanhai Ma <sup>2</sup>

<sup>1</sup>People's Hospital of Xinjiang Uygur Autonomous Region, Xinjiang 830002, China

<sup>2</sup>Xinjiang Uygur Autonomous Region Hospital of Traditional Chinese Medicine, Xinjiang 831199, China

Correspondence should be addressed to Quanhai Ma; lvxiaoyi@xju.edu.cn

Received 19 January 2020; Accepted 4 February 2020; Published 7 May 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Jianguo Li and Quanhai Ma. 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.

Multimodality brain image registration technology is the key technology to determine the accuracy and speed of brain diagnosis and treatment. In order to achieve high-precision image registration, a fast subpixel registration algorithm based on single-step DFT combined with phase correlation constraint in multimodality brain image was proposed in this paper. Firstly, the coarse positioning at the pixel level was achieved by using the downsampling cross-correlation model, which reduced the Fourier transform dimension of the cross-correlation matrix and the multiplication of the discrete Fourier transform matrix, so as to speed up the coarse registration process. Then, the improved DFT multiplier of the matrix multiplication was used in the neighborhood of the coarse point, and the subpixel fast location was achieved by the bidirectional search strategy. Qualitative and quantitative simulation experiment results show that, compared with comparison registration algorithms, our proposed algorithm could greatly reduce space and time complexity without losing accuracy.

## 1. Introduction

Medical image registration technology is a widely used image processing technology in the field of medicine image analysis [1]. It plays an important role in human 3D modeling, multisource medical image fusion, the lesion feature detection and extraction, and other auxiliary diagnoses [2]. Brain medical diagnosis has a high demand for accuracy, and brain CT image registration technology is the key technology to determine the accuracy and speed of brain diagnosis and treatment [3]. The research motivation of this article is shown in Figure 1.

Different forms of images express different information and different functions. Combining the two can simultaneously express information from many aspects of the human body in one image. The internal structure and function of the human body can be reflected through the image, providing intuitive human anatomy, physiology, and pathology information. At this time, the image configuration

technology needs to solve the problem of position registration of fusion between images. When there is moderate noise in the image and there is translation and scaling between the multimodality images, phase correlation image registration technology is an effective method for subpixel image registration. This paper proposes an improved algorithm based on Guizar-Sicairos registration, which can quickly search for the offset between registered images and greatly reduce the time and space complexity of registration without losing the registration accuracy.

Medical image registration can be divided into single-mode image registration and multimode image registration from the imaging mode. Single modality means that multiple images to be registered are acquired by the same imaging technology, and multimodality means that the registered images are acquired by different imaging technologies [4]. Since the imaging principles of different imaging modes are different, the images they acquire have different characteristics, and the sensitivity to different tissues is also

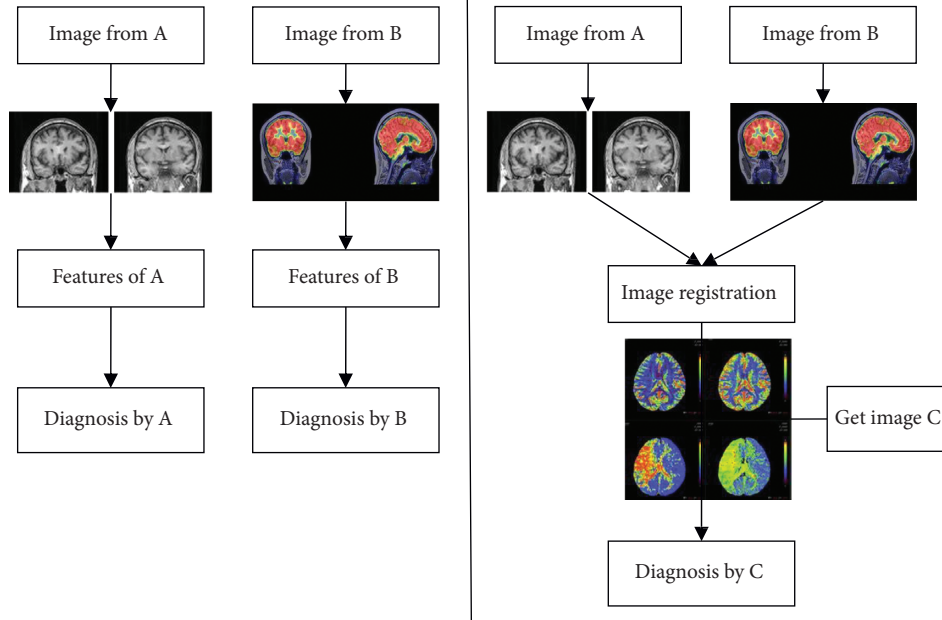


FIGURE 1: Motivation for this study.

different. Therefore, the object information of different modals is also different [5]. In order to help doctors better understand the disease situation, it is necessary to fuse a variety of multimodality images combined with more information so as to make diagnoses. For example, low-quality US images captured in real time during surgery and high-quality CT (or MRI) images before surgery are used to balance the accuracy and real time required clinically during computer-assisted intervention [6]. However, the premise of multimodal image fusion is to register multimodal images, so multimodal image registration technology is one of the research hotspots in medical image processing and is widely used in modern computer-aided interventional medicine [7]. In clinical applications, increasing clinical needs and image imaging modes have injected new impetus into multimodal medical image registration [8].

In the field of medical analysis, image registration achieves subpixel registration accuracy, but there is a huge challenge, especially in multimodality image registration [9]. In order to effectively register multimodal images, subpixel accurate feature-point location is needed. The inevitable error of the actual acquisition system makes it impossible to obtain completely accurate features of the two images, and the corresponding images are offset, stretched, and rotated in spatial domain. The research result shows that when the deviation of two complex images exceeds 0.1 pixels, the quality of the corresponding feature points will be seriously affected, which seriously affects the registration accuracy. Therefore, the high-precision registration of two multimodalities images is the first step of medical image analysis. We need to further study the key issues of multimodality registration algorithms, find the optimal algorithms that are more in line with clinical development, and further improve them reasonably.

Nowadays, typical image registration methods include registration algorithm based on control points, registration algorithm based on image features, and registration algorithm based on regional cross-correlation [10]. In particular, the registration method based on regional cross-correlation is to use cross-correlation technology to get the relative offset between image pairs, which has good robustness. Since the regional cross-correlation algorithm based on spectral operation improves the efficiency of registration, it is a common fast regional cross-correlation algorithm. Therefore, the phase correlation-based registration method has been widely studied due to its advantages of high accuracy, low computational complexity, small amount of computation, strong antinoise, and optical blur invariance [11]. The mainstream model of brain CT image registration is based on the maximum mutual information as a measurement and combined with efficient and high-precision optimization search algorithm. Because of the large amount of mutual information calculation and the slow speed of registration, the performance of the optimization search algorithm has a great impact on the efficiency and accuracy of registration. As for multimodality brain CT images, it is an effective way to improve the speed and accuracy of registration.

Phase correlation method includes spatial domain-based phase correlation and frequency domain-based phase correlation. The early correlation-registration algorithm mainly uses image translation parameters, and its image registration accuracy can achieve pixel level. Then, on this basis, Fourier-Mellin transform is used to expand image registration to the case of rotation, translation, and scaling, but its registration accuracy is also only pixel level [12]. Chen et al. proposed Fourier transform based on matrix multiplication, which can be used for subpixel registration of multimodality image [13]. Its accuracy is

better than that of the traditional subpixel registration method, but the processing efficiency of large-scale multimodality medicine data is not high because the calculation efficiency of this improved method is obviously low in the process of pixel-level displacement. In order to solve the problem of mismatch (inappropriate problem) between registration accuracy and computational complexity, Claus et al. proposed a novel and effective image registration algorithm, which has the same registration accuracy as the standard fast Fourier transform and is considered to be one of the most reliable algorithms in the image registration algorithm based on the phase correlation method. For ease of description, the registration algorithm is called as SSDFT for the image registration algorithm based on the single-step discrete Fourier transform [14].

In this paper, we improve the performance of single-step discrete Fourier transform by reducing the dimension of Fourier transform cross-correlation matrix and the number of DFT matrix multiplication used to locate the peak value. Compared with other studies, the innovation of this study is summarized as follows:

- (1) Compared with single-step discrete Fourier transform (SSDFT), the improved algorithm proposed in this paper can quickly search the offset between the registered images
- (2) The method proposed enhances the registration performance of single-step discrete Fourier transform and greatly reduces the time and space complexity of registration without losing the registration accuracy
- (3) The algorithm proposed is robust and insensitive to noise.

The organization structure of this paper is as follows: Section 2 describes the single-step DFT registration algorithm in detail and introduces the multimodality medical image registration theory and implementation process; Section 3 proposes our improved multimodality registration algorithm, and the process of image registration is given in Section 4. Section 5 selects the different multimodality images with manual transformation to test and verify our proposed registration algorithm and gives the simulation results and analysis; we summarize the whole paper in the last section.

## 2. Materials and Methods

**2.1. Multimodality Brain Image Registration.** Image registration refers to comparing and matching two images  $F(x)$  and  $R(y)$  obtained at different times or under different conditions. According to the spatial transformation relationship obtained from the corresponding point location information of the two images, we can define a similarity measurement function that maximizes the similarity between the two images after the spatial transformation. In other words, each point on the image  $F(x)$  has a corresponding unique point on the image  $R(y)$ , and these two

points should be for the same physical space location. The mathematical model of image registration is shown in the following equation:

$$S(T) = S(R(y), F(T(x))), \quad (1)$$

where  $S$  is the similarity measurement function;  $T$  is the transformation space;  $R(y)$  represents the reference image, and  $F(T(x))$  represents a transformed frequency domain image. The main task of image registration is to find the optimal spatial transformation function  $T$  to make  $S$  reach the maximum, which can achieve an exact matching between the registered image and the reference image. The model is written as follows:

$$\hat{T} = \arg \max_T S(T), \quad (2)$$

The process of registration is also the process of solving the global optimal value of the similarity measure function and its corresponding spatial transformation parameters. The search range of the parameters is called as the search space, and the number of parameters is called the degree of freedom in the spatial transformation model. The number of parameters is related to the spatial transformation model, and the degrees of freedom of different transformation models are also different [15]. Take 3D rigid body transformation as an example, spatial transformation matrix can denoted as  $T = (t_x, t_y, t_z, \alpha, \beta, \gamma)$ , where  $t_x, t_y, t_z$  are the displacement offset of the registered image with respect to the three directions of the coordinate axes  $x, y, z$ ;  $\alpha, \beta, \gamma$  represent the rotation angles in three directions around the coordinate axes  $x, y, z$ .

The calculation of image registration parameters can be roughly divided into two types: image registration parameters based on grayscale and image registration parameters based on features. Gray-based image registration uses the gray data for registration, which can effectively avoid errors caused by feature extraction [16]. A predefined registration measurement function in the process of medicine image registration is designed to measure the difference between two images and then search for an optimal transformation to maximize the similarity of both images. This method has the characteristics of high accuracy and strong robustness and can achieve automatic registration without preprocessing [17]. Correlation methods and mutual information methods are commonly used at present. The framework of gray-based registration is shown in Figure 2.

Feature-based image registration methods use feature sets extracted from images to establish a correspondence relationship between feature sets [18]. Registration parameter is the key to success, but these methods require manual participation to complete the extraction of image features. Feature-based image registration method is divided into three steps: feature extraction, feature matching, and spatial transformation. Corresponding image features are firstly extracted based on the characteristic of the image, such as corners, edges, and curvature [19]. Secondly, a matching algorithm is used to match the corresponding features between the registered image and the reference image. Finally, the best matching function of the two images is achieved by changing the transformation parameters between the

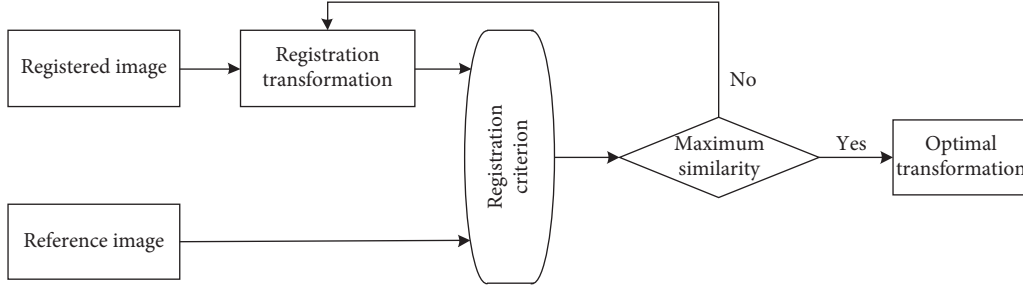


FIGURE 2: The framework of gray-based registration.

registered image and the reference image. The basic steps of the feature-based image registration method are shown in Figure 3.

In feature-based image registration methods, corner points are often used to express features of medicine image. A corner point is a point where the gray level of an image changes drastically and can also be defined as the intersection of two edges in an image. Corners have the advantages of stability and abundant information. In addition, their corners also have the advantages of rotation invariance, affine invariance, and scale invariance, which is very suitable for matching medical images with different modalities [20].

**2.2. Single-Step Discrete Fourier Transform.** Correlation algorithm is a statistics-based registration [5]. It is assumed that the time interval between the two images taken successively is small enough, and there is only a small linear displacement between the images [21]; suppose that the gray distribution function of the two images is  $f_1$  and  $f_2$ , and  $d_x$  and  $d_y$  are displacement offsets of  $f_2$  relative to  $f_1$  in  $x$  and  $y$  axes, respectively. The normalized mean square error (NRMSE) [22] between  $f_1$  and  $f_2$  can be expressed as follows:

$$E^2 = \min_{\alpha, x_0, y_0} \frac{\sum_{x,y} |\alpha g(x - x_0, y - y_0) - f(x, y)|^2}{\sum_{x,y} |f(x, y)|^2} \quad (3)$$

$$= 1 - \frac{\max |R_{f_1 f_2}(x_0, y_0)|}{\sum_{x,y} |f_1(x, y)|^2 \sum_{x,y} |f_1(x, y)|^2},$$

where  $r_{f_1 f_2}$  represents the cross-correlation coefficient between  $f_1$  and  $f_2$ . The cross-correlation value is defined as follows:

$$R_{f_1 f_2}(d_x, d_y) = \sum_{n \times v} f_1 * f_2$$

$$= \sum_{n \times v} \hat{F}_1(n, w) \hat{F}_2^*(n, w) \cdot \exp \left[ i 2\pi \left( \frac{d_x n}{M_1} + \frac{d_y w}{M_2} \right) \right], \quad (4)$$

where  $\hat{F}_1(v, w)$  is the Fourier transform of image  $f_1$ . In order to get the accurate peak position, the traditional

subpixel image registration algorithm has huge storage cost and time consumption because of scaling the image and processing the whole upsampling matrix. In order to overcome this performance limitation, there are two steps proposed by Guizar-Sicairos algorithm [23] to improve the efficiency of registration: (1) the fast Fourier transform (FFT) with the upsampling coefficient  $\epsilon_0 = 2$  is used to calculate the cross-correlation surface peak coordinates between images so as to obtain the initial subpixel motion estimation; (2) then, the accurate peak value is searched in a small window area near the initial estimation. Through the discrete Fourier transform of the small window area with the initial estimation as the center and its size as  $(1.5\epsilon, 1.5\epsilon)$ , an upsampling cross-correlation surface is obtained, without the need of zero-padding to the product. For the realization of this process, equation (1) can also be rewritten as the matrix product of size  $(1.5\epsilon, M_1)$ ,  $(M_1, M_2)$ , and  $(M_2, 1.5\epsilon)$ , respectively. Therefore, the peak location can be found on the result matrix with size  $(1.5\epsilon, 1.5\epsilon)$ . It can be seen that, compared with the conventional FFT image registration method, the computational complexity of this method is greatly improved.

### 3. The Improved Single-Step DFT for Brain Image Registration

Although Guizar-Sicairos algorithm is a novel and fast algorithm for subpixel registration, its main disadvantage is that most of the registration time is spent in the first step to find the initial estimation. To solve this problem, this paper improves the Guizar-Sicairos registration algorithm to reduce the time cost. Therefore, the improved algorithm is to reduce the time of initial estimation of peak position and the time of accurate registration. The framework of brain image registration is shown in Figure 4. Due to the consistency of neighborhood structures in multimodal brain images, the directions around key points where gray values change severely are used as dominant orientations. To adapt for multimodality registration, the SURF descriptor is modified according to the gradient reversals. Due to the great difference between different brain images, the existing algorithm can achieve pixel-level registration. For example, literature [8] makes full use of the flexibility of NSCT for image decomposition and the accuracy of SURF for feature location, as well as the quickness of SURF for feature extraction. The main work of this paper is to design a novel and

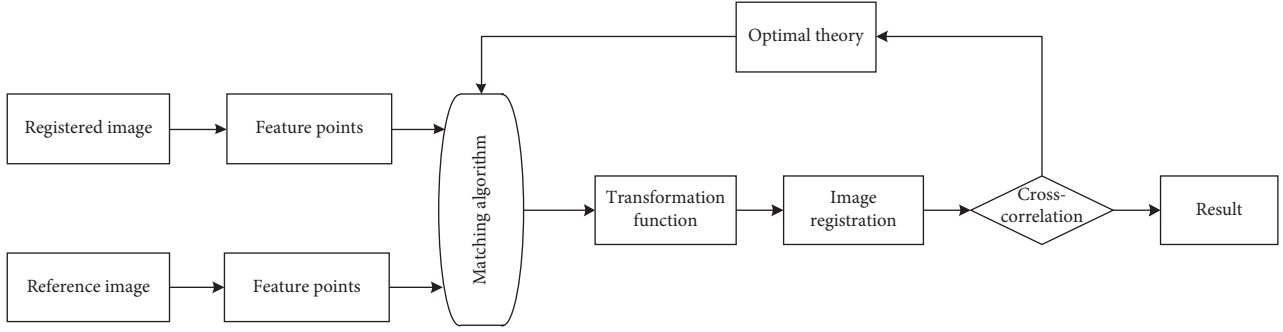


FIGURE 3: The framework of feature-based registration.

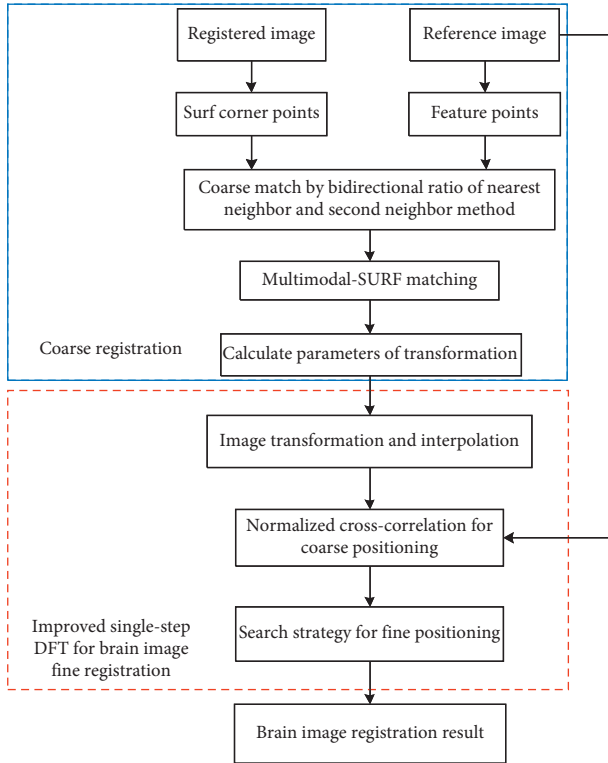


FIGURE 4: The framework of brain image registration.

fast subpixel large-scale translation image registration algorithm. Therefore, rough registration is represented by a blue box, and our proposed method is located in a red box.

Assuming that  $F_1(u, v)$  and  $F_2(u, v)$  are corresponding Fourier transforms, the following equation can be obtained:

$$F_2(u, v) = e^{-j2\pi(ux_0+vy_0)} F_1(u, v). \quad (5)$$

The cross-power spectrum of two images is defined as follows:

$$P(u, v) = \frac{F_1(u, v)F_2^*(u, v)}{|F_1(u, v)F_2^*(u, v)|} = e^{j2\pi(ux_0+vy_0)}. \quad (6)$$

The translation theorem of Fourier transform ensures that the phase of cross-power spectrum between images is equal to the phase difference between images. If the inverse discrete Fourier transform is applied to  $P(u, v)$  in frequency

domain, then the unit pulse function will be obtained at  $(x_0, y_0)$ :

$$F^{-1}\left(e^{j2\pi(ux_0+vy_0)}\right) = \delta(x_0, y_0). \quad (7)$$

It can be seen that the amplitude of the transformed surface is almost zero except for the amplitude at  $(x_0, y_0)$ , so it can be used to measure the translation between two images. Since the normalization is adopted in formula (7), the phase correlation has excellent anti-interference characteristics. After the above analysis, the right side of formula (4) can be rewritten as follows:

$$R_{f_1 f_2} = \sum_{v, \omega} \sum_{n, m} \hat{F}_1(v, \omega) \hat{F}_2^*(v, \omega) \delta(v - K_1 m) \delta(\omega - K_2 n) \cdot \exp\left[i2\pi\left(\frac{d_x v}{M_1} + \frac{d_y \omega}{M_2}\right)\right], \quad (8)$$

where  $K_1$  and  $K_2$  are the downsampling coefficients along the  $x$  direction and  $y$  direction, respectively, and  $\delta$  is the Kronecker impulse function. Taking advantage of the selection feature of function  $\delta$ , thus equation (5) can be written as follows:

$$R_{f_1 f_2} = \sum_{n, m} \hat{F}_1(K_1 m, K_2 n) \hat{F}_2^*(K_1 m, K_2 n) \cdot \exp\left[i2\pi\left(\frac{d_x v}{M_1/K_1} + \frac{d_y \omega}{M_2/K_2}\right)\right]. \quad (9)$$

It can be seen that the equation represents the cross-correlation of the images after dimension reduction and sampling, where the size of the cross-correlation matrix before and after sampling is  $M_1 \times M_2$  and  $M_1/K_1 \times M_2/K_2$ , respectively. By comparing equation (4) and equation (6), it can be inferred that as long as the peak value falls within the downsampling cross-correlation matrix, the peak positions in the matrix are the same. Therefore, the key idea of the improved algorithm in this paper is to do downsampling of the Fourier transform for the two registered images and then use the same Guizar-Sicairos method to find the initial estimation of the peak position. The complexity of the time complexity of the improved registration algorithm is  $O(K_1^{-1}K_2^{-1}M_1M_2\varepsilon)$ .

In order to obtain accurate reconstruction based on downsampling in frequency domain, the proposed



registration algorithm in this paper needs to calculate an overlapping form of cross-correlation. Because the cross-power spectrum performs  $K$  downsampling in two directions and the cross-correlation time spectrum is located near  $M/K$  in spatial domain, the overlap of cross-correlation function may change the initial peak position, resulting in the wrong estimation of the downsampling matrix. In order to solve this problem, the improved algorithm in this paper normalizes the original cross-correlation function before downsampling, and its expression can be denoted as follows [8]:

$$R_{f_1 f_2}(d_x, d_y) = e^{i2\pi(vd_x/M_1 + \omega d_y/M_2)}. \quad (10)$$

It can be seen that the inverse Fourier transform of the complex exponential is a delta function, which is to say that the position of the single peak is related to the offset  $(d_x, d_y)$ . In order to reduce the edge effect, a window function can be used in the Fourier transform to intercept, or zero-padding. Therefore, the peak curve obtained by the cross-correlation function is independent of the image. As long as the peak position in the original normalized cross-correlation before downsampling is less than  $M/K$ , the overlapping cross-correlation will not change the initial peak position. Experimental results show that the cross-correlation surface has an obvious peak value, but the peak value surface is very broad, not much larger than other peaks. The cross-correlation surface has a sharp peak value. It means these local peaks may be considered as the position of global peaks in the wrong position area. Therefore, using the inverse Fourier transform of normalized cross-power spectrum, the matching performance will be very accurate and stable even in the presence of noise.

In order to correctly locate the peak value in the overlapping regularization phase correlation matrix, it is assumed that the subpixel translation boundaries along the  $x$ -axis and  $y$ -axis are  $d_x = M_1/K_1$  and  $d_y = M_2/K_2$ , respectively. If the sampling coefficient is 2, it can be obtained the maximum subpixel displacement  $d_x = M_1/2$  and  $d_y = M_2/2$  in two directions. In addition, the upper limits of the sampling coefficient are  $M_1/2$  and  $M_2/2$ , which will make the subpixel displacement less than 2. If these upper limits are not satisfied, the initial peak value will fall outside the boundary  $M_1/K_1 \times M_2/K_2$  of the original cross-power spectrum. It will make the peak value near these values and mistake the surface peak position, so this method cannot determine the subpixel displacement robustly.

Through the above theoretical analysis, it can be seen that the improved algorithm proposed in this paper mainly uses the downsampling cross-correlation function to solve the overlapping model and reduces the Fourier transform dimension of the cross-correlation matrix and the multiplication number of the discrete Fourier transform matrix to speed up the registration process.

#### 4. Brain Image Fine-Search Strategy for Registration

The second step of Guizar-Sicairos method is to search for an accurate peak value in a small window area near the initial estimate. Combined with the above analysis, we can achieve

the goal of acceleration by reducing the number of multiplication and addition of complex matrix (cumulative multiplication) needed to locate the peak value in the upsampling cross-correlation window. In addition, according to the derivation of literature [14], the correlation function can be written as the product of three matrices, namely:

$$\begin{aligned} R_{f_2 f_1} &= \sum_{\omega} \left[ e^{i2\pi(d_y \omega/M_2)} \right]_{1.5\epsilon \times M_1} \sum_{\omega} \left[ \hat{F}_1(v, \omega) \times \hat{F}_2^*(v, \omega) \right]_{M_1 \times M_2} \\ &\quad \cdot \sum_v \left[ e^{i2\pi(d_x v/M_1)} \right]_{M_2 \times 1.5\epsilon} \\ &= A_{1.5\epsilon \times M_1} B_{M_1 \times M_2} C_{M_2 \times 1.5\epsilon}. \end{aligned} \quad (11)$$

The aim is to find the subpixel peak in the result matrix  $R_{f_1 f_2}$ . Through the correlation surface analysis, it can be seen that there is a correlation peak in the cross-correlation, and its upsampling form is similar to a parabola shape, and the contour is monotonically increasing. Therefore, the improved algorithm adopts a forward and backward search strategy to reduce the total number of multiplication and addition operations of complex matrix, and its search process is referred to literature [4].

The main steps of this improved algorithm are as follows: (1) coarse positioning: the algorithm uses the improved strategy proposed in Section 3 to calculate the peak coordinates of cross-correlation surfaces between images and obtain the pixel-level translation  $(x, y)$ ; (2) fine positioning: the DFT of improved matrix multiplication and search strategy are used to obtain the  $n$ -times sampled and neighborhood area  $1.5 \times 1.5$  of the coarse positioning point  $(x, y)$ , and the pixel-level translation  $(\bar{x}, \bar{y})$  is obtained by calculating the phase correlation of the upsampling area. Considering the resampling multiple  $n$ , the subpixel-level translation  $(\bar{x}/n, \bar{y}/n)$  is obtained. Therefore, the translation  $\Delta = (x_1, y_1)$  combined with phase correlation and resampled image registration is written as follows:

$$\begin{cases} x_1 = x + \frac{\bar{x}}{n}, \\ y_1 = y + \frac{\bar{y}}{n}. \end{cases} \quad (12)$$

In this paper, phase correlation is used to obtain coarse position points in the original image and fine position point is obtained after resampling  $n$ -times. Because the phase correlation-based coarse positioning has the pixel-level accuracy, setting the fine positioning area as the size of  $1.5 \times 1.5$  with the coarse positioning point as the center can ensure the subpixel-level accurate positioning point is in this area. In order to obtain high positioning accuracy and not increase too much calculation cost, the inverse Fourier transform will be carried out in the region with the size of  $150 \times 150$  to obtain the fine positioning peak value if the upsampling times  $n$  is taken as 100. In other words, the positioning accuracy will reach 0.001 pixels.

It can be seen from the theoretical analysis that the improved fine-searching process in this paper will also reduce the calculation time for large-scale multimode brain imaging registration. However, both the Guizar-Sicairos algorithm and its improved algorithms use the same inverse DFT matrix, and most of the calculation time is spent on the generation of the matrix, rather than looking for the peak value. Therefore, the proposed improvement algorithm is limited to large-scale multimode brain imaging registration, but its calculation time will not exceed the original registration algorithm.

## 5. Simulation Experiment and Results

**5.1. Experimental Data.** This experiment uses the brain PET/CT data provided by the Second Affiliated Hospital of China Medical University (CMU) to experimentally test our proposed registration algorithm. The data obtained by PET/CT imaging equipment include a PET image volume data and a CT image volume data. The specific information of the experimental data is shown in Table 1. In addition, this experiment also applies the registration algorithm to the brain data from Vanderbilt University and will give the registration result so as to verify the accuracy of the proposed algorithm. In order to verify the registration accuracy of multimodal data acquired in the real environment, the CT brain image data acquired from a patient in Guangzhou General Military Hospital (GMH) are also selected, and the registration results are given, where the image sample is shown in Figure 5. The resolution, pixel size, and grayscale range of each image data are shown in Table 1.

**5.2. Evaluation Index.** In order to measure the registration accuracy well in the experiment, two completely aligned different modal images are selected. We adopted a random deformation operation to one image and then take the deformed image as a floating image (registered image), and the other is a reference image. The introduced random deformation operation is the gold standard to test whether the deformation obtained by registration is accurate. We measure the accuracy of image registration by calculating the difference between the gold standard and the registration estimation. In our experiment, the target registration error (THE) is the valuation index, and its definition is as follows:

$$\text{THE} = \sqrt{\frac{1}{|\Omega_I|} \sum_{x_i \in \Omega_I} (I_1(x_i) - I_2(x_i))^2}, \quad (13)$$

where  $I_1$  and  $I_2$  are two images that need to calculate THE;  $I_1(x)$  and  $I_2(x)$  are the gray values of the same points in two images;  $\Omega_I$  is the image domain of  $I$ ; and  $|\Omega_I|$  is the number of pixels. THE measures the similarity of the two images by calculating the root mean square error of gray value. The smaller the THE value of the reference image and the registered image is, the closer the gray value of the same point is in the two images, the more similar the whole image is, and the better the registration effect is.

**5.3. Experiment and Results.** In order to verify the effectiveness of our proposed registration algorithm for multimodality brain images, this paper performs experimental simulation on MATLAB 7.6 platform. To fairly test the performance of the improved image registration algorithm in this paper, the Guizar-Sicairos algorithm [23], the RSTT registration algorithm proposed by Foroosh and Berci [21], the normalized cross-correlation algorithm (NCC), and the Fast-RST methods proposed by Zhou et al. [24] were selected. This paper uses the medicine images with different sizes as simulation images, and images with different scales are implemented by interpolation. In this paper, the registration accuracy of the Guizar-Sicairos method and the improved algorithm proposed in this paper are both 0.01 pixels. The calculation time required for the initial coarse peak estimation and fine-search step is shown in Figures 6 and 7. It can be seen that compared with the Guizar-Sicairos method, our proposed registration algorithm in this paper greatly reduces the time to obtain the initial peak position.

For images with different sizes, the sampling coefficient is set to  $M/32$ . If the image size is not considered, the time required for peak estimation is roughly the same as the time required for a matrix of size  $32 \times 32$ . For large-size images with a size of  $512 \times 512$ , the estimated time is close to 2 milliseconds, while the traditional Guizar-Sicairos method requires about 200 seconds. During the fine registration process, our improved method in this paper improves the performance of the Guizar-Sicairos algorithm, as shown in Figure 7. In the entire simulation experiment, the step size  $\sigma$  is set to  $0.3\epsilon$ , and the performance comparison results of all algorithms are shown in Figure 8.

Overall, the registration time of our proposed registration method is faster than that of other comparison methods. In addition, it can be seen from the experimental results that the registration error of the Guizar-Sicairos method is 0.000471, and the error of our improved method in this paper is 0.000460. The experimental process also shows that our algorithm in this paper can adapt to various translation situations. It only needs to adaptively calculate the sampling step in the initial positioning step, while the improved algorithm in [3] can only process weak translation. In order to quantitatively evaluate the accuracy of the registration algorithm, some evaluation criterion is designed for quantitative analysis. Table 2 and Figure 9 show the effectiveness under different registration algorithms. According to the definition of target registration error (THE), if the evaluation index is close to 1, this registration algorithm is the most accurate [20]. In addition, by changing the required subpixel accuracy, this paper also tests the registration accuracy and calculation time of different comparison algorithm, as shown in Table 3 and Figure 10, where the image size is  $512 \times 512$ . Experimental results show that our improved algorithm in this paper will greatly reduce the time and space complexity for multimodality brain image registration and obtain the same subpixel accuracy as the original Guizar-Sicairos algorithm.

TABLE 1: Image characteristic for different modalities.

Data sources	Modality	Resolution	Pixel size	Grayscale
Vanderbilt	CT	$512 \times 512 \times 15$	0.415/0.415/6	0~4096
	MRI	$512 \times 416 \times 19$	0.4462/0.4492/5	0~1149
CMU	CT	$512 \times 512 \times 28$	0.6535/0.6535/4	-32,768~32767
	PET	$128 * 128 * 16$	2.59072/2.59072/8	-32,768~32767
GMH	CT	$512 \times 512 \times 15$	0.4151/0.4151/6	0~4096
	MRI	$512 \times 416 \times 19$	0.4462/0.4492/5	0~1149

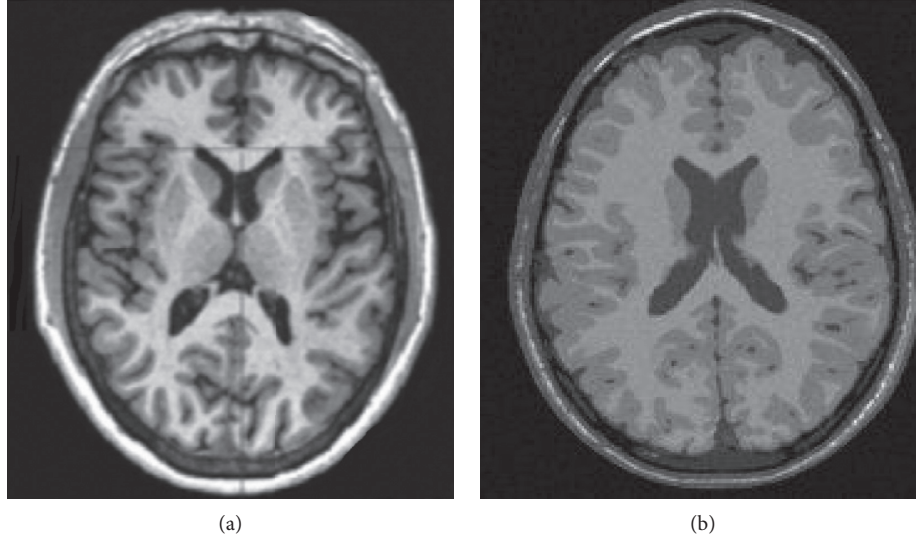


FIGURE 5: Image sample for different modalities.

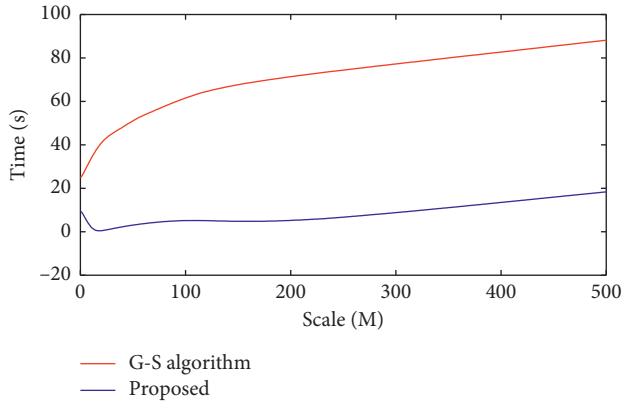


FIGURE 6: Running time for initial estimation.

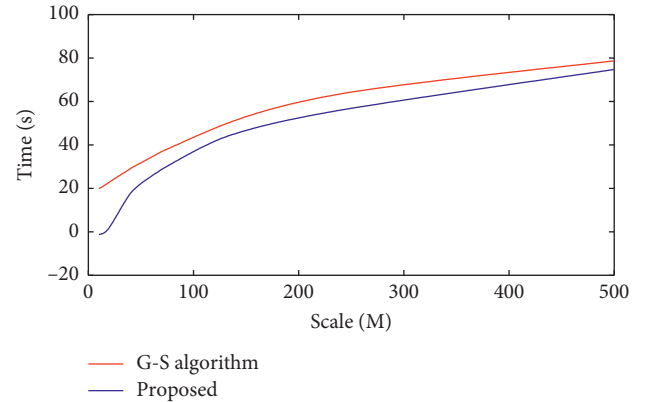


FIGURE 7: Running time for fine registration.

**5.4. Discussion.** Although Guizar-Sicairos algorithm is a novel and fast algorithm for subpixel registration, its main disadvantage is that most of the registration time is spent in the first step to find the initial estimation. To solve this problem, this paper improves the Guizar-Sicairos registration algorithm to reduce the time cost. Therefore, our improved algorithm is to reduce the time of initial estimation of peak position and the time of accurate registration. Due to the consistency of neighborhood structures in multimodal brain images, the directions around key points where gray values change severely are used as dominant orientations. To

adapt for multimodality registration, the SURF descriptor is modified according to the gradient reversals. Due to the great difference between different brain images, the existing algorithm can achieve pixel-level registration. For example, literature [8] makes full use of the flexibility of NSCT for image decomposition and the accuracy of SURF for feature location, as well as the quickness of SURF for feature extraction.

Multimodality brain image registration technology is the key technology to determine the accuracy and speed of brain diagnosis and treatment. In order to achieve high-precision

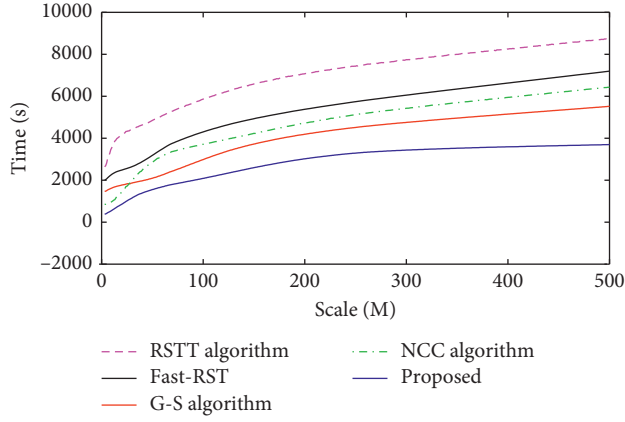


FIGURE 8: Comparison of total time for different algorithms.

TABLE 2: Comparison of registration accuracy for different algorithms.

Proposed	G-S algorithm	RSTT	Fast-RST	NCC
0.996	0.989	0.981	0.893	0.861

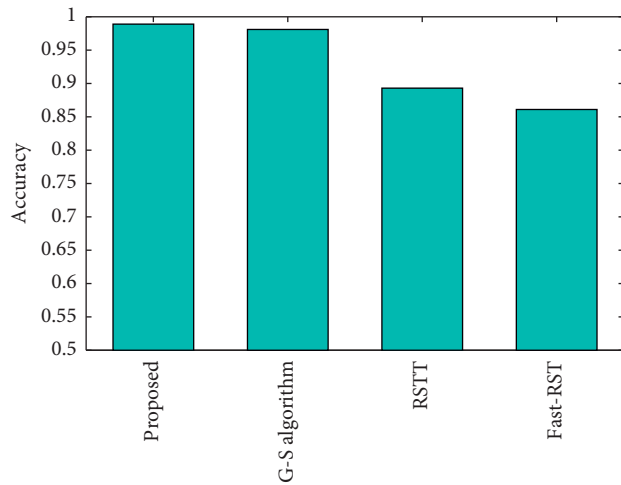


FIGURE 9: Schematic comparison of registration accuracy of each algorithm.

image registration, a novel and fast subpixel large-scale translation image registration algorithm was proposed. Firstly, the coarse positioning at the pixel level was achieved by using the downsampling cross-correlation model, which reduced the Fourier transform dimension of the cross-correlation matrix and the multiplication of the discrete Fourier transform matrix, so as to speed up the coarse registration process. Then, the improved DFT multiplier of the matrix multiplication was used in the neighborhood of the coarse point, and the subpixel fast location was achieved by the bidirectional search strategy. Simulation experiment results show that, compared with common image registration algorithms, our proposed algorithm could greatly reduce space and time complexity without losing accuracy.

TABLE 3: Comparison results between total time and registration accuracy for different algorithms.

Accuracy	G-S algorithm	Proposed
0.1	0.547	0.157
0.01	0.631	0.291
0.001	2.989	1.489
0.0001	151.560	75.130

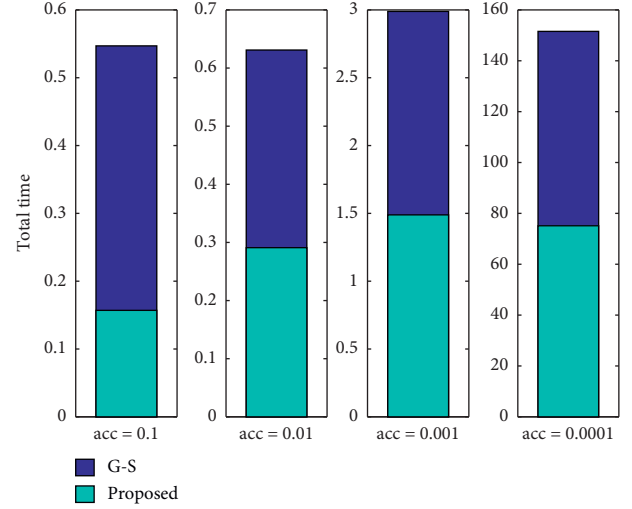


FIGURE 10: Schematic comparison of the total running time of each algorithm and registration accuracy.

## 6. Conclusion

When there is moderate noise in the image and there is translation and scaling between the multimodality images, phase correlation image registration technology is an effective method for subpixel image registration. This paper proposes an improved algorithm based on Guizar-Sicairos registration, which can quickly search for the offset between registered images and greatly reduce the time and space complexity of registration without losing the registration accuracy. Theoretical analysis and experimental verification show that our proposed multimodality brain image registration algorithm has high matching accuracy and antinoise performance, can be well applied to medicine image registration with big-scale translation, and is suitable for medicine analysis engineering applications.

## Data Availability

All relevant data are within the paper.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## References

- [1] M. L. Uss, B. Vozel, V. A. Dushepa, V. A. Komjak, and K. Chehdi, "A precise lower bound on image subpixel registration accuracy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3333–3345, 2014.



- [2] J. B. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 2009.
- [3] A. Yousef, J. Li, and M. Karim, "High-speed image registration algorithm with subpixel accuracy," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1796–1800, 2015.
- [4] L. K. Arata, A. P. Dhawan, and J. P. Broderick, "Three-dimensional anatomical model-based segmentation of MR brain images through principal axes registration," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 11, pp. 1069–1078, 1995.
- [5] S. Liao and A. C. S. Chung, "Feature based nonrigid brain MR image registration with symmetric alpha stable filters," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 106–119, 2010.
- [6] D. Fan, E. Shen, L. Li et al., "Small baseline stereo matching method based on phase correlation," *Journal of Geomatics Science and Technology*, vol. 30, pp. 154–157, 2013.
- [7] X. Tong, Y. Xu, Z. Ye et al., "An improved phase correlation method based on 2-D plane fitting and the maximum kernel density estimator," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1953–1957, 2015.
- [8] H. Foroosh, M. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.
- [9] X. Lin, F. T. Qiu, and S. Ruan, "A topology preserving non-rigid registration algorithm with integration shape knowledge to segment brain subcortical structures from MRI images," *Pattern Recognition*, vol. 43, no. 7, pp. 2418–2427, 2010.
- [10] S. Liao and A. C. S. Chung, "Nonrigid brain MR image registration using uniform spherical region descriptor," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 53–66, 2012.
- [11] P. Bing, X. X. Hui-Min, and D. Fu-Long, "Performance of sub-pixel registration algorithms in digital image correlation," *Measurement Science and Technology*, vol. 17, no. 6, pp. 1615–1621, 2006.
- [12] B. Pan and B. Wang, "Digital image correlation with enhanced accuracy and efficiency: a comparison of two subpixel registration algorithms," *Experimental Mechanics*, vol. 56, no. 8, pp. 1395–1409, 2016.
- [13] W. Chen, Z. Jiang, L. Tang, Y. Liu, and Z. Liu, "Equal noise resistance of two mainstream iterative sub-pixel registration algorithms in digital image correlation," *Experimental Mechanics*, vol. 57, no. 6, pp. 979–996, 2017.
- [14] D. Claus, M. Fritzsche, B. Timmerman, P. Bryanston-Cross, and D. Iliescu, "High-resolution digital holography utilized by the subpixel sampling method," *Applied Optics*, vol. 50, no. 24, pp. 4711–4719, 2011.
- [15] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1185–1199, 2004.
- [16] L. Barazzetti, M. M. Scaioni, and M. Gianinetto, "Automatic co-registration of satellite time series via least squares adjustment," *European Journal of Remote Sensing*, vol. 47, no. 1, pp. 55–74, 2014.
- [17] C.-L. Tsai, W. C. Warger II, G. S. Laevsky, and C. A. Dimarzio, "Alignment with sub-pixel accuracy for images of multi-modality microscopes using automatic calibration," *Journal of Microscopy*, vol. 232, no. 1, pp. 164–176, 2008.
- [18] R. J. Althof, M. G. J. Wind, and J. T. Dobbins, "A rapid and automatic image registration algorithm with subpixel accuracy," *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, pp. 308–316, 1997.
- [19] O. Clatz, H. Delingette, I.-F. Golby et al., "Robust nonrigid registration to capture brain shift from intraoperative MRI," *IEEE Transactions on Medical Imaging*, vol. 24, no. 11, pp. 1417–1427, 2005.
- [20] D. V. Iosifescu, M. E. Shenton, S. K. Warfield et al., "An automated registration algorithm for measuring MRI subcortical brain structures," *NeuroImage*, vol. 6, no. 1, pp. 13–25, 1997.
- [21] H. Foroosh and M. Balci, "Sub-pixel registration and estimation of local shifts directly in the fourier domain," in *Proceedings of the International Conference on Image Processing (ICIP)*, Singapore, October 2004.
- [22] L. Thurfjell, Y. Lau, and B. Hutton, "Improving efficiency of multi-modality registration of brain scans based on mutual information," *Nuclear Medicine Communications*, vol. 20, no. 4, pp. 373–382, 1999.
- [23] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33, no. 2, pp. 156–158, 2008.
- [24] F. Zhou, W. Yang, and Q. Liao, "A coarse-to-fine subpixel registration method to recover local perspective deformation in the application of image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 53–66, 2012.



## Research Article

# A Novel Radial Basis Neural Network-Leveraged Fast Training Method for Identifying Organs in MR Images

Min Xu<sup>1,2,3</sup>, Pengjiang Qian<sup>3,4</sup>, Jiamin Zheng<sup>4</sup>, Hongwei Ge<sup>2</sup>,  
and Raymond F. Muzic Jr.<sup>5</sup>

<sup>1</sup>School of Internet of Things Technology, Wuxi Institute of Technology, Wuxi 214121, China

<sup>2</sup>School of Internet of Things, Jiangnan University, Wuxi 214122, China

<sup>3</sup>Jiangsu Key Lab of Media Design and Software Technology, Jiangnan University, Wuxi 214122, China

<sup>4</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China

<sup>5</sup>Department of Radiology and Case Center for Imaging Research, University Hospitals, Case Western Reserve University, Cleveland, OH 44106, USA

Correspondence should be addressed to Pengjiang Qian; [qianpjiang@jiangnan.edu.cn](mailto:qianpjiang@jiangnan.edu.cn)

Received 29 January 2020; Revised 11 April 2020; Accepted 20 April 2020; Published 5 May 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Min Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a new method for fast organ classification and segmentation of abdominal magnetic resonance (MR) images. Magnetic resonance imaging (MRI) is a new type of high-tech imaging examination fashion in recent years. Recognition of specific target areas (organs) based on MR images is one of the key issues in computer-aided diagnosis of medical images. Artificial neural network technology has made significant progress in image processing based on the multimodal MR attributes of each pixel in MR images. However, with the generation of large-scale data, there are few studies on the rapid processing of large-scale MRI data. To address this deficiency, we present a fast radial basis function artificial neural network (Fast-RBF) algorithm. The importance of our efforts is as follows: (1) The proposed algorithm achieves fast processing of large-scale image data by introducing the  $\epsilon$ -insensitive loss function, the structural risk term, and the core-set principle. We apply this algorithm to the identification of specific target areas in MR images. (2) For each abdominal MRI case, we use four MR sequences (fat, water, in-phase (IP), and opposed-phase (OP)) and the position coordinates  $(x, y)$  of each pixel as the input of the algorithm. We use three classifiers to identify the liver and kidneys in the MR images. Experiments show that the proposed method achieves a higher precision in the recognition of specific regions of medical images and has better adaptability in the case of large-scale datasets than the traditional RBF algorithm.

## 1. Introduction

Magnetic resonance imaging (MRI) is a new type of high-tech imaging examination fashion in recent years. It has the advantages of no ionizing radiation, no bone artifacts, and multidirectional and multiparameter imaging [1]. Therefore, the generation of an end-to-end intelligent disease diagnosis system based on MRI is an inevitable direction for the development of intelligent medicine. To achieve the goal of effective intelligent medical treatment, this paper studies the classification of abdominal organs based on MRI.

There are many techniques for medical image processing [2–5]. Gordillo et al. [6] divided the existing MR image pro-

cessing technologies into the following three categories: The first type is threshold-based methods, which classify the segmentation objects (such as pixels) of the MR image by comparing them with different thresholds [7–9]. The second type is region-based methods, which divide several mutually exclusive regions according to preset rules and then categorize pixels with the same attributes into the same region [10, 11]. The third type is pixel-based classification methods, which mainly classify the objects according to the MR multimodal attributes of each pixel. According to whether the training set is labeled or not, they can be subdivided into unsupervised, semisupervised, and supervised methods [12–14].

Of the third type of methods, an artificial neural network as a supervised learning model is applied to the field of medical imaging [15]. It is suitable for image processing without prior distribution assumptions; its application can be divided into three categories: The first type is to apply artificial neural networks directly to MR image processing. Lucht et al. [16] applied a neural network to the dynamic segmentation of MR breast images. Egmont-Petersen et al. [17] used neural networks and multiscale pharmacokinetic features to segment bone tumors in MR perfusion images. Zhang et al. [18] proposed a visual encoding model based on deep neural networks and transfer learning for brain activity measured by functional magnetic resonance imaging. The second type is to use a convolutional neural network or its improved algorithm to segment MR images [19–22]. Khalilia et al. [20] used convolutional neural networks to automatically perform brain tissue segmentation in fetal MRI. Wang et al. [21] used dynamic pixelwise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI. The third type is to use hybrid neural networks to segment MR images. Glass et al. [23] used a hybrid artificial neural network to segment the inversion recovery image of a normal human brain. Alejo et al. [24] used a hybrid artificial neural network to design an accurate computer-aided method capable of performing region segmentation. Reddick et al. [25] used a hybrid neural network to propose a fully automatic method for segmentation and classification of multispectral MR images.

Based on the review of the above literature, great progress has been made in the use of artificial neural networks for medical image segmentation. However, with the higher resolution requirements of MR images and the increasing size of the dataset, research on fast artificial neural network training for large medical image datasets is still lacking. In response to this phenomenon, this paper proposes the Fast-RBF algorithm, which has fast processing capabilities for large datasets. We applied this method to MRI-based abdominal organ classification and segmentation. The results showed that this method achieved significant results. The main contributions of this paper are as follows:

- (1) The Fast-RBF algorithm with a large-sample-processing capability is proposed by introducing the  $\varepsilon$ -insensitive loss function and structural risk term and using the core-set principle [26]. This method not only retains the strong nonlinear fitting ability and simple learning rules of RBF artificial neural networks but can also process a large dataset quickly, which improves the processing speed and efficiency.
- (2) For each abdominal MRI case, we use four MR sequences (fat, water, IP, and OP) and the position coordinates  $(x, y)$  of each pixel as the input of the algorithm. We use three classifiers to identify the liver, kidneys, and other tissues. The proposed algorithm has better adaptability and runs faster in large dataset scenarios than the traditional RBF neural network algorithm.

The remainder of this paper is divided into four parts: Section 2 introduces RBF neural networks and the relation-

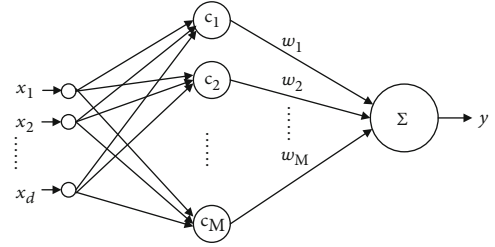


FIGURE 1: The model of an RBF neural network.

ship between RBF neural networks and linear models; Section 3 introduces the Fast-RBF neural network with its large-sample-processing ability; Section 4 verifies the validity of the proposed algorithm on medical image processing; and Section 5 summarizes the full text.

## 2. Related Work

**2.1. RBF Neural Network.** RBF neural networks consist of an input layer, an implicit layer, and an output layer, as shown in Figure 1. Among them,  $\mathbf{x}_i \in R^d, y \in R$ , the number of hidden layer nodes is  $M$ , and the nonlinear mapping  $f : R^d \rightarrow R$  is performed by the RBF neural network.

In an RBF neural network, the input layer receives the training samples; the hidden layer node performs a nonlinear transformation through the radial basis function that maps the input space to a new space. If the radial basis function is defined as a Gaussian function, let  $\mathbf{c}_i \in R^d$  denote the center of the Gaussian function and let  $\delta_i$  represent the kernel width of the Gaussian function. This function can be expressed as

$$\varphi(\|\mathbf{x} - \mathbf{c}_i\|) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_i\|^2}{\delta_i}\right). \quad (1)$$

The nodes of the output layer implement a linear weighted combination in this new space. Let  $w_i$  be the connection weight of the hidden layer and the output layer and  $\varphi(\bullet)$  be the radial basis function; then, the mapping function of  $R^d \rightarrow R$  is

$$y = f(\mathbf{x}) = \sum_{i=1}^M w_i \varphi(\|\mathbf{x} - \mathbf{c}_i\|). \quad (2)$$

**2.2. RBF Neural Network and Linear Model.** According to the introduction above, the RBF neural network has 3 parameters: the center vector of the radial basis function  $\mathbf{c}_i = [c_{i1}, c_{i2}, \dots, c_{id}]^T$ , the kernel width  $\delta_i$ , and the weight of the output layer  $w_i$ . Among them,  $\mathbf{c}_i$  and  $\delta_i$  can be determined by the fuzzy C-means (FCM) clustering algorithm [27], and  $w_i$  is obtained by the gradient descent learning algorithm. Let  $\mu_{ji}$ , which is obtained by the FCM clustering algorithm, denote the fuzzy membership of sample  $\mathbf{x}_j$  for the  $i$ th class,  $n$  represent the size of the training sample, and  $M$  indicate the number of hidden layer nodes; then, the center of the radial basis function  $\mathbf{c}_{ik}$  and the kernel width  $\delta_i$  can be expressed by equations (3) and (4):

$$c_{ik} = \frac{\sum_{j=1}^n \mu_{ji} x_{jk}}{\sum_{j=1}^n \mu_{ji}}, \quad (3)$$

$$\delta_i = \frac{\sum_{j=1}^n \mu_{ji} \|\mathbf{x}_j - \mathbf{c}_i\|^2}{\sum_{j=1}^n \mu_{ji}}. \quad (4)$$

Let  $\tilde{\mathbf{x}}_i = \varphi(\|\mathbf{x} - \mathbf{c}_i\|)$ ,  $i = 1, 2, \dots, M$ .

$$\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_M]^T. \quad (5)$$

The center  $\mathbf{c}_i$  and the kernel width  $\delta_i$  of the radial basis function are obtained by equations (3) and (4), the input sample is mapped to the new space  $f: R^d \rightarrow R^M$ , and the conversion from the input layer to the hidden layer is a non-linear mapping.

Let  $\mathbf{p} = [w_1, w_2, \dots, w_M]^T$ ; then, the neural network function can be expressed as

$$y = \mathbf{p}^T \tilde{\mathbf{x}}. \quad (6)$$

It can be seen from equation (6) that when the radial basis function hidden layer is estimated, the output of the network can be converted into a linear model.

### 3. Fast-RBF Algorithm

**3.1. Fast-RBF Principle.** First, the  $\varepsilon$ -insensitive loss function corresponding to the RBF linear model is introduced. To minimize the value of the  $\varepsilon$ -insensitive loss function,  $\varepsilon$  is solved as the constraint term of the optimization problem. Then, the structural risk term and the Gaussian kernel are introduced to construct the RBF neural network optimization model with large-sample processing. The specific steps are as follows.

*Step 1.* From equations (3) and (4), the values of  $\mathbf{c}_i$  and  $\delta_i$  are obtained; then, from equation (5), the model input  $\tilde{\mathbf{x}}$  is obtained.

*Step 2.* Introducing the  $\varepsilon$ -insensitive loss function.

First, the definition of the  $\varepsilon$ -insensitive loss function is given:

**Definition 1.** The  $\varepsilon$ -insensitive loss function is defined as [28]

$$L^\varepsilon(\mathbf{x}, y, f) = |y - f(\mathbf{x})|_\varepsilon = \max(0, |y - f(\mathbf{x})| - \varepsilon), \quad (7)$$

where  $\mathbf{x} \in R^d, y \in R$ .

For the linear model of equation (6), its corresponding  $\varepsilon$ -insensitive loss function can be expressed as

$$\sum_{i=1}^n |y_i^o - y_i|_\varepsilon = \sum_{i=1}^n \max(0, |y_i^o - y_i| - \varepsilon) = \sum_{i=1}^n \max(0, |\mathbf{p}^T \tilde{\mathbf{x}}_i - y_i| - \varepsilon), \quad (8)$$

where  $y_i^o$  represents the neural network output value and  $y_i$  represents the real output value.

For equation (8), the constraints of  $\mathbf{p}^T \tilde{\mathbf{x}}_i - y_i < \varepsilon$  and  $y_i - \mathbf{p}^T \tilde{\mathbf{x}}_i < \varepsilon$  are not always satisfied, so the relaxation factors  $\xi_i$  and  $\xi_i^*$  are introduced, and the following constraints can be obtained:

$$\begin{cases} y_i - \mathbf{p}^T \tilde{\mathbf{x}}_i < \varepsilon + \xi_i, & \xi_i \geq 0, \\ \mathbf{p}^T \tilde{\mathbf{x}}_i - y_i < \varepsilon + \xi_i^*, & \xi_i^* \geq 0. \end{cases} \quad (9)$$

The purpose of this algorithm is to minimize the value of the  $\varepsilon$ -insensitive loss function represented by equation (8). The value of the  $\varepsilon$ -insensitive parameter will directly affect the accuracy of the modeling. Therefore, the parameter  $\lambda$  is introduced, and  $\varepsilon$  is used as the constraint term in the optimization problem. Combined with equation (9), the optimization problem can be expressed equivalently as

$$\min 2\lambda\varepsilon + \frac{\lambda}{\mu n} \sum_{i=1}^n (\xi_i^2 + \xi_i^{*2}), \quad \text{s.t.} \begin{cases} y_i - \mathbf{p}^T \tilde{\mathbf{x}}_i < \varepsilon + \xi_i, \\ \mathbf{p}^T \tilde{\mathbf{x}}_i - y_i < \varepsilon + \xi_i^*, \end{cases} \quad (10)$$

where the parameter  $\mu$  is the balance factor and  $\xi_i, \xi_i^* \geq 0$  is automatically satisfied.

*Step 3.* Introducing structural risk items and kernel functions.

A support vector machine is an implementation of the principle of structural risk minimization [28]; the method proposed in this paper learns the implementation method of the support vector machine and introduces a regularization term to minimize the risk in the algorithm structure. The kernel method is an important component of a support vector machine [28], which is used to improve the computational ability of the linear learner. The method proposed in this paper also introduces a kernel function. After introducing the regular term and the kernel function, the optimization problem can be expressed by

$$\min_{\mathbf{p}, \varepsilon, \xi_i, \xi_i^*} \|\mathbf{p}\|^2 + 2\lambda\varepsilon + \frac{\lambda}{\mu n} \sum_{i=1}^n (\xi_i^2 + \xi_i^{*2}), \quad \text{s.t.} \begin{cases} y_i - \mathbf{p}^T \varphi(\tilde{\mathbf{x}}_i) < \varepsilon + \xi_i, \\ \mathbf{p}^T \varphi(\tilde{\mathbf{x}}_i) - y_i < \varepsilon + \xi_i^*, \\ i = 1, 2, \dots, n. \end{cases} \quad (11)$$

*Step 4.* Formula derivation.

By introducing the Lagrange multiplier, the Lagrangian function of equation (11) can be expressed as

$$\begin{aligned} L = & \|\mathbf{p}\|^2 + 2\lambda\varepsilon + \frac{\lambda}{\mu n} \sum_{i=1}^n (\xi_i^2 + \xi_i^{*2}) \\ & + \sum_{i=1}^n \alpha_i (y_i - \mathbf{p}^T \varphi(\tilde{\mathbf{x}}_i) - \varepsilon - \xi_i) \\ & + \sum_{i=1}^n \alpha_i^* (\mathbf{p}^T \varphi(\tilde{\mathbf{x}}_i) - y_i - \varepsilon - \xi_i^*). \end{aligned} \quad (12)$$

The matrix form of the corresponding dual problem of equation (12) is

$$\left\{ \max [\alpha^T \alpha^{*T}] \begin{bmatrix} \frac{2}{\lambda} y \\ -\frac{2}{\lambda} y \end{bmatrix} - [\alpha^T \quad \alpha^{*T}] \tilde{\mathbf{K}} \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}, \quad \text{s.t.} [\alpha^T \quad \alpha^{*T}] \mathbf{1} = 1, \quad \alpha, \alpha^* \geq 0, \right. \quad (13)$$

where  $\alpha, \alpha^*$  are the Lagrange coefficients and  $\tilde{\mathbf{K}}$  is the kernel function. They are

$$\begin{aligned} y &= \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \\ \alpha &= \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \\ \alpha^* &= \begin{bmatrix} \alpha_1^* \\ \vdots \\ \alpha_n^* \end{bmatrix}, \\ \tilde{\mathbf{K}} &= \begin{bmatrix} \tilde{k}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) \end{bmatrix} = \begin{bmatrix} \mathbf{K} + \frac{\mu n}{\lambda} \mathbf{I} & -\mathbf{K} \\ -\mathbf{K} & \mathbf{K} + \frac{\mu n}{\lambda} \mathbf{I} \end{bmatrix}, \end{aligned} \quad (14)$$

where  $\mathbf{K}$  is the Gaussian kernel function.

The values of the variables obtained by the solution are

$$\begin{cases} \mathbf{p} = \lambda \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\tilde{\mathbf{x}}_i), \\ \xi_i = \alpha_i \mu n, \\ \xi_i^* = \alpha_i^* \mu n. \end{cases} \quad (15)$$

In addition, because  $\sum_{i=1}^n (\alpha_i + \alpha_i^*) = 1, \mu = \sum_{i=1}^n (\xi_i + \xi_i^*)/n$ .

#### Step 5. Prediction.

The prediction function is shown in the following equation:

$$\begin{aligned} y &= \mathbf{p}^T \varphi(\tilde{\mathbf{x}}_{\text{test}}) = \lambda \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi^T(\tilde{\mathbf{x}}_i) \varphi(\tilde{\mathbf{x}}_{\text{test}}) \\ &= \lambda \sum_{i=1}^n (\alpha_i - \alpha_i^*) \tilde{\mathbf{K}}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_{\text{test}}). \end{aligned} \quad (16)$$

If it is used for classification,

$$y = \text{sign}(\mathbf{p}^T \varphi(\tilde{\mathbf{x}}_{\text{test}})). \quad (17)$$

If  $y > 0$ , it belongs to the positive class, and if  $y < 0$ , it belongs to the negative class.

It can be seen from this section that the algorithm proposed in this paper is a quadratic programming problem.

**3.2. The Center-Constrained MEB Problem.** In 2002, Bădoiu and Clarkson proposed a minimum enclosing ball (MEB)  $(1 + \xi)$ -approximation algorithm based on the core set in the literature [26]. The algorithm uses a subset of the input set, which is called the core set, to solve the optimization problem. The algorithm can obtain the same good approximation results as the original input set to improve the efficiency of the algorithm. Tsang et al. [29] suggested that the MEB problem is related to many kernel problems. Eligible quadratic programming (QP) problems can be solved quickly by the core-set algorithm. The following section briefly introduces the center-constrained minimum enclosing ball (CC-MEB) algorithm. Next, we will introduce the relationship between the proposed algorithm and CC-MEB and implement the fast algorithm proposed in this paper.

Given the training sample  $S = \{\varphi(\mathbf{x}_i)\}_{i=1}^m$ , where  $\mathbf{x}_i \in R^d$  and  $\varphi$  is the kernel mapping associated with a given kernel  $\mathbf{K}$ , adding one dimension  $\delta_i$  to each  $\varphi(\mathbf{x}_i)$  forms a set  $S = \{(\varphi^T(\mathbf{x}_i), \delta_i)\}_{i=1}^m$ . By setting the coordinate of the last one-dimensional center point to be 0, that is, the CC-MEB's coordinate is  $[\mathbf{c}, 0]$ , then the optimization problem of finding the smallest CC-MEB that contains all the samples in the set  $S$  is

$$\left\{ \min_{\mathbf{c}, R} R^2, \quad \text{s.t.} (\varphi(\mathbf{x}_i) - \mathbf{c})^2 + \delta_i^2 \leq R^2, \right. \quad (18)$$

where  $i = 1, 2, \dots, m$ .

Let  $\mathbf{\Delta} = [\delta_1^2, \dots, \delta_m^2]^T \geq 0$ ; then, the matrix form of the corresponding dual problem of equation (18) is

$$\left\{ \max_{\boldsymbol{\beta}} \boldsymbol{\beta}^T (\text{diag}(\mathbf{K}) + \mathbf{\Delta}) - \boldsymbol{\beta}^T \mathbf{K} \boldsymbol{\beta}, \quad \text{s.t.} \boldsymbol{\beta} \geq 0, \boldsymbol{\beta}^T \mathbf{1} = 1, \right. \quad (19)$$

where the kernel matrix is

$$\mathbf{K}_{m \times m} = [k(\mathbf{x}_i, \mathbf{x}_j)] = [\varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}_j)]. \quad (20)$$

Using the optimal solution  $\boldsymbol{\beta}$  to obtain the values for the radius  $R$  and center point  $\mathbf{c}$ ,

$$\begin{cases} R = \sqrt{\boldsymbol{\beta}^T (\text{diag}(\mathbf{K}) + \mathbf{\Delta}) - \boldsymbol{\beta}^T \mathbf{K} \boldsymbol{\beta}} \\ \mathbf{c} = \sum_{i=1}^m \beta_i \varphi(\mathbf{x}_i). \end{cases} \quad (21)$$

The formula for the distance from any point to the center point is

$$\|\mathbf{c} - \varphi(\mathbf{x}_l)\|^2 + \delta_l^2 = \|\mathbf{c}\|^2 - 2(\mathbf{K}\boldsymbol{\beta})_l + k_{ll} + \delta_l^2. \quad (22)$$

Because  $\boldsymbol{\beta}^T \mathbf{1} = 1$ , any real number  $\eta$  is added to equation

Input: Dataset  $\{Z_i = (\tilde{\mathbf{x}}_i, y_i)\}_{i=1}^n$ , approximation parameter  $\xi$ , parameter  $\eta$ , parameter  $\lambda$ , parameter  $\mu$ , and kernel width  $\delta$ , where  $\xi = 1e-6$  and  $\eta = \max(0, \max(\text{diag}(\tilde{\mathbf{K}}) - (2/C) \begin{bmatrix} y \\ -y \end{bmatrix}))$

Output: Core-set  $Q$ , Lagrangian coefficient  $\tilde{\alpha}$

Training steps:

Step 1: Randomly select 20 samples to form the initial core set  $Q_0$ ;  
Generate the center  $\mathbf{c}_0$  and radius  $R_0$  of the initial CC-MEB according to equation (21) and set the number of iterations  $t$  to be 0

Step 2: Randomly select 59 samples and calculate the distance from any sample to the center of the CC-MEB according to equation (22). If there is no sample  $\tilde{\mathbf{x}}$  outside  $CC\text{-}MEB(\mathbf{c}_t, (1 + \xi)R_t)$ , proceed to step 6

Step 3: Find the farthest sample from the center  $\mathbf{c}_t$  in step 2 and add the sample to core-set  $Q_{t+1} = Q_t \cup \{\tilde{\mathbf{x}}\}$

Step 4: Solve the new CC-MEB, recorded as  $MEB(Q_{t+1})$ , and  $\mathbf{c}_{t+1} = \mathbf{c}_{MEB}(Q_{t+1})$ ,  $R_{t+1} = R_{MEB}(Q_{t+1})$

Step 5: Set  $t = t + 1$  and return to step 2

Step 6: Terminate the training and return the required output

Prediction step:

Input the test sample  $\tilde{\mathbf{x}}_{\text{test}}$  into the following:  
 $y_{\text{test}} = \mathbf{p}^T \varphi(\tilde{\mathbf{x}}_{\text{test}}) = \lambda \mathbf{c}^T \varphi(\tilde{\mathbf{x}}_{\text{test}})$

ALGORITHM 1: The Fast-RBF algorithm.

(19), which does not affect the value of  $\beta$ . The original dual form is changed to

$$\begin{cases} \max_{\beta} \beta^T (\text{diag}(\mathbf{K}) + \Delta - \eta \mathbf{1}) - \beta^T \mathbf{K} \beta, & \text{s.t. } \beta \geq 0, \beta^T \mathbf{1} = 1, \Delta \geq 0. \end{cases} \quad (23)$$

Reference [29] pointed out that any QP problem that satisfies equation (23) can be regarded as an CC-MEB problem, which can be solved by the core-set fast algorithm

**3.3. Relationship between Fast-RBF and CC-MEB.** Equation (13) is the QP form of Fast-RBF. Let  $\tilde{\alpha} = [\alpha^T \quad \alpha^{*T}]^T$ ; then,

$$\Delta = -\text{diag}(\tilde{\mathbf{K}}) + \eta \mathbf{1} + \frac{2}{\lambda} \begin{bmatrix} y \\ -y \end{bmatrix}, \quad (24)$$

where the real number  $\eta$  should be large enough so that  $\Delta \geq 0$ . Thus, equation (13) can be written as follows:

$$\begin{cases} \max_{\tilde{\alpha}} \tilde{\alpha}^T (\text{diag}(\tilde{\mathbf{K}}) + \Delta - \eta \mathbf{1}) - \tilde{\alpha}^T \tilde{\mathbf{K}} \tilde{\alpha} \\ \text{s.t.} \quad \tilde{\alpha}^T \mathbf{1} = 1. \end{cases} \quad (25)$$

This form is equivalent to the CC-MEB problem from equation (23), and the problem can be solved using the core-set fast algorithm [29].

According to formula (25), the center of sphere  $\mathbf{c}$  can be calculated as  $\mathbf{c} = \sum_{i=1}^{2*n} \tilde{\alpha}_i \tilde{\varphi}(\tilde{\mathbf{x}}_i)$ . In the formula, when  $i = 1, \dots, n$ , then  $\tilde{\varphi}(\tilde{\mathbf{x}}_i) = \varphi(\tilde{\mathbf{x}}_i)$ ; when  $i = n + 1, \dots, 2n$ , then  $\tilde{\varphi}(\tilde{\mathbf{x}}_i) = -\varphi(\tilde{\mathbf{x}}_i)$ , and the derivation is available:

$$\mathbf{c} = \sum_{i=1}^{2*n} \tilde{\alpha}_i \tilde{\varphi}(\tilde{\mathbf{x}}_i) = \sum_{i=1}^n \alpha_i \varphi(\tilde{\mathbf{x}}_i) + \sum_{i=1}^n \alpha_i^* (-\varphi(\tilde{\mathbf{x}}_i)) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\tilde{\mathbf{x}}_i). \quad (26)$$

Therefore, the value of  $\mathbf{p}$  in equation (15) is  $\mathbf{p} = \lambda \mathbf{c}$ .

**3.4. The Implementation of Fast-RBF.** Algorithm 1 describes the steps of the Fast-RBF algorithm, and the flow chart is shown in Figure 2.

## 4. Experimental Results and Analysis

In this paper, the effectiveness of the proposed method is verified by comparing it with the traditional RBF algorithm on MR images. The experiment is divided into two stages: the parameter optimization stage and the modeling stage. In the parameter optimization stage, the grid search method is used to obtain the optimal parameters of each algorithm based on the training set. In the modeling stage, the training set is modeled using optimal parameters, and the test set is used to obtain the performance of each algorithm.

The experiment is verified from the following four aspects:

- (1) Verify that the size of the core set of the Fast-RBF algorithm is much smaller than the training set's scale, which can speed up the modeling time of the algorithm
- (2) Verify that the prediction capability of the Fast-RBF algorithm is comparable to the prediction capability of the RBF algorithm
- (3) Verify that the modeling time of the Fast-RBF algorithm on large datasets is much smaller than that of the RBF algorithm

For the experimental environment, the operating system is Windows 10; the processor is an Intel i5 2.71 GHz CPU; the memory is 8 GB; and the main application software is MATLAB R2015a.

**4.1. Experimental Preparation.** The use case in this section is from MRI scans of five subjects recruited by the University Hospitals Cleveland Medical Center Institutional Review Board. Before the experiment, a block diagram is first used



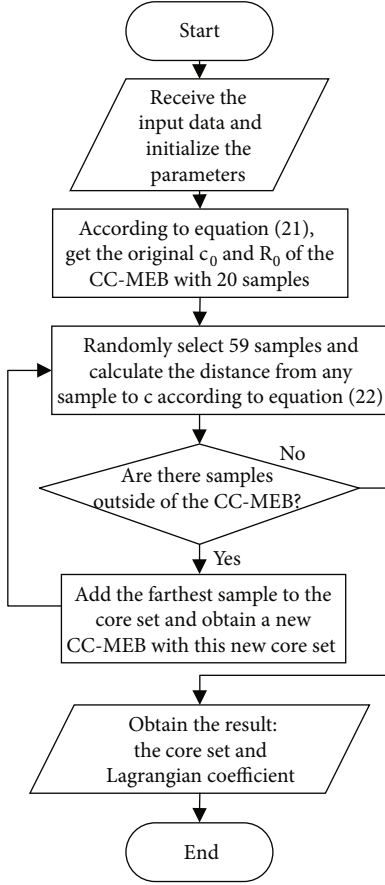


FIGURE 2: Flow chart of the algorithm.

to frame the area to be identified, as shown in Figure 3. Next, we train and test the data of region of interest in abdominal organ map. The experiment is to classify the liver and kidneys of the region of interest in the abdominal organ map.

For each case, we extract the local textural features from four different types of abdominal 3D MR images, namely, fat, water, in-phase (IP), and opposed-phase (OP), as the input of the algorithm. Noise cannot be avoided in these actual data, and this noise will affect the final image recognition effect. Therefore, this paper adopts the method proposed in [30, 31] to design a convolution kernel as shown in Table 1, preprocesses the experimental data, and implements feature extraction.

In addition, we also consider the pixel spacing of the MR images and adopt the grid division strategy. Let  $(x, y)$  represent the position information of the pixel. That is, we combine the features that we extracted and obtain a six-dimensional feature.

**4.2. Experimental Method.** We define the liver of the region of interest in abdominal MR images as class A, the kidneys as class B, and other tissues as class C. Therefore, this is a multi-classification problem. We train “liver (class A)-kidney (class B),” “liver (class A)-other tissue (class C),” and “kidney (class B)-other tissue (class C)” to obtain three classification results; the final result is then determined by voting. The voting method is as follows:

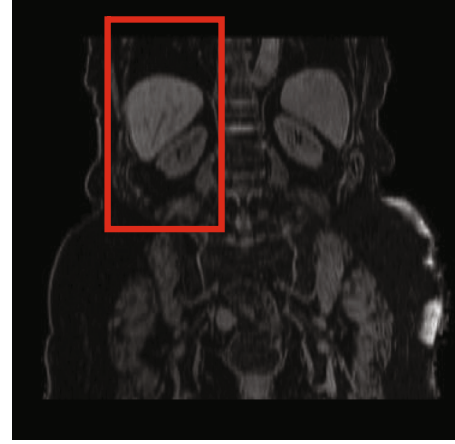


FIGURE 3: Areas to be identified.

TABLE 1: Convolutional kernel  $CK_{3 \times 3}$ .

0.1	0.1	0.1
0.1	0.2	0.1
0.1	0.1	0.1

Let  $A = B = C = 0$ .

The classification is  $(A, B)$  if it belongs to A, and  $A = A + 1$ ; otherwise,  $B = B + 1$ .

The classification is  $(A, C)$  if it belongs to A, and  $A = A + 1$ ; otherwise,  $C = C + 1$ .

The classification is  $(B, C)$  if it belongs to B, and  $B = B + 1$ ; otherwise,  $C = C + 1$ .

The final sample belongs to the class with the largest values of A, B, and C.

The classification accuracy is used to measure the performance of the algorithm.

$$\text{prediction accuracy} = \frac{\text{the number of correctly classified test samples}}{\text{total number of test samples}}. \quad (27)$$

**4.3. Experimental Results.** Cases 1-4 contain a total of 59,904 data points. We randomly selected 10,000 data points, 20,000 data points, 30,000 data points, 40,000 data points, 50,000 data points, and 59,904 data points for training. Case 5, which contains 16,896 data points, was used as the test set. The experiment was repeated 10 times for each training set size to verify the advantages of the proposed method.

**4.3.1. Core Set of the Fast-RBF Algorithm.** Figure 4 shows the average values of the total number of core-set samples for the three classifiers at different training set sizes. Figure 4 shows that the total number of core sets is between 240 and 300, which is much smaller than the sample size. Replacing all the samples with the core sets in the model construction step will greatly improve the operational efficiency.

**4.3.2. Prediction Ability of the Fast-RBF Algorithm.** It can be seen from Table 2 that both algorithms can achieve a good generalization performance. However, with the increase in

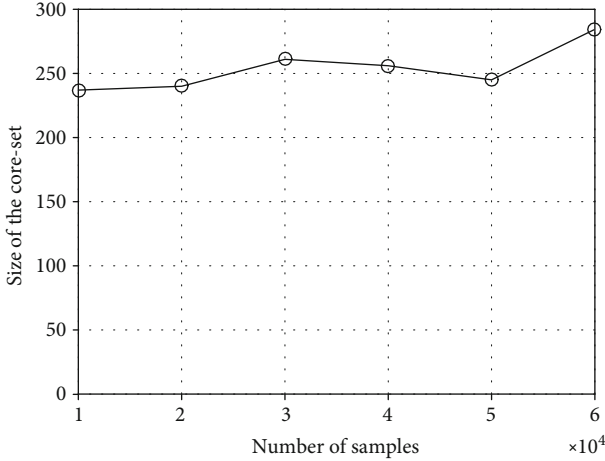


FIGURE 4: The size of the core set at different sample sizes.

TABLE 2: Prediction accuracy and standard deviation of the two algorithms at different dataset sizes.

Size of the dataset	Prediction accuracy and standard deviation	
	RBF	Fast-BRF
10,000	$0.9345 \pm 0.0083$	$0.9458 \pm 0.0153$
20,000	$0.9389 \pm 0.0063$	$0.9551 \pm 0.0132$
30,000	$0.9381 \pm 0.0104$	$0.9496 \pm 0.0098$
40,000	—	$0.9467 \pm 0.0111$
50,000	—	$0.9432 \pm 0.0112$
59,904	—	$0.9552 \pm 0.0066$

the amount of training data, the RBF algorithm requires more samples to participate in the modeling step, so it is more constrained. When the data size exceeds 30,000 data points, it can no longer be solved. The Fast-RBF algorithm uses core-set technology to solve the problem. Key samples are added to the core set one by one, and the average size of the core set does not exceed 300, so it can process a larger dataset and can achieve a generalization ability comparable to that of the RBF algorithm. The organ classification results are shown in Figure 5.

**4.3.3. Time Performance of the Fast-RBF Method.** Table 3 shows that the modeling time required by the two algorithms has a stable growth with increasing sample size. When the size of the dataset is 30,000 data points, the average modeling time of the RBF algorithm is 7,580 seconds, while the average time of the Fast-RBF algorithm is 15.2609 seconds. The modeling time of the Fast-RBF algorithm is much smaller than that of the RBF algorithm. In addition, when the size of the dataset is more than 30,000 data points, the RBF algorithm will not run.

**4.4. Experimental Conclusion.** It is known from experiments that the Fast-RBF algorithm can be used for organ recognition in MR images. The advantage of the proposed algorithm is that it requires far less modeling time than the RBF algo-

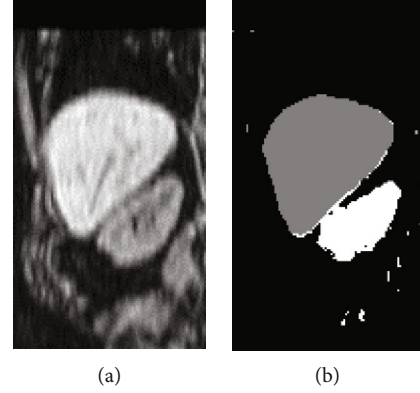


FIGURE 5: Test results. (a) Original picture. (b) Organ classification results.

TABLE 3: Average modeling time and standard deviation of the two algorithms under different data sizes.

Size of the dataset	Modeling time and standard deviation of each method (s)	
	RBF	Fast-BRF
10,000	$210.5813 \pm 3.5134$	$10.1719 \pm 7.0177$
20,000	$737.1344 \pm 7.1357$	$12.8016 \pm 4.0126$
30,000	$7.58E+03 \pm 164.0596$	$15.2609 \pm 8.2559$
40,000	—	$16.5953 \pm 9.1518$
50,000	—	$17.5953 \pm 9.1983$
59,904	—	$22.9781 \pm 7.0587$

rithm in large datasets under the premise of ensuring the prediction accuracy. The algorithm has strong practicability.

## 5. Conclusion

Our studies are based on MRI of challenging body sections of the abdomen. We proposed the Fast-RBF algorithm, which is suitable for the rapid training of a large dataset. By introducing the  $\epsilon$ -insensitive loss function, learning the structural risk term and kernel method of the support vector machine, and using the core-set principle, the proposed algorithm can meet the needs of large sample sizes. This method can quickly process large datasets and is suitable for medical image processing.

The method proposed in this paper is a supervised learning method. The training samples need to be labeled, and the workload of data preparation is large. In the future, we will further study the semisupervised abdominal image recognition method in which only a small number of class labels are needed to achieve image processing.

## Data Availability

Data sharing is not available for our study, as the experimental data were afforded by our collaboration partners at the University Hospitals Cleveland Medical Center, OH, USA. Without permission, we cannot share any of our data with others.

## Additional Points

*Human and Animal Rights.* The experimental data are from five subjects recruited by the University Hospitals Cleveland Medical Center Institutional Review Board. We used these data to be compatible with the board's requirements.

## Consent

Informed consent was obtained from all the individual participants included in this study.

## Disclosure

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, USA.

## Conflicts of Interest

We declare that there are no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grants 61772241 and 61702225, by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under grant 18KJB520048, by the Science and Technology Demonstration Project of Social Development of Wuxi under grant WX18IVJN002, and by JiangSu 333 expert engineering. The research in this publication was also supported by the National Cancer Institute of the National Institutes of Health, USA, under award number R01CA196687.

## References

- [1] N. Upadhyay and A. D. Waldman, "Conventional MRI evaluation of gliomas," *The British Journal of Radiology*, vol. 84, pp. S107–S111, 2011.
- [2] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7405–7415, 2016.
- [3] H. Zhang, Z. Gao, L. Xu et al., "A meshfree representation for cardiac medical image computing," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–12, 2018.
- [4] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 865–878, 2017.
- [5] M. Li, S. Dong, Z. Gao et al., "Unified model for interpreting multi-view echocardiographic sequences without temporal information," *Applied Soft Computing*, vol. 88, article 106049, 2020.
- [6] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," *Magnetic Resonance Imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.
- [7] Y. J. Zhang and J. J. Gerbrands, "Transition region determination based thresholding," *Pattern Recognition Letters*, vol. 12, no. 1, pp. 13–23, 1991.
- [8] C. Yan, N. Sang, and T. Zhang, "Local entropy-based transition region extraction and thresholding," *Pattern Recognition Letters*, vol. 24, no. 16, pp. 2935–2941, 2003.
- [9] P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, "Tumour volume determination from MR images by morphological segmentation," *Physics in Medicine and Biology*, vol. 41, no. 11, pp. 2437–2446, 1996.
- [10] R. Adams and L. Bishof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.
- [11] Y. Yang, D. Tian, and B. Wu, "A fast and reliable noise-resistant medical image segmentation and bias field correction model," *Magnetic Resonance Imaging*, vol. 54, pp. 15–31, 2018.
- [12] J. Nie, Z. Xue, T. Liu et al., "Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov Random Field," *Computerized Medical Imaging and Graphics*, vol. 33, no. 6, pp. 431–441, 2009.
- [13] M. Havaei, A. Davy, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
- [14] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [15] A. S. Miller, B. Blott, and T. K. Hames, "Review of neural network applications in medical imaging and signal processing," *Medical & Biological Engineering & Computing*, vol. 30, no. 5, pp. 449–464, 1992.
- [16] R. Lucht, S. Delorme, and G. Brix, "Neural network-based segmentation of dynamic MR mammographic images," *Magnetic Resonance Imaging*, vol. 20, no. 2, pp. 147–154, 2002.
- [17] M. Egmont-Petersen, A. F. Frangi, W. J. Niessen et al., "Segmentation of bone tumor in MR perfusion images using neural networks and multiscale pharmacokinetic features," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, pp. 80–83, Barcelona, Spain, September 2000.
- [18] C. Zhang, K. Qiao, L. Wang et al., "A visual encoding model based on deep neural networks and transfer learning for brain activity measured by functional magnetic resonance imaging," *Journal of Neuroscience Methods*, vol. 325, pp. 108318–108319, 2019.
- [19] N. J. Tustison, B. B. Avants, and J. C. Gee, "Learning image-based spatial transformations via convolutional neural networks: a review," *Magnetic Resonance Imaging*, vol. 64, pp. 142–153, 2019.
- [20] N. Khalili, N. Lessmann, E. Turk et al., "Automatic brain tissue segmentation in fetal MRI using convolutional neural networks," *Magnetic Resonance Imaging*, vol. 64, pp. 77–89, 2019.
- [21] Z. Wang, L. Xie, and J. Qi, "Dynamic pixel-wise weighting-based fully convolutional neural networks for left ventricle segmentation in short-axis MRI," *Magnetic Resonance Imaging*, vol. 66, pp. 131–140, 2020.
- [22] A. K. Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 63–74, 2019.

- [23] J. O. Glass, W. E. Reddick, O. Goloubeva, V. Yo, and R. G. Steen, "Hybrid artificial neural network segmentation of precise and accurate inversion recovery (PAIR) images from normal human brain," *Magnetic Resonance Imaging*, vol. 18, no. 10, pp. 1245–1253, 2000.
- [24] R. Pérez de Alejo, J. Ruiz-Cabello, M. Cortijo et al., "Computer-assisted enhanced volumetric segmentation magnetic resonance imaging data using a mixture of artificial neural networks," *Magnetic Resonance Imaging*, vol. 21, no. 8, pp. 901–912, 2003.
- [25] W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton, "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 911–918, 1997.
- [26] M. Bădoiu and K. L. Clarkson, "Optimal core-sets for balls," *Computational Geometry: Theory and Applications*, vol. 40, no. 1, pp. 14–22, 2008.
- [27] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, 1995.
- [28] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, New York, NY, USA, 2013.
- [29] I. W. Tsang, J. T. Kwok, and J. M. Zurada, "Generalized core vector machines," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1126–1140, 2006.
- [30] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.
- [31] J. Zheng, J. Cao, Z. Wang et al., "Semi-automatic synthetic computed tomography generation for abdomens using transfer learning and semi-supervised classification," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 9, pp. 1878–1886, 2019.

## Research Article

# A Combined Ultrasonic Backscatter Parameter for Bone Status Evaluation in Neonates

Weiying Mao <sup>1</sup>, Yang Du <sup>1</sup>, Chengcheng Liu <sup>2</sup>, Boyi Li <sup>3</sup>, Dean Ta <sup>3,4</sup>, Chao Chen,<sup>1</sup> and Rong Zhang <sup>1</sup>

<sup>1</sup>Department of Neonatology, Children's Hospital of Fudan University, Shanghai 201102, China

<sup>2</sup>Institute of Acoustics, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China

<sup>3</sup>Department of Electronic Engineering, Fudan University, Shanghai 200433, China

<sup>4</sup>Academy for Engineering and Technology, Fudan University, Shanghai 200433, China

Correspondence should be addressed to Rong Zhang; [zhang\\_rong@fudan.edu.cn](mailto:zhang_rong@fudan.edu.cn)

Received 2 March 2020; Accepted 30 March 2020; Published 1 May 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Weiying Mao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Metabolic bone disease (MBD) is one of the major complications of prematurity. Ultrasonic backscatter technique has the potential to be a portable and noninvasive method for early diagnosis of MBD. This study firstly applied CAS to neonates, which was defined as a linear combination of the apparent integrated backscatter coefficient (AIB) and spectral centroid shift (SCS). The objective was to evaluate the feasibility of ultrasonic backscatter technique for assessing neonatal bone health using AIB, SCS, and CAS. Ultrasonic backscatter measurements at 3.5 MHz, 5.0 MHz, and 7.5 MHz were performed on a total of 505 newborns within 48 hours after birth. The values of backscatter parameters were calculated and compared among gestational age groups. Correlations between backscatter parameters, gestational age, anthropometric indices, and biochemical markers were analyzed. The optimal predicting models for CAS were determined. The results showed term infants had lower SCS and higher AIB and CAS than preterm infants. Gestational age and anthropometric indices were negatively correlated with SCS ( $|r| = 0.45 - 0.57$ ,  $P < 0.001$ ), and positively correlated with AIB ( $|r| = 0.36 - 0.60$ ,  $P < 0.001$ ) and CAS ( $|r| = 0.56 - 0.69$ ,  $P < 0.001$ ). Biochemical markers yielded weak or nonsignificant correlations with backscatter parameters. CAS had relatively stronger correlations with the neonatal variables than AIB and SCS. At 3.5 MHz and 5.0 MHz, only gestational age ( $P < 0.001$ ) independently contributed to the measurements of CAS, and could explain up to 40.5% – 44.3% of CAS variation. At 7.5 MHz, the combination of gestational age ( $P < 0.001$ ), head circumference ( $P = 0.002$ ), and serum calcium ( $P = 0.037$ ) explained up to 40.3% of CAS variation. This study suggested ultrasonic backscatter technique was feasible to evaluate neonatal bone status. CAS was a promising parameter to provide more information about bone health than AIB or SCS alone.

## 1. Introduction

Metabolic bone disease (MBD) is one of the major complications of prematurity, characterized by disorders of calcium and phosphorus metabolism and inadequate bone mineralization. Premature infants are at high risk of MBD because approximately 80% fetal bone mineral accretion occurs in the third trimester of gestation [1]. The lack of bone mineral deposition can be aggravated after birth by rapid bone growth, prolonged parenteral nutrition, and exposure to specific drugs such as diuretics [2]. Despite improved nutritional and medical strategies, the incidence of MBD is estimated to

be 16% – 40% in infants with birth weight less than 1500 g [1]. MBD impacts both short-term and long-term prognosis of prematurity. In addition to rickets and spontaneous fractures [3], it leads to compromised lung development [3] and short stature in childhood [4]. Therefore, early diagnosis of MBD is critical.

However, it is difficult to recognize MBD with few symptoms in the early stage. Current diagnostic approaches mainly include serum biochemical markers and radiological examinations. Although the combination of serum alkaline phosphatase (AKP)  $> 900$  IU/L and phosphorus  $< 1.80$  mmol/L yields a sensitivity of 100% and a specificity of 70% for diagnosis



of low bone mineral density (BMD) [5], biochemical markers are not necessarily associated with BMD and can be affected by other diseases such as cholestasis. Repeated blood sampling is also not preferred for preterm infants. Bedside X-ray is convenient to show osteopenia and fractures, but is insensitive to a loss of bone mass less than 20% – 40% [6]. While dual energy X-ray absorptiometry (DEXA) and quantitative computed tomography (QCT) are widely used in adults to diagnose osteoporosis with high accuracy and sensitivity, they are restricted in neonates due to the risk of radiation and inconvenience to move. There is a pressing need for a valid, efficient, and noninvasive method to assess and monitor neonatal bone health.

Quantitative ultrasound (QUS) is a portable, low-cost, and radiation-free diagnostic technique which has been developed in the last decades. It not only reflects BMD [7–9], but also reflects bone microstructure [10–12] and mechanical properties [13–15]. In transmission mode, speed of sound (SOS) and bone transmission time (BTT) are commonly used parameters for neonatal bone status evaluation. Studies have shown that SOS and BTT are significantly higher in term infants than in preterm infants and positively correlated with gestational age [16–22]. On the other hand, ultrasonic transmission technique has its limitations. Scattering and dispersion are unavoidable in ultrasonic propagation, but are not considered in transmission measurements. And the requirement for a pair of transducers parallel to each other reduces the reproducibility.

Another modality of QUS, ultrasonic backscatter technique, has drawn more attention in recent years [23–29]. Unlike transmission technique, backscatter measurements are based on pulse-echo mode with a single transducer to both transmit and receive signals. It is easy to examine central skeletal sites [30, 31] and operate in incubators. Ultrasonic backscatter technique may be a promising approach to MBD screening [23–25], but few studies have been performed on neonates. Zhang et al. [32] firstly measured apparent backscatter coefficient (BSC) in 122 neonates and revealed significant correlations with gestational age, weight, and length at birth ( $|r| = 0.43 - 0.47$ ,  $P < 0.001$ ). Liu et al. [33] proposed a signal selection standard of apparent backscatter parameters for neonatal bone evaluation, including apparent integrated backscatter coefficient (AIB), frequency intercept of apparent backscatter (FIAB), frequency slope of apparent backscatter (FSAB), and spectral centroid shift (SCS). Our previous study [34] also suggested AIB, FIAB, and FSAB were feasible for assessing and monitoring neonatal bone status.

As different parameters reflect different properties of cancellous bone, the combination of backscatter parameters may provide more structural information. Recently, a new parameter, a linear combination of AIB and SCS (CAS), was introduced by Tang et al. [35]. They recruited 1262 adults and found that CAS was significantly correlated with BMD ( $|r| = 0.73 - 0.84$ ,  $P < 0.05$ ). The correlation coefficients of CAS were higher than that of AIB and SCS ( $|r| = 0.48 - 0.69$ ,  $P < 0.05$ ). There was no report about CAS for bone status evaluation in neonates.

We designed this study to evaluate the feasibility of ultrasonic backscatter technique for assessing neonatal bone

TABLE 1: Characteristics of the transducers used in this study.

Model	Diameter (inch)	Central frequency (MHz)	–6 dB bandwidth (band range) (MHz)
V546	0.250	3.5	3.25 (1.60 – 4.91)
V543	0.250	5.0	3.91 (3.20 – 7.11)
V122	0.375	7.5	5.65 (4.79 – 10.44)

health using AIB, SCS, and CAS. To the best of our knowledge, CAS was applied to neonates for the first time.

## 2. Materials and Methods

**2.1. Participants.** Newborns were eligible for this study who were less than 48 hours after birth and hospitalized in the Department of Neonatology, Children’s Hospital of Fudan University, Shanghai, China between October 9, 2017 and May 30, 2019. Infants were excluded if born with congenital malformations, chromosomal abnormalities or inherited metabolic diseases. A total of 505 infants were enrolled, including 268 males and 237 females. All participants were divided into four groups according to gestational age: PRE-1 for preterm infants with gestational age less than 28 weeks, PRE-2 for those with gestational age between 28 weeks and 31<sup>+6</sup> weeks, PRE-3 for those with gestational age between 32 weeks and 36<sup>+6</sup> weeks, and TERM for term infants born at  $\geq 37$  weeks of gestation. For each infant, anthropometric indices were measured at birth including birth weight, length, and head circumference. Blood sampling was taken immediately after they were admitted to the hospital. Serum calcium, phosphorus, and AKP were tested.

Informed consents were signed by parents of each participant before enrollment. The study protocol was approved by the Ethics Committee of Children’s Hospital of Fudan University (No. 25/2016).

**2.2. Ultrasonic Backscatter Measurements.** A novel ultrasonic backscatter bone diagnostic instrument (UBBD; Fudan University, Shanghai, China) was applied in this study. The backscatter signals were transmitted and received by a single planar transducer with central frequencies of 3.5 MHz, 5.0 MHz, and 7.5 MHz (Panametrics, Waltham, MA, USA) (Table 1). The transducers were excited by a bipolar short pulse with a voltage of approximately  $\pm 50$  V from the UBBD instrument. Ultrasonic backscatter measurements at 3.5 MHz, 5.0 MHz, and 7.5 MHz were carried out within 48 hours after birth for each participant, and were performed with only one operator in order to avoid measurement errors caused by different operators. The transducers were placed on the medial part of the heel and coupled by ultrasonic gel (Aloka Medical Equipment, Shanghai, China), where the surface was flat and soft tissue was thin atop the calcaneus. Each measurement was finished in 5 seconds. The instrument further conducted signal preprocessing, amplification and a 14-bit analog to digital conversion with a sampling frequency of 50.0 MHz. To reduce random noise, 128 waveforms were averaged in the time domain and the signals were stored for offline analysis.

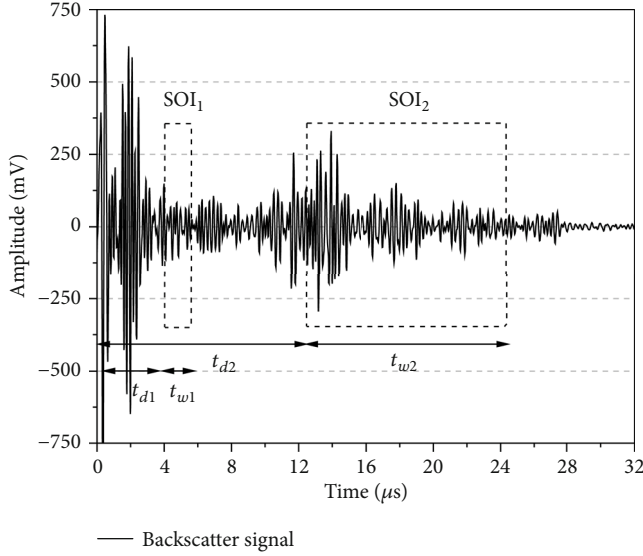


FIGURE 1: A typical ultrasonic backscatter signal from a female infant born at 29 weeks of gestation at 5.0 MHz frequency. The delay ( $t_d$ ) and duration ( $t_w$ ) of signals of interest (SOI) varies for AIB and SCS. For the SOI<sub>1</sub> of AIB,  $t_{d1} = 4 \mu s$  and  $t_{w1} = 2 \mu s$  at 3.5 MHz,  $1.4 \mu s$  at 5.0 MHz, and  $0.92 \mu s$  at 7.5 MHz. For the SOI<sub>2</sub> of SCS,  $t_{d2} = 12.5 \mu s$  and  $t_{w2} = 12 \mu s$ .

**2.3. Ultrasonic Backscatter Parameter Calculation.** MATLAB R2018b (MathWorks, Natick, MA, USA) was used for signal analysis and parameter calculation. Figure 1 illustrates a typical ultrasonic backscatter signal acquired from a female infant born at 29 weeks of gestation at 5.0 MHz. Different rectangular windows were set to select the signals of interest (SOI) for AIB and SCS. The gate delay ( $t_d$ ) of SOI<sub>1</sub> for AIB was  $4 \mu s$  to avoid intervening signals from soft tissue and cortical bone. The duration ( $t_w$ ) was  $2 \mu s$  at 3.5 MHz,  $1.4 \mu s$  at 5.0 MHz, and  $0.92 \mu s$  at 7.5 MHz. The SOI<sub>2</sub> for SCS located in the time range of  $12.5 - 24.5 \mu s$  at all the frequencies. The locations of SOIs depended on the previous study by Liu et al. [33], as well as optimization of the correlations between backscatter parameters, anthropometric indices and biochemical markers.

AIB was defined as the integrated value of apparent backscatter transfer function (ABTF) over the  $-6$  dB frequency bandwidth [36, 37]:

$$ABTF = 10 \log_{10} \left( \frac{P_s(f)}{P_r(f)} \right), \quad (1)$$

$$AIB = \frac{\int_{f_{\min}}^{f_{\max}} ABTF df}{f_{\max} - f_{\min}}. \quad (2)$$

SCS was the downshift of the spectral centroid of the backscatter signal ( $f_s$ ) compared to that of the reference signal ( $f_r$ ). It was calculated as [38, 39].

$$SCS = f_s - f_r = \frac{\int_{f_{\min}}^{f_{\max}} f \cdot P_s(f) df}{\int_{f_{\min}}^{f_{\max}} P_s(f) df} - \frac{\int_{f_{\min}}^{f_{\max}} f \cdot P_r(f) df}{\int_{f_{\min}}^{f_{\max}} P_r(f) df}. \quad (3)$$

In the formulas above,  $f_{\max}$  and  $f_{\min}$  were the upper and lower limit of the  $-6$  dB frequency bandwidth of the transducer;  $P_s(f)$  and  $P_r(f)$  referred to the power spectrum of the backscatter signal and the reference signal, respectively. The reference signal was a reflected signal from a polished steel plate immersed in pure water. The power spectrum was obtained through a fast Fourier transform.

CAS was a linear combination of AIB and SCS, which was defined as [35].

$$CAS = \omega AIB - SCS. \quad (4)$$

According to Tang et al. [35], the coefficient  $\omega$  was a positive number depended on when the correlation between CAS and BMD achieved best. Considering that BMD of the newborns was unable to measure directly and fetal bone matured with increased gestational age, the value of  $\omega$  in this study was determined by optimization of the correlation between CAS and gestational age instead. The values of  $\omega$  were 0.041 at 3.5 MHz, 0.030 at 5.0 MHz, and 0.072 at 7.5 MHz.

For each participant, AIB, SCS and CAS were calculated at all the frequencies of 3.5 MHz, 5.0 MHz and 7.5 MHz. Different frequency bands were put into the same formula for each parameter.

**2.4. Statistical Analyses.** We used SPSS 22.0 (IBM, Armonk, NY, USA) for statistical analysis. The normality of all the variables was checked by the Shapiro-Wilk test. None obeyed normal distribution. Descriptive data were presented as median and quartile. Differences among subgroups based on gestational age were examined using the Kruskal-Wallis  $H$  test followed by all pairwise comparisons. Correlations between ultrasonic backscatter parameters, gestational age, anthropometric indices, and biochemical markers were calculated by simple linear regression (Spearman's rank correlation). Relative contributions of gestational age, anthropometric indices, and biochemical markers to the measurements of CAS were determined using multiple linear regression. The optimal models for predicting CAS were produced by forward stepwise multiple regression. A  $P$  value less than 0.05 indicated statistical significance.

### 3. Results

Table 2 summarizes the baseline characteristics of the participants. No significant difference in gender was found across subgroups. Birth weight, length, head circumference, and serum calcium of the enrolled infants increased with gestational age. Serum phosphorus and AKP at birth decreased with gestational age.

As Table 3 demonstrates, AIB and CAS were significantly higher and SCS was significantly lower in term infants compared with any group of preterm infants at 3.5 MHz, 5.0 MHz, and 7.5 MHz. That was also shown in PRE-3 when compared with either PRE-1 or PRE-2. There was no significant difference in the backscatter parameters between PRE-1 and PRE-2 except for CAS at 7.5 MHz.

TABLE 2: Baseline characteristics of the participants at birth.

Gestational age group	PRE-1 < 28 weeks	PRE-2 28 – 31 <sup>+6</sup> weeks	PRE-3 32 – 36 <sup>+6</sup> weeks	TERM ≥ 37 weeks	Total
Number	25	139	195	146	505
Male	14 (56.0)	70 (50.4)	109 (55.9)	75 (51.4)	268 (53.1)
Gestational age (d)	191 (187, 193)	211 (204, 219)	240 (232, 248)	274 (266, 283)	239 (218, 263)
Birth weight (g)	990 (933, 1115)	1440 (1245, 1625)	2000 (1800, 2290)	3305 (2954, 3653)	1980 (1533, 2830)
Length (cm)	35 (34, 36)	38 (37, 40)	42 (41, 44)	49 (47, 50)	42 (39, 46)
Head circumference (cm)	25.0 (24.0, 25.0)	27.5 (26.0, 28.0)	30.0 (29.0, 31.0)	34.0 (33.0, 35.0)	30.0 (28.0, 33.0)
Alkaline phosphatase (IU/L)	261 (207, 317)	226 (173, 273)	204 (168, 238)	167 (139, 199)	197 (160, 246)
Calcium (mmol/L)	1.81 (1.66, 1.93)	1.98 (1.87, 2.12)	2.10 (1.99, 2.22)	2.17 (2.01, 2.29)	2.07 (1.94, 2.21)
Phosphorus (mmol/L)	2.06 (1.91, 2.24)	1.99 (1.78, 2.27)	1.86 (1.67, 2.03)	1.79 (1.62, 1.94)	1.88 (1.67, 2.09)

Data are n (%) or median (P<sub>25</sub>, P<sub>75</sub>).

TABLE 3: The values of ultrasonic backscatter parameters among different gestational age groups.

Gestational age group	PRE-1 < 28 weeks	PRE-2 28 – 31 <sup>+6</sup> weeks	PRE-3 32 – 36 <sup>+6</sup> weeks	TERM ≥ 37 weeks
AIB (dB)				
3.5 MHz	-51.64 (-54.36, -48.22)	-50.71 (-53.62, -46.78)	-44.54 (-48.42, -40.65) <sup>ab</sup>	-41.94 (-44.95, -37.94) <sup>abc</sup>
5.0 MHz	-55.78 (-58.61, -50.29)	-54.01 (-58.00, -49.56)	-50.84 (-54.22, -47.74) <sup>ab</sup>	-47.28 (-50.49, -44.85) <sup>abc</sup>
7.5 MHz	-51.29 (-53.66, -49.08)	-49.60 (-51.82, -47.85)	-48.28 (-50.27, -46.09) <sup>ab</sup>	-46.74 (-48.62, -43.98) <sup>abc</sup>
SCS (MHz)				
3.5 MHz	-0.24 (-0.35, -0.10)	-0.28 (-0.40, -0.19)	-0.43 (-0.51, -0.31) <sup>ab</sup>	-0.51 (-0.61, -0.42) <sup>abc</sup>
5.0 MHz	-0.35 (-0.50, -0.14)	-0.41 (-0.53, -0.29)	-0.63 (-0.74, -0.49) <sup>ab</sup>	-0.72 (-0.83, -0.59) <sup>abc</sup>
7.5 MHz	-0.26 (-0.43, -0.09)	-0.43 (-0.65, -0.29)	-0.67 (-0.85, -0.50) <sup>ab</sup>	-0.81 (-1.04, -0.63) <sup>abc</sup>
CAS				
3.5 MHz	-1.87 (-2.01, -1.71)	-1.76 (-1.93, -1.54)	-1.38 (-1.57, -1.20) <sup>ab</sup>	-1.17 (-1.38, -1.03) <sup>abc</sup>
5.0 MHz	-1.25 (-1.50, -1.10)	-1.16 (-1.36, -1.00)	-0.91 (-1.09, -0.71) <sup>ab</sup>	-0.68 (-0.86, -0.56) <sup>abc</sup>
7.5 MHz	-3.37 (-3.61, -3.18)	-3.12 (-3.34, -2.85) <sup>a</sup>	-2.78 (-2.96, -2.53) <sup>ab</sup>	-2.51 (-2.77, -2.31) <sup>abc</sup>

<sup>abc</sup>Data are median (P<sub>25</sub>, P<sub>75</sub>). Significantly different from PRE-1 ( $P < 0.001$ ). Significantly different from PRE-2 ( $P < 0.001$ ). Significantly different from PRE-3 ( $P < 0.001$ ).

Figure 2 is the scatterplot of AIB, SCS, and CAS associated with gestational age at 3.5 MHz, 5.0 MHz, and 7.5 MHz. Table 4 lists the correlation coefficients of the backscatter parameters with anthropometric indices and biochemical markers at all the frequencies.

Gestational age, birth weight, length, and head circumference were negatively correlated with SCS ( $|r| = 0.45 - 0.57$ ,  $P < 0.001$ ), and positively correlated with AIB ( $|r| = 0.36 - 0.60$ ,  $P < 0.001$ ) and CAS ( $|r| = 0.56 - 0.69$ ,  $P < 0.001$ ). Biochemical markers, especially serum phosphorus, yielded relatively weak correlations with backscatter parameters ( $|r| = 0.18 - 0.26$  for AKP,  $P < 0.001$ ;  $|r| = 0.17 - 0.34$  for calcium,  $P < 0.001$ ;  $|r| = 0.06 - 0.14$  for phosphorus,  $P < 0.05$  or not significant). In most cases, CAS had stronger correlations with the neonatal variables than AIB and SCS.

Table 5 shows the correlations between gestational age, anthropometric indices, and biochemical markers. Gestational age and anthropometric indices had strong positive correlations with each other ( $|r| = 0.86 - 0.96$ ,  $P < 0.001$ ) and weak to moderate correlations with biochemical markers ( $|r| = 0.17 - 0.43$ ,  $P < 0.001$ ). Correlations among biochemical markers were quite weak or nonsignificant.

To find the variables independently influencing CAS, gestational age, anthropometric indices, and biochemical markers were included in multiple regression analysis, as shown in Table 6. At 3.5 MHz, gestational age was the only variable that significantly contributed to the measurements of CAS ( $P < 0.001$ ). At 5.0 MHz, serum calcium ( $P = 0.040$ ) also made an independent contribution to CAS measurements besides gestational age ( $P < 0.001$ ). However, forward stepwise regression revealed that only gestational age was entered into the optimal model for predicting CAS at both 3.5 MHz and 5.0 MHz (Table 7). It could explain up to 44.3% and 40.5% of the variation of CAS in neonates, respectively. At 7.5 MHz, gestational age ( $P < 0.001$ ), head circumference ( $P = 0.002$ ), and serum calcium ( $P = 0.037$ ) were independent factors that influenced CAS measurements, and the combination could explain up to 40.3% of the variation.

## 4. Discussion

4.1. *Explanation of Ultrasonic Backscatter Parameters.* AIB is “apparent” backscatter parameter which represents

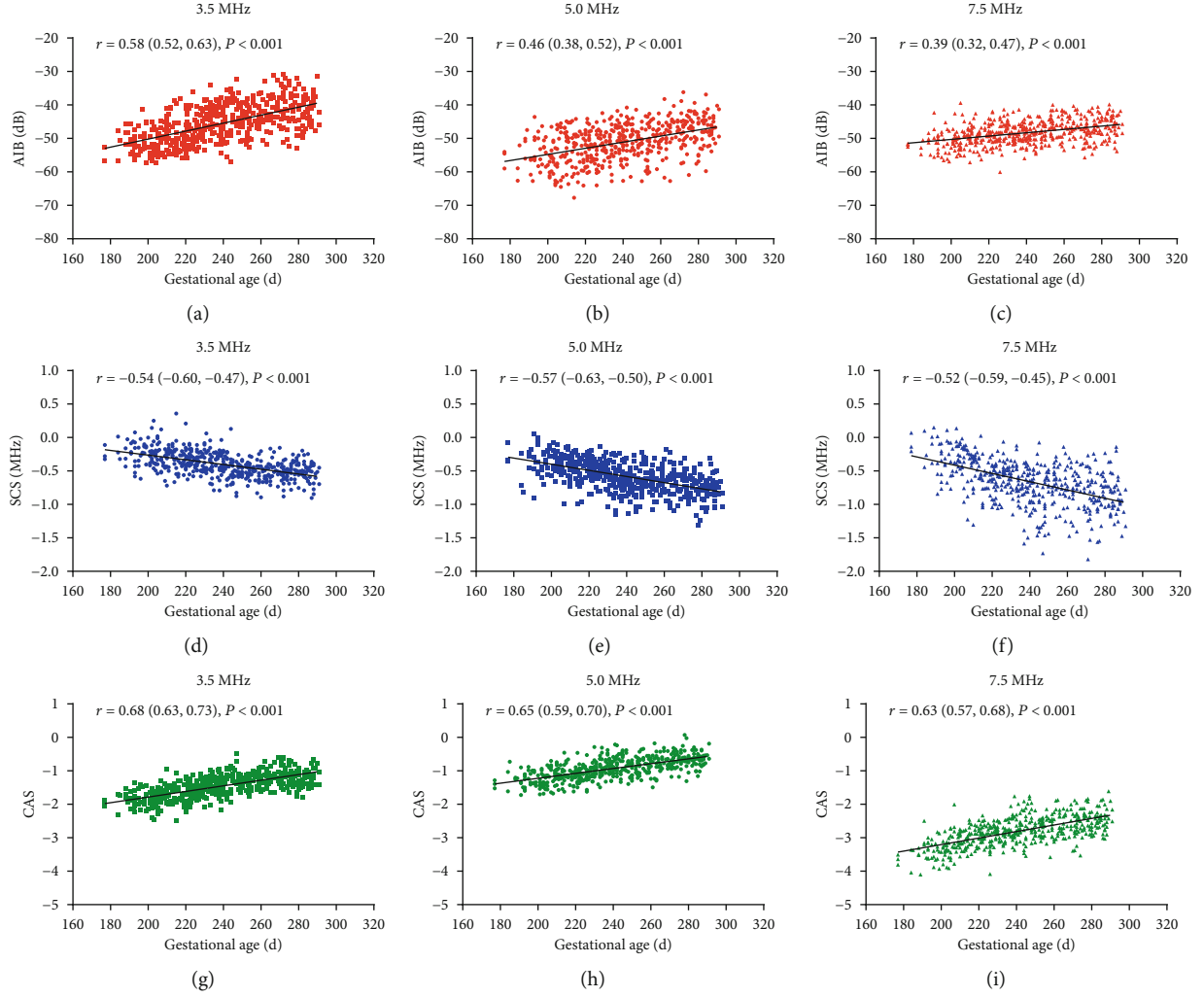


FIGURE 2: The scatterplots of AIB (a-c), SCS (d-f), and CAS (g-i) associated with gestational age at 3.5 MHz, 5.0 MHz, and 7.5 MHz ( $n = 505$ ). Gestational age is negatively correlated with SCS ( $|r| = 0.52 - 0.57$ ,  $P < 0.001$ ), and positively correlated with AIB ( $|r| = 0.39 - 0.58$ ,  $P < 0.001$ ) and CAS ( $|r| = 0.63 - 0.68$ ,  $P < 0.001$ ).

frequency-averaged backscatter power without compensation for attenuation in ultrasonic propagation [25, 27, 36, 37, 40]. AIB is convenient for in vivo measurements at a lower cost as it is unnecessary to measure the attenuation coefficient at the investigated position using transmission technique with two transducers. The values of AIB depend primarily on the comprehensive effects of backscatter and attenuation [36, 37, 41]. Attenuation is determined by the attenuation coefficient, as well as ultrasonic propagation length presented as  $t_d$  and  $t_w$  in this study. When propagation length is small, attenuation is weak and backscatter dominates the observed effects. Consequently, we selected short  $t_d$  and  $t_w$  for AIB. AIB is expected to be positively correlated with BMD in this case as the backscatter effects are more pronounced with higher BMD [33, 36, 37].

SCS is a downshift in the center frequency of the backscattered spectrum caused by attenuation within a scattering medium [42]. Stronger attenuation leads to larger magnitude of the downshift. Since the attenuation coefficient increases with BMD, the correlation between SCS and BMD is consis-

tently negative [31, 38, 39, 42]. Relatively long  $t_d$  and  $t_w$  are preferred in order to improve the sensitivity to detect low BMD [33].

CAS is a linear combination of AIB and SCS, varying in the same direction as AIB according to the formula. As backscatter is the dominant effect that influences AIB while attenuation mainly affects SCS, CAS reflects both backscatter and attenuation effects of cancellous bone. It may be a promising backscatter parameter to provide more information about bone status than AIB or SCS alone.

**4.2. Correlations with Gestational Age.** A large number of studies on bone specimens from adults or animals have confirmed AIB and SCS were not only significantly correlated with BMD obtained from DEXA or QCT [26, 27, 31, 38, 39, 42–44], but also provided complementary information about bone microstructure and mechanical properties such as bone volume fraction, trabecular separation, thickness, and number density [23, 25, 26, 45, 46]. In the present study, newborns with older gestational age had lower SCS and

TABLE 4: Correlation coefficients of backscatter parameters with anthropometric indices and biochemical markers.

	AIB			SCS			CAS		
	3.5 MHz	5.0 MHz	7.5 MHz	3.5 MHz	5.0 MHz	7.5 MHz	3.5 MHz	5.0 MHz	7.5 MHz
Birth weight									
Spearman $r$	0.60	0.47	0.39	-0.54	-0.56	-0.51	0.69	0.65	0.62
$P$ value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Length									
Spearman $r$	0.53	0.42	0.36	-0.52	-0.51	-0.45	0.63	0.59	0.56
$P$ value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Head circumference									
Spearman $r$	0.57	0.44	0.39	-0.54	-0.53	-0.50	0.67	0.62	0.60
$P$ value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Alkaline phosphatase									
Spearman $r$	-0.23	-0.20	-0.20	0.18	0.21	0.18	-0.25	-0.26	-0.26
$P$ value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Calcium									
Spearman $r$	0.31	0.28	0.17	-0.24	-0.28	-0.30	0.33	0.34	0.32
$P$ value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Phosphorus									
Spearman $r$	-0.11	-0.09	-0.14	0.09	0.13	0.06	-0.12	-0.14	-0.14
$P$ value	0.017	0.056	0.002	0.054	0.004	0.217	0.008	0.001	0.002

TABLE 5: Correlations between gestational age, anthropometric indices, and biochemical markers.

Variables	Spearman $r$						
	Gestational age	Birth weight	Length	Head circumference	Alkaline phosphatase	Calcium	Phosphorus
Gestational age	1	—	—	—	—	—	—
Birth weight	0.90 <sup>a</sup>	1	—	—	—	—	—
Length	0.86 <sup>a</sup>	0.93 <sup>a</sup>	1	—	—	—	—
Head circumference	0.91 <sup>a</sup>	0.96 <sup>a</sup>	0.91 <sup>a</sup>	1	—	—	—
Alkaline phosphatase	-0.38 <sup>a</sup>	-0.29 <sup>a</sup>	-0.28 <sup>a</sup>	-0.28 <sup>a</sup>	1	—	—
Calcium	0.43 <sup>a</sup>	0.39 <sup>a</sup>	0.35 <sup>a</sup>	0.38 <sup>a</sup>	-0.12 <sup>b</sup>	1	—
Phosphorus	-0.30 <sup>a</sup>	-0.17 <sup>a</sup>	-0.19 <sup>a</sup>	-0.21 <sup>a</sup>	0.13 <sup>b</sup>	-0.04 <sup>c</sup>	1

<sup>a</sup> $P < 0.001$ ; <sup>b</sup>  $P < 0.01$ ; <sup>c</sup> not significant,  $P = 0.338$ .

TABLE 6: Multivariate analysis of gestational age, anthropometric indices, and biochemical markers for the measurement of CAS.

Variables	3.5 MHz		5.0 MHz		7.5 MHz	
	Regression coefficient	$P$ value	Regression coefficient	$P$ value	Regression coefficient	$P$ value
Gestational age (d)	0.006	< 0.001	0.005	< 0.001	0.008	< 0.001
Birth weight (g)	0.00006	0.191	0.0001	0.100	-0.00001	0.813
Length (cm)	0.001	0.825	-0.002	0.751	-0.013	0.152
Head circumference (cm)	0.008	0.383	0.002	0.834	0.034	0.002
Alkaline phosphatase (IU/L)	-0.00004	0.855	0.000	0.407	0.000	0.159
Calcium (mmol/L)	0.075	0.227	0.122	0.040	0.168	0.037
Phosphorus (mmol/L)	0.049	0.157	0.016	0.628	0.065	0.147

higher AIB and CAS at birth, suggesting that term infants had better bone status than preterm infants. Moderate correlations were found between gestational age and the backscatter parameters which were also shown in previous studies

about neonates [32–34, 47]. There was a decreasing trend in the correlation coefficients between AIB and gestational age at higher frequencies which was likely attributed to heavy attenuation [36, 37, 41], but the frequency-dependent



TABLE 7: The optimal models for predicting CAS.

	Independent variables*	RMSE	Adjusted R <sup>2</sup>
3.5 MHz	GA	0.26	0.443
	0.008GA – 3.463		
5.0 MHz	GA	0.25	0.405
	0.007GA – 2.705		
7.5 MHz	GA, H, CA	0.33	0.403
	0.006GA + 0.028H + 0.172CA – 5.508		

GA gestational age, H head circumference, CA serum calcium, RMSE root mean square error of the regression, R<sup>2</sup> square of the adjusted correlation coefficient of the regression. \*  $P < 0.001$  for GA,  $P = 0.003$  for H,  $P = 0.033$  for CA.

variation of correlation coefficients was not obvious for SCS and CAS. The correlation coefficients of CAS were higher than that of AIB and SCS at all the frequencies, similar to the results of Tang and his colleagues [35]. It might be concluded that AIB, SCS and CAS were feasible to evaluate neonatal bone health at birth. CAS was probably more effective than AIB or SCS alone. Note that both BMD and bone mineral content (BMC) of newborns increased with gestational age [6], so it was reasonable to treat gestational age as an index of the degree of bone maturity.

**4.3. Correlations with Anthropometric Indices.** Anthropometric indices (birth weight, length and head circumference) were closely related to gestational age, reflecting fetal growth and maturation as well. Similarly, anthropometric indices had moderate correlations with AIB, SCS, and CAS and were able to reflect skeletal development. In accordance with gestational age, the correlation coefficients between AIB and anthropometric indices tended to decrease at higher frequencies though without significance. CAS had relatively stronger correlations with anthropometric indices than AIB and SCS at 3.5 MHz, 5.0 MHz, and 7.5 MHz. Among these anthropometric indices, head circumference was generally considered an index of neurodevelopment which was different from birth weight and length. However, Akcakus et al. [48] have reported positive correlations between head circumference and whole-body BMD and BMC of term infants at birth. Studies on ultrasonic backscatter technique also found that head circumference was significantly correlated with SCS, AIB, FIAB, and FSAB [33, 34].

**4.4. Correlations with Biochemical Markers.** Although serum calcium, phosphorus, and AKP were moderately correlated with gestational age, they were weakly or nonsignificantly correlated with ultrasonic backscatter parameters, consistent with Liu et al. [47] and our previous study [34]. Serum calcium might not be a useful marker for inadequate bone mineralization because its level usually remained normal in the early stage of MBD as a result of secondary hyperparathyroidism [49]. In contrast, the diagnostic power of serum phosphorus and AKP is still controversial. Some studies revealed that low phosphorus levels in combination with high AKP levels increased the sensitivity and specificity of MBD screening [5, 6], but there were also studies that

reported routine measurements of serum AKP and phosphorus were useless in predicting bone mineralization outcome in premature infants [50]. The validity of biochemical markers for assessing neonatal bone health required further researches.

**4.5. Optimal Predicting Models for CAS.** Simple linear regression demonstrated that gestational age, anthropometric indices, and biochemical markers were significantly associated with each other. Multiple regression demonstrated only gestational age independently contributed to the measurements of CAS at all the frequencies. As discussed above, gestational age increased with the degree of fetal maturity and positively correlated with BMD and BMC of newborns, so it was not surprising that gestational age played an important role in the predicting models of CAS. Anthropometric indices and biochemical markers were not independent factors influencing CAS mainly on account of multicollinearity. It was noteworthy that head circumference and serum calcium at 7.5 MHz were also entered into the predicting model. Ultrasound with higher frequencies (i.e., 7.5 MHz) provided a better resolution for the measurement of bone microstructure. Considering the tiny bone size of premature infants, the model at 7.5 MHz was supposed to achieve better performance in predicting backscatter properties. However, the optimal model at 7.5 MHz explained up to 40.3% of the variation of CAS, slightly lower than that of 3.5 MHz and 5.0 MHz. Therefore, the optimal frequency range for neonates and corresponding mechanisms deserved more attention in the future.

## 5. Limitations

One potential limitation of this study was the lack of direct indicators of bone status for neonates. None of the anthropometric indices or biochemical markers was completely reliable substitute for BMD and microstructural parameters. If there were comparative data that directly reflected bone status, the results would be more convincing. Moreover, follow-up studies remained to be conducted in view that MBD typically arose within 6 – 16 weeks after birth [1].

## 6. Conclusions

We performed ultrasonic backscatter measurements on 505 newborns at birth at 3.5 MHz, 5.0 MHz, and 7.5 MHz frequencies. The CAS, which was defined as a linear combination of AIB and SCS, was applied to neonates for the first time. Results indicated that AIB, SCS and CAS were significantly correlated with gestational age and anthropometric indices. CAS had relatively stronger correlations than AIB or SCS alone. Gestational age made significantly independent contributions to CAS at all the frequencies, and the optimal predicting models could explain up to 40.3% – 44.3% of the variation of CAS. This study suggested ultrasonic backscatter technique was feasible to evaluate neonatal bone status. CAS was a promising parameter to provide more information about bone health.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

This work was supported by the National Natural Science Foundation of China [grant numbers 81601504, 11874289, 11827808, 11525416].

## References

- [1] S. E. Rustico, A. C. Calabria, and S. J. Garber, "Metabolic bone disease of prematurity," *Journal of Clinical & Translational Endocrinology*, vol. 1, no. 3, pp. 85–91, 2014.
- [2] S. Ukarapong, S. K. B. Venkatarayappa, C. Navarrete, and G. Berkovitz, "Risk factors of metabolic bone disease of prematurity," *Early Human Development*, vol. 112, pp. 29–34, 2017.
- [3] S. Viswanathan, W. Khasawneh, K. McNelis et al., "Metabolic bone disease: a continued challenge in extremely low birth weight infants," *JPEN: Journal of Parenteral and Enteral Nutrition*, vol. 38, no. 8, pp. 982–990, 2014.
- [4] M. S. Fewtrell, T. J. Cole, N. J. Bishop, and A. Lucas, "Neonatal factors predicting childhood height in preterm infants: Evidence for a persisting effect of early metabolic bone disease?," *Journal of Pediatrics*, vol. 137, no. 5, pp. 668–673, 2000.
- [5] M. C. Backström, T. Kouri, A.-L. Kuusela et al., "bone isoenzyme of serum alkaline phosphatase and serum inorganic phosphate in metabolic bone disease of prematurity," *Acta Paediatrica*, vol. 89, no. 7, pp. 867–873, 2000.
- [6] C. M. Harrison and A. T. Gibson, "osteopenia in preterm infants," *Archives of Disease in Childhood: Fetal and Neonatal Edition*, vol. 98, no. 3, pp. F272–F275, 2013.
- [7] R. Barkmann, S. Dencks, P. Laugier et al., "Femur ultrasound (FemUS)—first clinical results on hip fracture discrimination and estimation of femoral BMD," *Osteoporosis International*, vol. 21, no. 6, pp. 969–976, 2010.
- [8] P. H. F. Nicholson and R. Alkalay, "Quantitative ultrasound predicts bone mineral density and failure load in human lumbar vertebrae," *Clinical biomechanics*, vol. 22, no. 6, pp. 623–629, 2007.
- [9] S. Chaffai, F. Peyrin, S. Nuzzo, R. Porcher, G. Berger, and P. Laugier, "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: relationships to density and microstructure," *Bone*, vol. 30, no. 1, article S8756328201006500, pp. 229–237, 2002.
- [10] K. A. Wear, "Nonlinear attenuation and dispersion in human calcaneus in vitro: statistical validation and relationships to microarchitecture," *The Journal of the Acoustical Society of America*, vol. 137, no. 3, pp. 1126–1133, 2015.
- [11] K. A. Wear, S. Nagaraja, M. L. Dreher, and S. L. Gibson, "Relationships of quantitative ultrasound parameters with cancellous bone microstructure in human calcaneus in vitro," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. 1605–1612, 2012.
- [12] M. A. Hakulinen, J. S. Day, J. Töyräs, H. Weinans, and J. S. Jurvelin, "Ultrasonic characterization of human trabecular bone microstructure," *Physics in Medicine and Biology*, vol. 51, no. 6, pp. 1633–1648, 2006.
- [13] G. Haiat, F. Padilla, and P. Laugier, "Sensitivity of QUS parameters to controlled variations of bone strength assessed with a cellular model," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 55, no. 7, pp. 1488–1496, 2008.
- [14] J. Töyräs, M. T. Nieminen, H. Kröger, and J. S. Jurvelin, "Bone mineral density, ultrasound velocity, and broadband attenuation predict mechanical properties of trabecular bone differently," *Bone*, vol. 31, no. 4, pp. 503–507, 2002.
- [15] C. Langton, "Prediction of mechanical properties of the human calcaneus by broadband ultrasonic attenuation," *Bone*, vol. 18, no. 6, pp. 495–503, 1996.
- [16] H.-L. Chen, H.-I. Tseng, S.-N. Yang, and R.-C. Yang, "bone status and associated factors measured by quantitative ultrasound in preterm and full-term newborn infants," *Early Human Development*, vol. 88, no. 8, pp. 617–622, 2012.
- [17] M. S. Fewtrell, K. L. Loh, S. Chomtho, K. Kennedy, J. Hawdon, and A. Khakoo, "Quantitative ultrasound (QUS): a useful tool for monitoring bone health in preterm infants?," *Acta Paediatrica*, vol. 97, no. 12, pp. 1625–1630, 2008.
- [18] J. Mercy, B. Dillon, J. Morris, A. J. Emmerson, and M. Z. Mughal, "Relationship of tibial speed of sound and lower limb length to nutrient intake in preterm infants," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 92, no. 5, pp. F381–F385, 2007.
- [19] H. McDevitt, "Quantitative ultrasound assessment of bone in preterm and term neonates," *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 90, no. 4, pp. F341–F342, 2005.
- [20] X.-p. Liao, W.-l. Zhang, J. He, J.-h. Sun, and P. Huang, "Bone measurements of infants in the first 3 months of life by quantitative ultrasound: the influence of gestational age, season, and postnatal age," *Pediatric Radiology*, vol. 35, no. 9, pp. 847–853, 2005.
- [21] E. Ritschl, K. Wehmeier, F. De Terlizzi et al., "Assessment of Skeletal Development in Preterm and Term Infants by Quantitative Ultrasound," *Pediatric Research*, vol. 58, no. 2, pp. 341–346, 2005.
- [22] A. Rubinacci, G. E. Moro, G. Boehm, F. De Terlizzi, G. L. Moro, and R. Cadossi, "Quantitative ultrasound for the assessment of osteopenia in preterm infants," *European Journal of Endocrinology*, vol. 149, no. 4, pp. 307–315, 2003.
- [23] Y. Li, B. Li, Y. Li et al., "The Ability of Ultrasonic Backscatter Parametric Imaging to Characterize Bovine Trabecular Bone," *Ultrasonic Imaging*, vol. 41, no. 5, pp. 271–289, 2019.
- [24] B. K. Hoffmeister, A. M. Viano, J. Huang et al., "Ultrasonic backscatter difference measurements of cancellous bone from the human femur: Relation to bone mineral density and microstructure," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3642–3653, 2018.
- [25] J. P. Karjalainen, J. Töyräs, O. Riekkinen, M. Hakulinen, and J. S. Jurvelin, "Ultrasound Backscatter Imaging Provides Frequency-Dependent Information on Structure, Composition and Mechanical Properties of Human Trabecular Bone," *Ultrasound in Medicine & Biology*, vol. 35, no. 8, pp. 1376–1384, 2009.
- [26] M. K. H. Malo, J. Töyräs, J. P. Karjalainen, H. Isaksson, O. Riekkinen, and J. S. Jurvelin, "Ultrasound backscatter

- measurements of intact human proximal femurs—Relationships of ultrasound parameters with tissue structure and mineral density,” *Bone*, vol. 64, pp. 240–245, 2014.
- [27] B. K. Hoffmeister, D. P. Johnson, J. A. Janeski et al., “Ultrasonic characterization of human cancellous bone in vitro using three different apparent backscatter parameters in the frequency range 0.6–15.0 mhz,” *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 55, no. 7, pp. 1442–1452, 2008.
  - [28] C. Roux, V. Roberjot, R. Porcher, S. Kolta, M. Dougados, and P. Laugier, “Ultrasonic Backscatter and Transmission Parameters at the Os Calcis in Postmenopausal Osteoporosis,” *Journal of Bone and Mineral Research*, vol. 16, no. 7, pp. 1353–1362, 2001.
  - [29] K. A. Wear and B. S. Garra, “Assessment of bone density using ultrasonic backscatter,” *Ultrasound in Medicine & Biology*, vol. 24, no. 5, pp. 689–695, 1998.
  - [30] J. Litniewski, L. Cieslik, M. Lewandowski, R. Tymkiewicz, B. Zienkiewicz, and A. Nowicki, “Ultrasonic scanner for in vivo measurement of cancellous bone properties from backscattered data,” *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 59, no. 7, pp. 1470–1477, 2012.
  - [31] B. S. Garra, M. Locher, S. Felker, and K. A. Wear, “Measurements of Ultrasonic Backscattered Spectral Centroid Shift From Spine In Vivo: Methodology and Preliminary Results,” *Ultrasound in Medicine & Biology*, vol. 35, no. 1, pp. 165–168, 2009.
  - [32] R. Zhang, D. Ta, C. Liu, and C. Chen, “Feasibility of bone assessment with ultrasonic backscatter signals in neonates,” *Ultrasound in Medicine & Biology*, vol. 39, no. 10, pp. 1751–1759, 2013.
  - [33] C. Liu, R. Zhang, Y. Li, F. Xu, D. Ta, and W. Wang, “An Ultrasonic Backscatter Instrument for Cancellous Bone Evaluation in Neonates,” *Engineering*, vol. 1, no. 3, pp. 336–343, 2015.
  - [34] W. Mao, Y. Du, C. Liu et al., “Ultrasonic Backscatter Technique for Assessing and Monitoring Neonatal Cancellous Bone Status in Vivo,” *IEEE Access*, vol. 7, pp. 157417–157426, 2019.
  - [35] T. Tang, C. Liu, F. Xu, and D. Ta, “Correlation between the combination of apparent integrated backscatter–spectral centroid shift and bone mineral density,” *Journal of Medical Ultrasonics*, vol. 43, no. 2, pp. 167–173, 2016.
  - [36] B. K. Hoffmeister, J. A. Mcpherson, M. R. Smathers, P. L. Spinolo, and M. E. Sellers, “Ultrasonic backscatter from cancellous bone: the apparent backscatter transfer function,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 62, no. 12, pp. 2115–2125, 2015.
  - [37] B. K. Hoffmeister, “Frequency dependence of apparent ultrasonic backscatter from human cancellous bone,” *Physics in Medicine and Biology*, vol. 56, no. 3, pp. 667–683, 2011.
  - [38] C. Liu, F. Xu, D. Ta et al., “Measurement of the Human Calcaneus In Vivo Using Ultrasonic Backscatter Spectral Centroid Shift,” *Journal of Ultrasound in Medicine*, vol. 35, no. 10, pp. 2197–2208, 2016.
  - [39] Y.-q. Jiang, C.-c. Liu, R.-y. Li et al., “Analysis of Apparent Integrated Backscatter Coefficient and Backscattered Spectral Centroid Shift in Calcaneus in vivo for the Ultrasonic Evaluation of Osteoporosis,” *Ultrasound in Medicine and Biology*, vol. 40, no. 6, pp. 1307–1317, 2014.
  - [40] B. K. Hoffmeister, P. L. Spinolo, M. E. Sellers, P. L. Marshall, A. M. Viano, and S.-R. Lee, “effect of intervening tissues on ultrasonic backscatter measurements of bone: an in vitro study,” *The Journal of the Acoustical Society of America*, vol. 138, no. 4, pp. 2449–2457, 2015.
  - [41] C. Liu, T. Tang, F. Xu et al., “Signal of Interest Selection Standard for Ultrasonic Backscatter in Cancellous Bone Evaluation,” *Ultrasound in Medicine & Biology*, vol. 41, no. 10, pp. 2714–2721, 2015.
  - [42] K. A. Wear, “Characterization of trabecular bone using the backscattered spectral centroid shift,” *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 50, no. 4, pp. 402–407, 2003.
  - [43] J. P. Karjalainen, O. Riekkinen, J. Töyräs et al., “Multi-site bone ultrasound measurements in elderly women with and without previous hip fractures,” *Osteoporosis International*, vol. 23, no. 4, pp. 1287–1295, 2012.
  - [44] B. K. Hoffmeister, C. I. Jones, G. J. Caldwell, and S. C. Kaste, “Ultrasonic characterization of cancellous bone using apparent integrated backscatter,” *Physics in Medicine and Biology*, vol. 51, no. 11, pp. 2715–2727, 2006.
  - [45] C. Liu, B. Li, Q. Diwu et al., “Relationships of Ultrasonic Backscatter With Bone Densities and Microstructure in Bovine Cancellous Bone,” *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 12, pp. 2311–2321, 2018.
  - [46] X. Chou, F. Xu, Y. Li, C. Liu, D. Ta, and L. H. Le, “Variability in Ultrasound Backscatter Induced by Trabecular Microstructure Deterioration in Cancellous Bone,” *BioMed Research International*, vol. 2018, 7 pages, 2018.
  - [47] C. Liu, B. Li, Y. Li et al., “Ultrasonic Backscatter Difference Measurement of Bone Health in Preterm and Term Newborns,” *Ultrasound in Medicine & Biology*, vol. 46, no. 2, pp. 305–314, 2020.
  - [48] M. Akcaks, E. Koklu, S. Kurtoglu, M. Kula, and S. Koklu, “The Relationship among Intrauterine Growth, Insulinlike Growth Factor I (IGF-I), IGF-Binding Protein-3, and Bone Mineral Status in Newborn Infants,” *American Journal of Perinatology*, vol. 23, no. 8, pp. 473–480, 2006.
  - [49] Y.-L. Hung, P.-C. Chen, S.-F. Jeng et al., “Serial measurements of serum alkaline phosphatase for early prediction of osteopaenia in preterm infants,” *Journal of Paediatrics and Child Health*, vol. 47, no. 3, pp. 134–139, 2011.
  - [50] J. Faerk, “Bone mineralisation in premature infants cannot be predicted from serum alkaline phosphatase or serum phosphate,” *Archives of Disease in Childhood - Fetal and Neonatal Edition*, vol. 87, no. 2, pp. 133F–1136, 2002.

## Research Article

# Epilepsy Detection in EEG Using Grassmann Discriminant Analysis Method

Hongbin Yu <sup>1</sup>, Chao Fan <sup>1</sup> and Yunting Zhang<sup>2</sup>

<sup>1</sup>School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214000, China

<sup>2</sup>Institute of Commerce of Digital Art Animation, Wuxi 214000, China

Correspondence should be addressed to Hongbin Yu; alexander.yuu@jiangnan.edu.cn

Received 16 January 2020; Accepted 17 March 2020; Published 1 May 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Hongbin Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Epilepsy is marked by seizures stemming from abnormal electrical activity in the brain, causing involuntary movement or behavior. Many scientists have been working hard to explore the cause of epilepsy and seek the prevention and treatment. In the field of machine learning, epileptic diagnosis based on EEG signal has been a very hot research topic; many methods have been proposed, and considerable progress has been achieved. However, resorting the epileptic diagnosis techniques based on EEG to the reality applications still faces many challenges. Low signal-to-noise ratio (SNR) is one of the most important methodological challenges for EEG data collection and analysis. This paper discusses an automated diagnostic method for epileptic detection using a Fréchet Mean embedded in the Grassmann manifold analysis. Fréchet mean-based Grassmann discriminant analysis (FMGDA) algorithm to implement the EEG data dimensionality reduction and clustering task. The method is resorted to reduce Grassmann data from high-dimensional data to a relative lower-dimensional data and maximize between-class distance and minimize within-class distance simultaneously. Every EEG feature is mapped into the Grassmann manifold space first and then resort the Fréchet mean to represent the clustering center to carry out the clustering work. We designed a detailed experimental scheme to test the performance of our proposed algorithm; the test is assessed on several benchmark datasets. Experimental results have delivered that our approach leads to a significant improvement over state-of-the-art Grassmann manifold methods.

## 1. Introduction

Epilepsy is a critical neurological disease that is caused by temporary abnormal discharges of the brain electrical activity, leading to uncontrollable movements and tremulous [1]. People with epilepsy are two or three times more likely to die early than the normal person [2]. It was estimated that approximately 1% people in the world suffer from epilepsy, and elderly patients are the majority [3]. Hence, epilepsy detection is of great significance in clinical therapy of epileptic patients.

Electroencephalogram (EEG) is most commonly used in epilepsy detection manner, since it contains valuable physiological information about the brain; it is also a valuable clinical tool for epilepsy evaluation and treatment [4]. Many automated diagnostic systems for epilepsy have been developed based on different technologies. In the area of artificial intelligence, many researchers have been focusing on the

research of EEG data dimensionality reduction and discriminant feature extracting from the raw EEG data as well as exploring high-performance clustering model to implement the epilepsy detection task.

Many feature extraction methods have been developed for the epilepsy detection including time-domain [5, 6], frequency-domain [7–12], time-frequency analysis [13], energy distribution in the time-frequency plane [14], wavelet features [15], and chaotic features such as entropies [6, 16]. There are still many classification methods that have been proposed to solve the epilepsy detection problems in recent years. Dehuri et al. proposed to use DE-RBFN method to implement the epilepsy detection task; EEG signals are decomposed with wavelet transform into different subbands, and statistical information is extracted from the wavelet coefficients to supply as the input to ensemble of DE-RBFNs [1]. Li et al. proposed to use the discrete wavelet transform



(DWT) in combination with EA method to extract significant features from raw EEG signals, and an effective network model called NNE is designed specifically to the task of epilepsy detection [2]. Jiang et al. propose to use multiple feature extraction method to obtain the multiview feature from the raw EEG, and the classical Takagi-Sugeno-Kang fuzzy system is used to implement the classification task [17]. Zhang et al. extracted the entropy as the feature and combine the SVM classifier to estimate the epileptic cases based on the EEG signal [18].

However, the aforementioned algorithms for epilepsy detection based on EEG data are all concerned in terms of Euclidean alike distance. In order to extend the study of EEG signal to non-Euclidean space and solve these problems existing in the EEG applications that the traditional Euclidean algorithms cannot solve, in this paper, we will cast the EEG processing problem into the Grassmann manifold to implement the epilepsy detection task.

Grassmann manifold have been a popular method in recent years for its strong capability in extracting discriminative information for image sets and videos, where the datasets are vectors instead of the normal used real value vectors. For instance, in the study of computer visions, except the traditional well-structured data, such as image data as well as video data, there exist some manifold-valued data. The movement of scattered key points in the video can be described by subspaces, i.e., the points on the so-called Grassmann manifold and the covariance feature descriptors of images are SPD manifold-valued data [19–21]. Thus, the clustering problem is completely different from those which instances are clustered into clusters according to their space structure in the euclidean spaces. However, in the Grassmann manifold space, the objects to be clustered themselves are subspaces (of the same dimension), i.e., the points on the abstract Grassmann manifold [22] which is completely different from the traditional Euclidean space with data vector form. A Grassmann manifold  $\mathcal{G}(p, D)$  is the space of all  $p$ -dimensional linear subspaces of  $\mathbf{R}^D (0 \leq p \leq D)$ . For  $p = 0$ , the Grassmann manifold becomes the Euclidean space itself. As Grassmann manifold is abstract, there are a number of ways to realize it, such as full column rank matrix representation method [19], orthogonal representation [23], symmetric idempotent matrix representation [24], and Stiefel manifold representation method [23]. Through those methods, some existing algorithms which are developed in the Euclidean space can be extended to the Grassmann manifold. Hamm and Lee [25] employed the projection metric to encode the Grassmannian points by Grassmannian kernels and developed Grassmann discriminant analysis on the kernel space. Further, Hamm and Lee [26] theoretically studied the relationship between projection kernel and the KL distance. Harandi et al. [27] proposed a graph embedding-based discriminant analysis approach on Grassmannian manifold which aims to simultaneously maximize discriminant power and preserve the geometrical structure of the manifold. Huang et al. [28] employed the projection metric to learn discriminant transformation on the Grassmann manifold. Wang et al. [21] extended classical LPP from the Euclidean space to Grassmann manifold. We adopt the Stie-

fel manifold representation strategy for the learning task in the Grassmann manifold space and proposed a Fréchet mean-based Grassmann discriminant analysis (FMGDA) algorithm for image sets recognition which has been presented in [29]. Comparing with the previously proposed Grassmann manifold algorithms, our proposed Fréchet mean algorithm has two important innovations: firstly, we solve the feature extraction problems by minimizing the within-class mean and maximizing the between-class mean simultaneously for Grassmannian points. Secondly, our proposed method generalized the classic LDA to non-Euclidean Grassmann manifolds, and our optimization problem can be characterized by the trace ratio problem.

EEG signal is an effective tool to study the firing mechanism of cortical neurons; however, due to intrinsic nature of lower SNR ratio, applications based on EEG signal are severely limited. To solve these problems that cannot be solved in the ordinary Euclidean space, we proposed to map the EEG features to the Grassmann manifold to find some clues to solve those problems and enhance the epilepsy diagnosis accuracy simultaneously. Most of the previously used algorithms for epilepsy detection are carried out in the Euclidean space, in order to extend the scope of EEG research and provide more ideas to solve the problems of epilepsy detection based on EEG signal; we proposed to implement the epilepsy detection task in the Grassmann manifold to do some exploratory research work.

Inspired by the already existing video classification research work in the Grassmann manifold space, we transform the ordinary epilepsy detection task into an image classification task, where the energy distribution in the time-frequency plane of EEG signal epochs is viewed as an image; thus, the epilepsy detection based on the EEG can be realized by implementing an image classification task. We adopt the FMGDA method to implement the epilepsy detection task. The spectrum features of each epoch EEG time series are extracted by short-time Fourier transform (STFT) first and get a spectrum matrix. The matrix is then used to construct a dataset vector, just as the video frames. The FMGDA algorithm finally carries out to classify each dataset.

We summarize our contributions as follows:

- (i) The energy distribution in the time-frequency plane of EEG signal can be used as the image for EEG classification
- (ii) EEG classification problem can be transformed to an image clustering problem in the Grassmann manifold space
- (iii) The Fréchet mean of EEG energy distribution in the time-frequency plane can be used to characterize the clustering center in the Grassmann manifold space

The paper is organized as follows: firstly, we review the traditional linear discriminant analysis method in Section 2. In Section 3, we introduce our Fréchet mean-based Grassmann discriminant analysis (FMGDA) method. In Section 4, we evaluate our proposed algorithms' performance on the epilepsy detection. In Section 5, we conclude our work.



## 2. Fréchet Mean-Based Grassmann Discriminant Analysis

In this section, we will introduce our Fréchet mean-based Grassmann discriminant analysis (FMGDA) algorithm which was proposed in [29]. We will use our FMGDA algorithm to implement the epilepsy detection task in the subsequent section.

**2.1. Linear Discriminant Analysis.** Linear discriminant analysis has been proposed for several years and gained considerable attentions for its superiority on the dimensionality reduction, feature extraction, and the classification research.

Given  $n$  samples  $x_1, x_2, x_3, \dots, x_n$  from  $c$  classes. LDA is to find a linear transformation which can maximize the between-class distance and minimize the within-class distance simultaneously in the transformed subspace. In other words, the linear mapping  $a$  can be obtained by solving the following optimization problem.

$$\hat{a} = \operatorname{argmax} \frac{a^T S_b a}{a^T S_w a}, \quad (1)$$

where

$$\begin{aligned} S_b &= \sum_{k=1}^c n_k (\mu^{(k)} - \mu) (\mu^{(k)} - \mu)^T, \\ S_w &= \sum_{k=1}^c \sum_{i=1}^{n_k} n_k (x_i^{(k)} - \mu^{(k)}) (x_i^{(k)} - \mu^{(k)})^T. \end{aligned} \quad (2)$$

Here,  $\mu^{(k)}$  denotes the mean vector of the  $k$ th class, and  $\mu$  denotes the centroid of all the sample instances.  $x_i^{(k)}$  is the  $i$ th instance from the  $k$ th class, and  $n_k$  denotes the instance number of the  $k$ th class. Matrices  $S_b$  and  $S_w$  are often called the between-class scatter matrix and within-class scatter matrix, respectively.

Substituting the sample instances into the formula, with some simple algebra manipulations, we can obtain the between-class scatter matrix and within-class scatter matrix as follows:

$$\begin{aligned} a^T S_b a &= \sum_{k=1}^c n_k a^T (\mu^{(k)} - \mu) (\mu^{(k)} - \mu)^T a, \\ &= \sum_{k=1}^c n_k \left\| a^T \mu^{(k)} - a^T \mu \right\|_2^2, \\ &= \sum_{k=1}^c n_k \delta_E(a^T \mu^{(k)}, a^T \mu), \\ a^T S_w a &= \sum_{k=1}^c \sum_{i=1}^{n_k} n_k \left\| a^T (x_i^{(k)} - \mu^{(k)}) \right\|_2^2, \\ &= \sum_{k=1}^c \sum_{i=1}^{n_k} n_k \left\| a^T x_i^{(k)} - a^T \mu^{(k)} \right\|_2^2, \\ &= \sum_{k=1}^c \sum_{i=1}^{n_k} n_k \delta_E(a^T x_i^{(k)}, a^T \mu^{(k)}), \end{aligned} \quad (3)$$

where the notation  $\delta_E(\cdot, \cdot)$  denotes the Euclidean distance between two regular data vectors. The solution of (1) is the generalized eigenvectors corresponding to the largest eigenvalues of the following:

$$S_b a = \lambda S_w a. \quad (4)$$

Thus, mapping vector  $a$  can be obtained by adopting the eigen-decomposition method on the matrix  $S_b^{-1} S_w$ , if  $S_b$  is nonsingular. There are at most  $c-1$  eigenvectors corresponding to nonzero values, since the rank of  $S_b$  is bounded from above by  $c-1$ . Thus, the reduced dimension by LDA is at most  $c-1$ .

**2.2. Grassmann Manifold.** In this section, we provide a brief summary of the basic Riemannian geometry of Grassmann manifold. More details can be found in [23, 30, 31]. A Grassmann manifold  $\mathcal{G}(p, D)$  is the space of all  $p$ -dimensional linear subspaces of  $\mathbb{R}^D$  ( $0 \leq p \leq D$ ). When  $p = 0$ , the Grassmann manifold becomes the Euclidean space itself. When  $p = 1$ , the Grassmann manifold consists of all the lines passing through the origin in  $\mathbb{R}^d$ . As Grassmann manifold is abstract, there are a number of ways to realize it for numerical learning purpose.

Assuming that  $\mathbb{R}_*^{d \times p}$  be the space of all  $d \times p$  matrix of full column rank,  $GL(p)$  denote the general group of nonsingular matrices of order  $p$  and  $\mathcal{O}(p)$  the group of all the  $p$ times orthogonal matrices.

(i) Representation by full column rank matrices [19]

$$\mathcal{G}(p, d) \cong \mathbb{R}_*^{d \times p} / GL(p) \quad (5)$$

(ii) The orthogonal representation [23]:

$$\mathcal{G}(p, d) \cong \mathcal{O}(d) / \mathcal{O}(p) \times \mathcal{O}(d-p) \quad (6)$$

(iii) Symmetric idempotent matrix representation [24]:

$$\mathcal{G}(p, d) \cong \left\{ P \in \mathbb{R}^{d \times d} : P^T = P, P^2 = P, \operatorname{rank}(P) = p \right\}. \quad (7)$$

(iv) The Stiefel manifold representation [25];

$$\mathcal{G}(p, d) \cong \mathcal{ST}(p, d) / \mathcal{O}(p) \quad (8)$$

where  $\mathcal{ST}(p, d) = \{X \in \mathbb{R}^{d \times p} : X^T X = I_p\}$ .

In our proposed FMGDA, we adopt the Stiefel manifold representation strategy to complete our Grassmann manifold-based algorithm. A point  $X$  on the Grassmann manifold  $\mathcal{G}(p, D)$  is a subspace spanned by the orthonormal columns of a  $D \times p$  matrix  $X$  such that  $X^T X = I_p$ , where  $I_p$  is the identity matrix of size  $p \times p$ .

Grassmann manifold has a nice property that it can be embedded into a space consisting of symmetric positive

semidefinite matrices. More precisely, let  $X \in \mathcal{G}(p, D)$ , we can define the following projection embedding

$$\Pi : \mathcal{G}(p, D) \longrightarrow \text{Sym}_+(D), \Pi(X) = XX^T, \quad (9)$$

where  $\text{Sym}_+(D)$  denotes the space of  $D \times D$  symmetric positive semidefinite matrices. Since  $\text{Sym}_+(D)$  can be understood as a Euclidean space, a natural metric for  $\text{Sym}_+(D)$  is the Frobenius norm. As such, we can define the following projection metric [25]:

$$\delta_p(X_1, X_2) = \frac{1}{\sqrt{2}} \|\Pi(X_1) - \Pi(X_2)\|_F^2, \quad (10)$$

where  $X_1$  and  $X_2$  are two Grassmann points and  $\Pi(X_i) = X_i X_i^T, i = 1, 2$ . As pointed in [32], the projection metric  $\delta_p(\cdot, \cdot)$  is able to approximate the true Grassmannian geodesic distance and become one of the most popular metrics for analyzing Grassmann manifold features [25, 27, 33].

### 3. FMGDA

In this section, we propose FMGDA, a supervised subspace learning for Grassmann manifold that maps a high-dimensional Grassmann point to a lower-dimensional Grassmann manifold.

Suppose we have a data set  $(\mathcal{X}, Y)$ , where  $\mathcal{X} = \{X_i\}_{i=1}^n$ ,  $X_i \in \mathcal{G}(p, D)$  is a Grassmann point,  $Y = [y_1, y_2, \dots, y_n]$  is the class indicator matrix.  $y_i(j) = 1$  if  $X_i$  belongs to the  $j$ th class and 0 otherwise. Our purpose is to learn the parameter  $A \in \mathbf{R}^{D \times d}$  of a mapping in the form  $f: X_i \in \mathcal{G}(p, D) \longrightarrow \mathcal{G}(p, d)$ , which is defined as:

$$f(X_i, A) = A^T X_i. \quad (11)$$

With this mapping  $f$ , the original high-dimensional Grassmann manifold can be transformed into a lower-dimensional Grassmann manifold. However,  $A^T X_i$  is not a valid Grassmann point since the parameter  $A$  is not an orthogonal matrix. To solve this problem, we temporarily employ the orthonormal components of  $A^T X_i$  defined by  $A^T X_i'$  to represent an orthonormal basis matrix, the transformed projection matrix. We now rewrite  $f(X_i, A) = A^T X_i$  to make  $f(X_i, A)$  a valid Grassmann point. The approach to get  $A^T X_i'$  will be thoroughly discussed in Section 4.

Let  $X^{(k)} = [X_1^{(k)}, X_2^{(k)}, \dots, X_{n_k}^{(k)}]$  be a set of Grassmann points from the  $k$ th class. Recalling the definition of LDA, we require the mean to capture the discriminant information. However, traditional Euclidean mean is not a valid Grassmann point. Therefore, special care must be taken into account to compute the subspace mean for Grassmann manifold. Fortunately, subspace mean on the Grassmann manifold has been studied in [34–36]. In particular, the Fréchet mean is commonly used to characterize the subspace mean of Grassmann manifold.

**Definition 1.** The Fréchet mean  $M^*$  for a set of points  $\{X_i\}_{i=1}^n, X_i \in \mathcal{G}(p, D)$  is the local minimizer of the cost function

$$M^* = \arg \min_M \sum_{i=1}^n \delta_p(X_i, M). \quad (12)$$

The above definition shows that the subspace mean depends heavily on the metric. If we assume all points come from Euclidean space and choose the Euclidean metric, the Fréchet mean has a closed form solution which is nothing but the traditional mean. Unfortunately, there is usually no closed solution for  $M^*$  with Riemannian metric, and the first-order gradient descent method [22] is commonly employed to find the solution. For Grassmann data points endowed with the projection metric, we have an analytic solution for the Fréchet mean which is characterized in the following lemma.

**Lemma 2.** The Fréchet mean  $M^*$  for a set of points  $\{X_i\}_{i=1}^n, X_i \in \mathcal{G}(p, D)$  is the  $p$  largest eigenvectors of  $\sum_{i=1}^n X_i X_i^T$ .

Let  $M^{(k)}$  be the class Fréchet mean of the  $k$ th samples  $X^{(k)}$  and  $M$  be the total Fréchet mean of  $\mathcal{X}$ . Similar to LDA, the within-class distance and between-class distance in the transformed low-dimensional Grassmann manifold are defined as

$$\begin{aligned} d_w(A) &= \sum_{k=1}^K \sum_{i=1}^{n_k} n_k \delta_p(f(X_i^{(k)}, A), f(M^{(k)}, A)), \\ d_b(A) &= \sum_{k=1}^K n_k \delta_p(f(M^{(k)}, A), f(M, A)), \end{aligned} \quad (13)$$

where  $f(X_i^{(k)}, A) = A^T X_i'$ ,  $f(M, A) = A^T M'$ , and  $f(M^{(k)}, A) = A^T M'^{(k)}$ . Note that  $A^T M'$  and  $A^T M'^{(k)}$  are the orthonormal components of  $A^T M$  and  $A^T M^{(k)}$ , respectively. By the definition of  $\delta_p(\cdot, \cdot)$ , we can explicitly write  $d_w(A)$  and  $d_b(A)$  as follows:

$$\begin{aligned} d_w(A) &= \sum_{k=1}^K \sum_{i=1}^{n_k} n_k \left\| A^T X_i'(k) X_i'(k)^T A - A^T M'^{(k)} M'^{(k)T} A \right\|_F^2, \\ d_b(A) &= \sum_{k=1}^K n_k \left\| A^T M'^{(k)} M'^{(k)T} A - A^T M' M'^T A \right\|_F^2. \end{aligned} \quad (14)$$

It should be pointed out that  $A^T M'^{(k)}$  does not exactly express the Fréchet mean of  $X^{(k)}$  in the lower dimensional manifold, but expresses it only approximately. Following the idea of LDA, we now arrive at the objective of FMGDA

$$\max_A \frac{d_b(A)}{d_w(A)}, \quad (15)$$

**Input:** Grassmann points  $\{X\}_{i=1}^n, X_i \in \mathcal{G}_{p,D}$   
**Output:** The mapping matrix  $A \in \mathbf{R}^{D \times d}$   
1: Initialize: Set the parameter  $A^{(0)}$   
2: Compute the class and total Fréchet mean  $M^{(k)}, M$  respectively according to Lemma 2.  
3: **While** not converged **do**  
4:   Normalized  $X_i^{(k)}, M, M^{(k)}$  according to **QR** decomposition  
5:   Compute  $B^{(k)}$  and  $Q_{ik}$   
6:   Solve the trace ration problem (20)  
7: **end while**

ALGORITHM 1: FMGDA algorithm.

which aims to maximize the between-class distance and minimize the within-class distance simultaneously.

#### 4. Iterative Optimization

The optimization problem (15) includes four variables  $A, X_i, M', M^{(k)}$  which is hard to find a closed solution. In the following, we propose an iterative solution for one of the four variables at a time by fixing the other and repeating for a certain number of iterations.

We follow the work in [21, 28] to obtain the orthonormal components of  $A^T X_i'(k), A^T M',$  and  $A^T M^{(k)}$ . Specifically, let  $A^T X_i^{(k)} = Q_{x_i} R_{x_i}$  be the QR decomposition, where  $Q_{x_i} \in \mathbf{R}^{d \times p}$  is an orthogonal matrix and  $R_{x_i} \in \mathbf{R}^{p \times p}$  is a nonsingular upper-triangular matrix. It is easy to show that  $Q_{x_i} = A^T X_i'(k)$ , where  $X_i'(k) = X_i^{(k)} R_{x_i}^{-1}$ . Note that  $Q_{x_i}$  and  $A^T X_i^{(k)}$  represent the same subspace and  $Q_{x_i}$  is orthonormal. Similar normalization procedure can be applied to get the other two variables  $M'$  and  $M^{(k)}$ .

Let  $B^{(k)} = M^{(k)} M'^{(k)T} - M' M'^T$ , we have

$$\begin{aligned} d_b(A) &= \sum_{k=1}^K n_k \left\| A^T M'^{(k)} M'^{(k)T} A - A^T M' M'^T A \right\|_F^2, \\ &= \sum_{k=1}^K n_k \left\| A^T B^{(k)} A \right\|_F^2, \\ &= \sum_{k=1}^K n_k \text{tr} \left( A^T B^{(k)} A A^T B^{(k)} A \right). \end{aligned} \quad (16)$$

Similarly, let  $Q_{ik} = X_i'(k) X_i'(k)^T - M'^{(k)} M'^{(k)T}$ , we have

$$\begin{aligned} d_w(A) &= \left\| A^T X_i'(k) X_i'(k)^T A - A^T M'^{(k)} M'^{(k)T} A \right\|_F^2, \\ &= \sum_{k=1}^K \sum_{i=1}^{n_k} n_k \left\| A^T Q_{ik} A \right\|_F^2, \\ &= \sum_{k=1}^K \sum_{i=1}^{n_k} n_k \text{tr} \left( A^T Q_{ik} A A^T Q_{ik} A \right). \end{aligned} \quad (17)$$

We now define a new objective  $g_t(A)$  in the  $t$ th iteration by using the last step  $A^{(t-1)}$  as follows

$$g_t(A) = \frac{\text{tr} \left( A^T \tilde{B}^{(t-1)} A \right)}{\text{tr} \left( A^T \tilde{Q}^{(t-1)} A \right)}, \quad (18)$$

where

$$\begin{aligned} \tilde{B}^{(t-1)} &= \sum_{k=1}^K n_k B^{(k)} A^{(t-1)} A^{(t-1)T} B^{(k)}, \\ \tilde{Q}^{(t-1)} &= \sum_{k=1}^K \sum_{i=1}^{n_k} n_k Q_{ik} A^{(t-1)} A^{(t-1)T} Q_{ik}. \end{aligned} \quad (19)$$

Clearly, the solution of  $t$ th iteration can be stated as

$$\max_A \frac{\text{tr} \left( A^T \tilde{B}^{(t-1)} A \right)}{\text{tr} \left( A^T \tilde{Q}^{(t-1)} A \right)}, \quad (20)$$

which is a trace ratio optimization problem has been extensively studied in [37–39]. In this paper, we use the algorithm proposed in [37] to solve the optimization problem (20). The whole procedure of FMGDA is summarized in Algorithm 1.

#### 5. Experiments

In this section, we will use our FMGDA to carry out the epilepsy detection task based on the EEG data. To evaluate the performance of our FMGDA, four other Grassmann manifold algorithms are introduced, and the classification accuracy comparison results are presented.

Each comparison experiments are run for 5 rounds; 10%, 20%, 40%, 60%, and 80% of all the data are used as the training data, respectively; each round are repeated for 10 times; the average accuracy results are reported.

**5.1. Experimental Setup.** The dataset we used to evaluate the algorithm are publicly available on the web from the University of Bonn, Germany (<http://www.meb.unibonn.de/epileptologie/science/physik/eeegdata.html>).

TABLE 1: Dataset description.

Subject	Groups	Size	Dataset description
Healthy	A	100	EEG signals from healthy people with eyes open
	B	100	EEG signals from healthy people with eyes closed
Epileptic	C	100	EEG signals obtained in the hippocampal formation of the opposite hemisphere of the brain during seizure-free intervals
	D	100	EEG signals obtained from within the epileptogenic zone during seizure-free intervals
	E	100	EEG signals measured during seizure

The complete data archive contains five groups of data (denoted by groups A to E), each group containing 100 single-channel EEG segments of 23.6 duration. The sampling rate of all data is 173.6 Hz. Groups A and B consist of segments acquired from surface EEG recording performed on five healthy volunteer subjects using standardized electrode placement scheme.

Recording was carried out when the subjects were relaxed in awoken state with eyes open (group A) and eyes closed (group B), respectively. Group C, group D, and group E are obtained from volunteer subjects with epilepsy. EEG signals in group C were recorded from the hippocampal formation of the opposite hemisphere of brain, while those in group D were measured during seizure-free intervals. Group E contains EEG signals recorded during seizure activity. In Table 1, we give a brief description of the five data sets.

In Figure 1, we present the energy distribution in the time-frequency plane of the data samples from the five group data. Five rows correspond to five different groups; in each row, there are five images which are computed by the same group EEG data. All the images are called energy distribution in the time-frequency plane. In each image, the vertical axis denotes the time, the horizontal axis represents the normalized frequency, and the color pixel denotes the energy value. The warmer color means the high energy value, and the colder color means the lower energy. The energy distribution in the time-frequency planes is computed by the short-time Fourier transform algorithm and the logarithmic operation then used to extract the logarithmic spectrogram value. All the images corresponding to the EEG data epochs are randomly chosen from the groups. Having obtained the power spectral density of EEG data, we then employ the logarithm to the spectrum power to get the logarithm feature. Finally, we process these logarithm power spectral density feature matrices as the same as the image set or video clip data.

We adopt the singular value decomposition (SVD) to get the basis of the matrix which consists of raw features of EEG spectrum set. More precisely, let  $\{\mathbf{X}_i\}_{i=1}^M$  be a spectrum set, where  $\mathbf{X}_i$  is a spectrum of EEG data with dimensionality  $m \times n$  and  $M$  is the number of spectrum sets. By vectorizing all the spectrum and stacking them along the column, we get a matrix  $\mathbf{Y} = [\text{vec}(\mathbf{X}_1), \text{vec}(\mathbf{X}_2), \dots, \text{vec}(\mathbf{X}_M)]$  which can be factorized as  $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  via SVD. We then choose the first  $p$  columns of  $\mathbf{U} \in \mathcal{G}(p, m \times n)$  as the Grassmannian point to represent the image set  $\{\mathbf{X}_i\}_{i=1}^M$ .

To validate the performance of our proposed FMGDA, five other Grassmann manifold algorithms are introduced

into our designed experiments to make comparisons with our algorithm on the clustering performance. The competitor algorithms used in the designed experiments are listed as follows:

- (i) KNN:  $k$  nearest neighbor classifier using projection metric for classifying Grassmann points without dimensionality reduction
- (ii) PML [28]: projection metric learning on the Grassmann manifold which can be seen as extension of Fisher LDA-like framework on the Grassmann manifold
- (iii) GGDA [27]: graph embedding discriminant analysis on Grassmannian manifolds
- (iv) GLPP [21]: locality preserving projections for Grassmann manifold
- (v) FMGDA: the proposed algorithm in this paper

We first use PML, GGDA, GLPP, and FMGDA to project the high-dimensional Grassmann points to a lower-dimensional manifold, then the KNN with  $k = 1$  is employed to classify Grassmann points.

**5.2. Experiment Results and Discussion.** From Figure 1, we can see the spectrum energy distribution of all the groups EEG data. Subject A data is different significantly with subject B data. We can conclude that eyes closed or open have obviously influence on the EEG data spectrum which can degrade the algorithm's performance on the epilepsy detection. Further, the intrinsic lower signal-to-noise ratio also results in the spectrum energy value to vary severely.

In the first designed experiment, we implement the five category classification task. Training data are randomly chosen from the total instances of each group data and the rest for test. We demonstrate the clustering accuracy results of the algorithms in the case of different numbers of training samples and data dimensionality in Table 2. The variable ratio denotes the percentage of all data we used for the model training, and the variable dim denotes the reduced dimensionality. From the experimental result, we can see that when the reduction dimensionality value is between 20 and 35, FMGDA has the best performance. As the dimensionality reduces and the performance decreases, the strange phenomenon also exists in other competitor algorithms. Classifying the EEG data into five different categories with only one



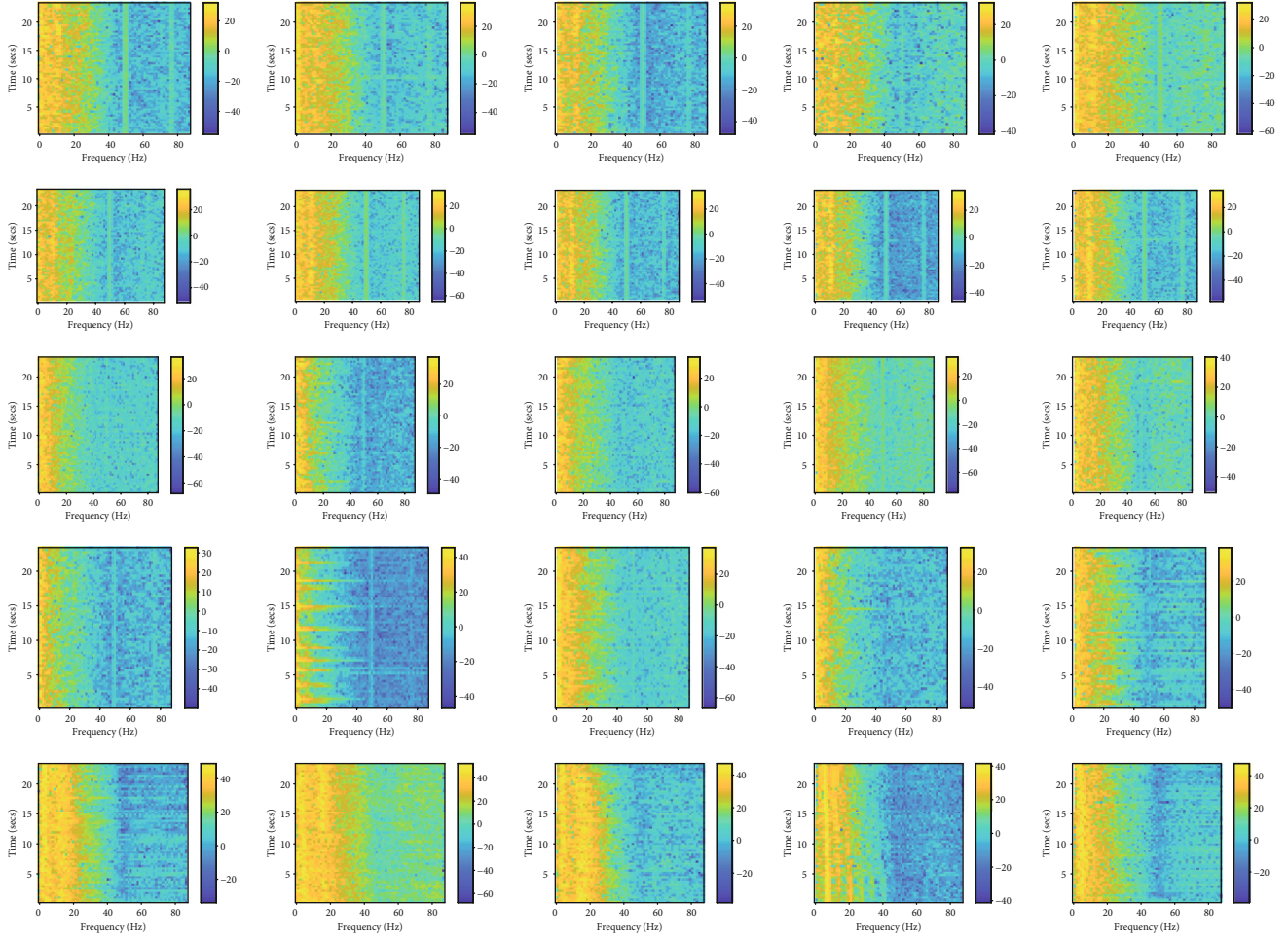


FIGURE 1: The energy distribution in the time-frequency plane feature of all EEG dataset.

channel data, our proposed FMGDA can achieve 70% accuracy when the setting value of the ratio and the reduction dimensions are 0.8 and 25, respectively. When the reduced dimensionality is between 20 and 35, we can get the considerable performance. The classification performance decreases as the dim increases; when the value of dim is bigger than 30, we guess that it is because too high dimensionality will introduce too much redundant information, thus degrading the algorithm's performance.

In the second designed experiment, we implement the binary classification experiment. We classify those healthy persons who are free in epilepsy with eyes open and closed. The experiment results are displayed in Table 3. From Table 3, we can get that when the dim value is equal to 15, we can get 70% classification accuracy which is significantly higher than the random probability. From this experiment, we can conclude that the state of the eyes has a very important influence on the classification of EEG signals.

In the third designed experiment, we compare the classification performance of our FMGDA with the competitor algorithms. We compare the five class classification and binary classification accuracy results, respectively. We firstly cluster the EEG data into five classes to try to recognize those

healthy persons whose eyes are open and those healthy persons whose eyes are closed, those persons who get the epilepsy in the hippocampal formation of the opposite hemisphere of the brain during seizure-free intervals, within the epileptogenic zone during seizure-free intervals and those persons who are during seizure. Secondly, we perform the binary clustering task to classify the EEG data into two classes corresponding to the healthy person with eyes open and the healthy person with eyes closed.

The comparison results are displayed in Figure 2. Classification accuracy shows that our FMGDA defeats the second place algorithm by a small margin. Even in the case of some parameter settings, the performance of our algorithm is weaker than that of the KNN or PML algorithms. The binary classification performance comparisons show that eyes open or closed have a significant effect on person's EEG signal, that is to say the state of the eyes has bad influence on the detection of epilepsy based on the EEG signal. The five category classification results demonstrate that none of the algorithms can get the classification accuracy greater than 75%. We guess that on the one hand, it is due to the complexity of the human brain and the low signal-to-noise ratio of the EEG signal itself, and on the other hand, it shows that our algorithm needs to be improved.



TABLE 2: The performance of FMGDA algorithms in the epilepsy detection of five categories.

Ratio	dim = 4	dim = 6	dim = 8	dim = 10	dim = 15	dim = 20	dim = 25	dim = 30	dim = 35	dim = 40	dim = 45	dim = 50	dim = 55	dim = 60
0.1	0.3800	0.3822	0.3689	0.4667	0.4489	0.4156	0.4289	0.3756	0.4022	0.3867	0.3733	0.3356	0.2622	0.2267
0.2	0.4050	0.3850	0.3725	0.4250	0.4525	0.4450	0.4575	0.4275	0.3950	0.4000	0.3750	0.3325	0.3175	0.2475
0.4	0.4233	0.3600	0.4367	0.4600	0.4733	0.4933	0.5400	0.4876	0.4500	0.4667	0.4200	0.3700	0.3100	0.2433
0.6	0.5150	0.3750	0.4000	0.4900	0.5000	0.4750	0.6200	0.5600	0.4650	0.4500	0.4100	0.3800	0.2600	0.2700
0.8	0.4600	0.4100	0.5000	0.5700	0.5300	0.5800	0.5900	0.5000	0.5200	0.4750	0.4300	0.3600	0.3300	0.2200

TABLE 3: The performance of FMGDA algorithms in the epilepsy detection of two categories. All the data were classified into two categories corresponding to the healthy person and epileptic person, respectively.

Ratio	dim = 4	dim = 6	dim = 8	dim = 10	dim = 15	dim = 20	dim = 25	dim = 30	dim = 35	dim = 40	dim = 45	dim = 50	dim = 55	dim = 60
0.1	0.6722	0.5500	0.6444	0.5833	0.5657	0.5444	0.5556	0.5778	0.5667	0.4833	0.5722	0.5167	0.4889	0.4833
0.2	0.7750	0.5625	0.6313	0.6250	0.64338	0.5938	0.6063	0.5438	0.5125	0.5250	0.5750	0.5188	0.5375	0.4688
0.4	0.5583	0.5833	0.6500	0.6583	0.5333	0.5833	0.5917	0.5833	0.5250	0.5750	0.6083	0.4333	0.5167	0.5667
0.6	0.6000	0.6125	0.6125	0.5500	0.6375	0.5375	0.6375	0.6625	0.5875	0.4125	0.4875	0.4750	0.5500	0.5250
0.8	0.6000	0.5000	0.6000	0.6750	0.7000	0.6250	0.6500	0.6250	0.6000	0.6250	0.3250	0.5250	0.5500	0.6500

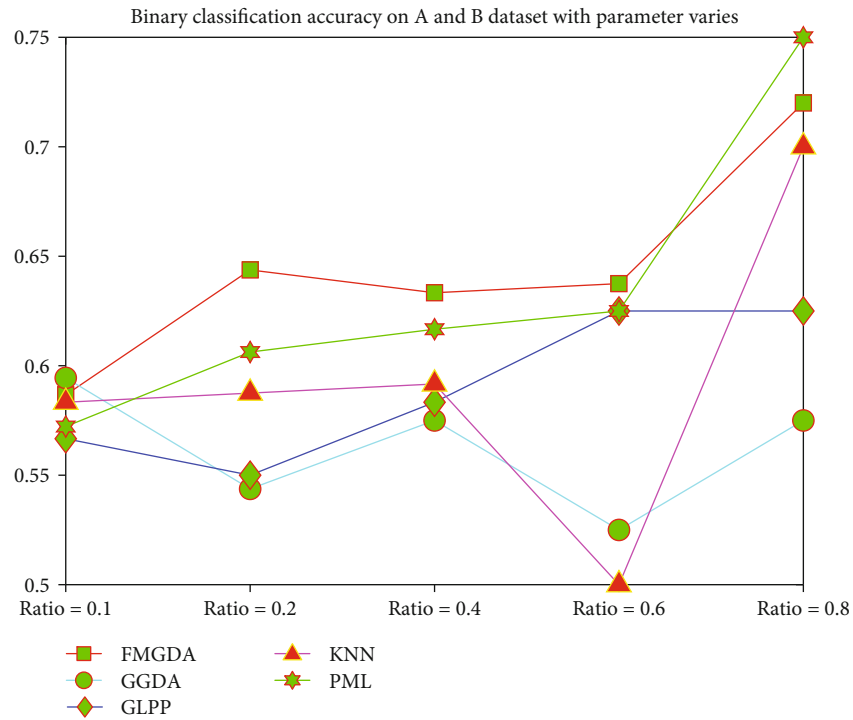
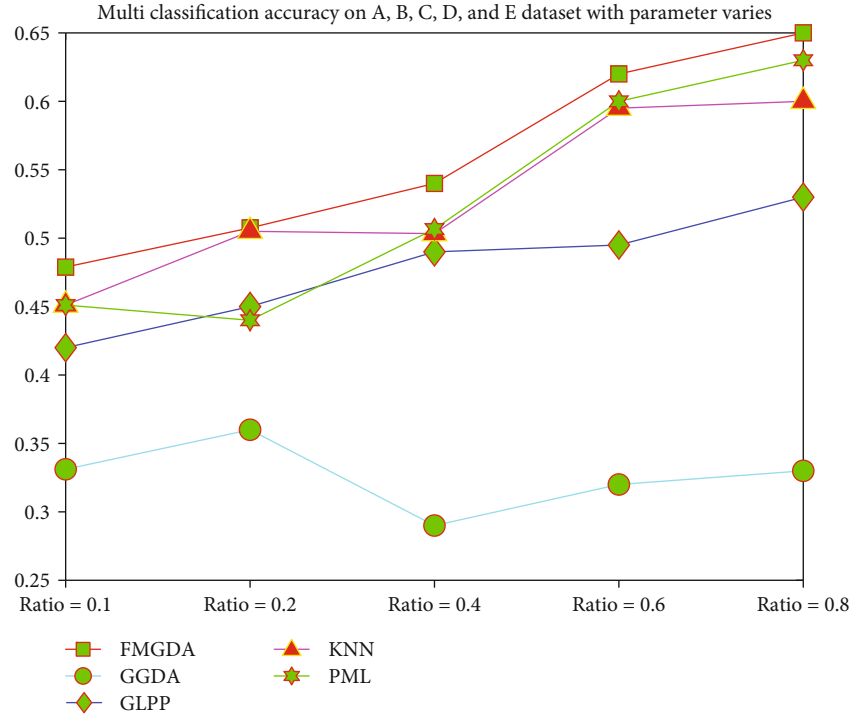
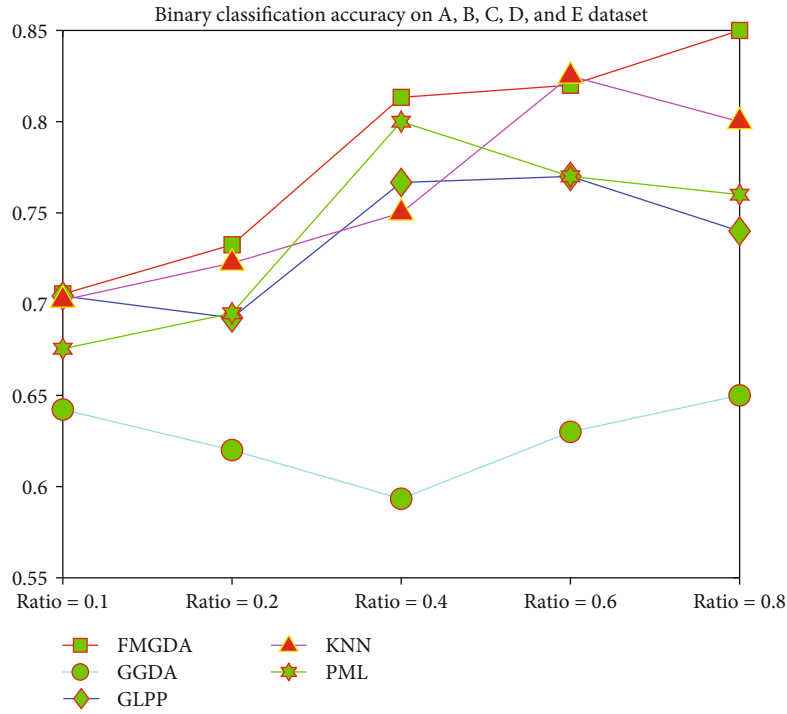


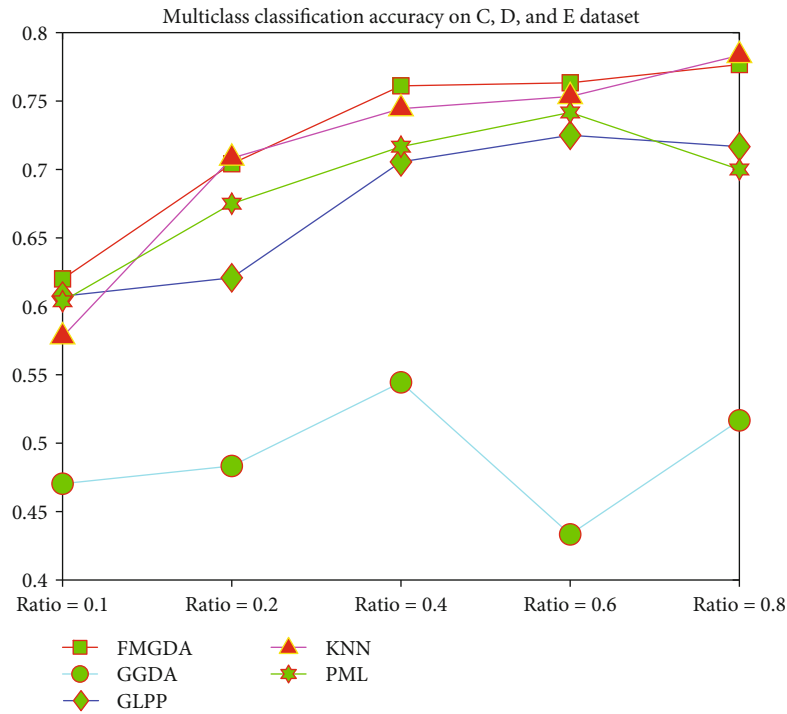
FIGURE 2: The classification accuracy on the EEG dataset with the parameter value varies. (a) Multiclass classification results over the all EEG dataset. (b) The binary classification results over group A and group B datasets.

We designed a fourth experiment to compare the performance of epilepsy detection based on Grassmann algorithms. We implement the binary classification task on all of the EEG data to classify the epilepsy persons from the healthy persons.

We compare the clustering performance of our FMGDA with the competitor algorithms. The comparison results are displayed in Figure 3. From the comparison results, we can get that our proposed FMGDA algorithm shows better



(a)



(b)

FIGURE 3: The classification accuracy on the EEG dataset with the parameter value varies. (a) Binary classification results over the all EEG dataset. (b) Multiclass classification results over group C, group D, and group E datasets.

performance on most of the experimental parameter setting case. The KNN and PML algorithms also get good clustering results and GGDA get the worst results.

In the fifth experiment, we test the algorithms' clustering performance of binary classification on the healthy and epi-

lepsy data which from the healthy persons with eyes open, from healthy persons with eyes closed, and from those persons who are during seizure, the classification accuracy is presented in Figure 4. From the experimental results, we can learn that KNN algorithm has the best performance over

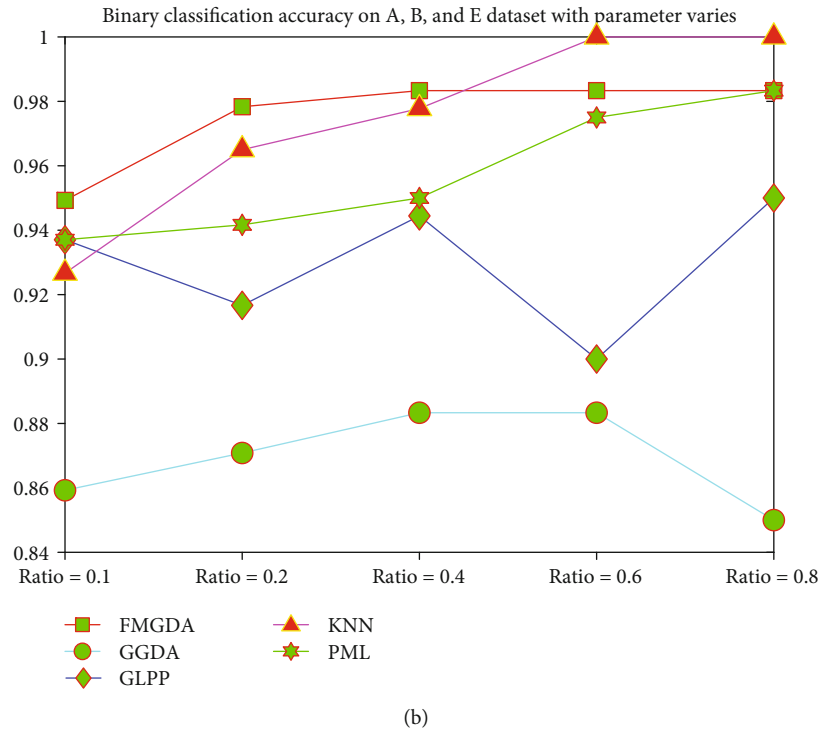
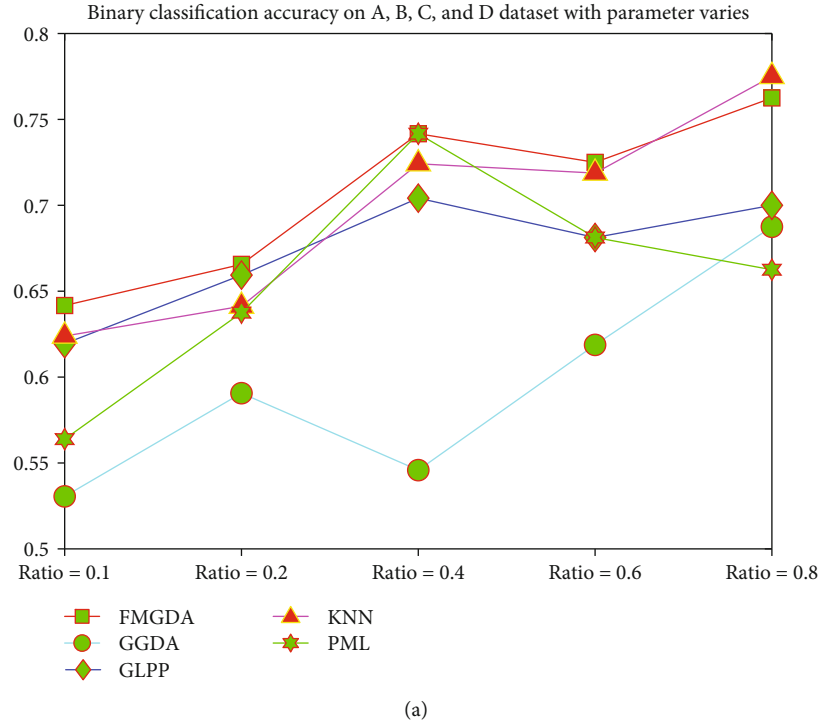


FIGURE 4: The classification accuracy on the EEG dataset with the parameter value varies. (a) Binary classification results over group A, group B, group C, and group D EEG dataset. (b) Binary classification results over group A, group B, and group E datasets.

the epilepsy detection task; our proposed FMGDA is inferior to KNN algorithm by a slight margin in some experiment cases but still better than the other three algorithms. The results also show that based on single-channel EEG data, patients, during seizures, can be detected with high accuracy and also those epileptic patients during seizure-free intervals.

## 6. Conclusion

In this paper, we proposed to use a Fréchet mean-based discriminant analysis method to implement the epilepsy detection based on the EEG data. Firstly, the short-time Fourier transform is used to extract the spectrogram feature of EEG



data to form a 2D power spectrum image. Secondly, we project the high-dimensional Grassmann spectrum image to a low-dimensional Grassmann manifolds. Thirdly, we adopt the Fréchet mean method to characterize the center of Grassmann points. Finally, the classical KNN method is adopted to detect the epilepsy based on the low-dimensional Grassmann points. Experimental results on benchmark datasets have demonstrated that our proposed method is useful in the application of epilepsy detection based on the EEG data and our proposed algorithm has certain advantages over the other competitor algorithms. However, there are still existing problems in our proposed algorithm. Firstly, the performance of our FMGDA on the epilepsy detection based on the EEG signal does not have great advantages, some state-of-art algorithms have better performance than our FMGDA. Secondly, the optimization algorithm involves the single value decomposition operations, which increase the complexity of our FMGDA; when the data size increases, the algorithm will be limited. We will focus on the enhance performance as well as reduce the optimization algorithm's complexity in our future work.

## Data Availability

Data were used to support this study are public dataset, which are available C at: <http://www.meb.unibonn.de/epileptologie/science/physik/eeegdata.html>.

## Conflicts of Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work; there is no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61772241 and 61702225, by the Natural Science Foundation of Jiangsu Province under Grant BK20160187, by 2016 Qinglan Project of Jiangsu Province, by 2016 Six Talent Peaks Project of Jiangsu Province, and by the Science and Technology Demonstration Project of Social Development of Wuxi under Grant WX18IVJN002.

## References

- [1] S. Dehuri, A. K. Jagadev, and S. B. Cho, "Epileptic Seizure Identification from Electroencephalography Signal Using DE- RBFNs Ensemble," *Procedia Computer Science*, vol. 23, pp. 84–95, 2013.
- [2] M. Li, W. Chen, and T. Zhang, "Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble," *Biomedical Signal Processing and Control*, vol. 31, pp. 357–365, 2017.
- [3] L. D. Iasemidis, Deng-Shan Shiao, W. Chaovalitwongse et al., "Adaptive epileptic seizure prediction system," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 5, pp. 616–627, 2003.
- [4] K. Fu, J. Qu, Y. Chai, and T. Zou, "Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals," *Biomedical Signal Processing and Control*, vol. 18, pp. 179–185, 2015.
- [5] R. Meier, H. Dittrich, A. Schulze-Bonhage, and A. Aertsen, "Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns," *Journal of Clinical Neurophysiology*, vol. 25, no. 3, pp. 119–131, 2008.
- [6] G. R. Minasyan, J. B. Chatten, M. J. Chatten, and R. N. Harner, "Patient-specific early seizure detection from scalp Electroencephalogram," *Journal of Clinical Neurophysiology*, vol. 27, no. 3, pp. 163–178, 2010.
- [7] A. Correa, E. Laciari, H. D. Patino, and M. E. Valentinuzzi, "Artifact removal from EEG signals using adaptive filters in cascade," *Journal of Physics: Conference Series*, vol. 90, article 012081, 2007.
- [8] İ. Güler and E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," *Journal of Neuroscience Methods*, vol. 148, no. 2, pp. 113–121, 2005.
- [9] A. Subasi, "EEG signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.
- [10] A. Aarabi, R. Fazel-Rezai, and Y. Aghakhani, "A fuzzy rule-based system for epileptic seizure detection in intracranial EEG," *Clinical Neurophysiology*, vol. 120, no. 9, pp. 1648–1657, 2009.
- [11] K. Polat and S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, 2007.
- [12] A. M. Chan, F. T. Sun, E. H. Boto, and B. M. Wingeier, "Automated seizure onset detection for accurate onset time determination in intracranial EEG," *Clinical Neurophysiology*, vol. 119, no. 12, pp. 2687–2696, 2008.
- [13] J. J. Niederhauser, R. Esteller, J. Echaz, G. Vachtsevanos, and B. Litt, "Detection of seizure precursors from depth-EEG using a sign periodogram transform," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 4, pp. 449–458, 2003.
- [14] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic seizure detection based on time-frequency analysis and artificial neural networks," *Computational Intelligence and Neuroscience*, vol. 2007, 13 pages, 2007.
- [15] B. Abibullaev, H. D. Seo, and M. S. Kim, "Epileptic spike detection using continuous wavelet transforms and artificial neural networks," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 8, no. 1, pp. 33–48, 2011.
- [16] Q. Yuan, W. Zhou, S. Li, and D. Cai, "Epileptic EEG classification based on extreme learning machine and nonlinear features," *Epilepsy Research*, vol. 96, no. 1–2, pp. 29–38, 2011.
- [17] Y. Jiang, Z. Deng, F. L. Chung et al., "Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 1, pp. 3–20, 2017.
- [18] Z. Zhang, Y. Zhou, Z. Chen, X. Tian, S. Du, and R. Huang, "Approximate entropy and support vector machines for

- electroencephalogram signal classification,” *Neural Regeneration Research*, vol. 8, no. 20, pp. 1844–1852, 2013.
- [19] P. A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, 2009.
  - [20] X. Pennec, P. Fillard, and N. Ayache, “A Riemannian framework for tensor computing,” *International Journal of Computer Vision*, vol. 66, no. 1, pp. 41–66, 2006.
  - [21] B. Y. Wang, Y. L. Hu, J. B. Gao et al., “Locality preserving projections for Grassmann manifold,” in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 2893–2900, Melbourne, Australia, 2017.
  - [22] P. A. Absil, R. Mahony, and R. Sepulchre, “Riemannian geometry of Grassmann manifolds with a view on algorithmic computation,” *Acta Applicandae Mathematicae*, vol. 80, no. 2, pp. 199–220, 2004.
  - [23] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.
  - [24] C. Yasuko, *Statistics on special manifolds*, Springer Science & Business Media, 174, 2012.
  - [25] J. Hamm and D. D. Lee, “Grassmann discriminant analysis: a unifying view on subspace-based learning,” *ICML '08: Proceedings of the 25th international conference on Machine learning*, 2008, pp. 376–383, Helsinki, Finland, July 2008.
  - [26] J. H. Hamm and D. D. Lee, *Extended Grassmann kernels for subspace-based learning*, Advances in neural information processing systems, 2009.
  - [27] M. T. Harandi, C. Sanderson, S. Shirazi, and B. C. Lovell, “Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching,” in *CVPR 2011*, pp. 2705–2712, Providence, RI, USA, June 2011.
  - [28] Z. W. Huang, R. P. Wang, S. G. Shan, and X. L. Chen, “Projection metric learning on Grassmann manifold with application to video based face recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 140–149, Hynes Convention Center in Boston, Massachusetts, USA, 2015.
  - [29] H. B. Yu, K. J. Xia, Y. Z. Jiang, and P. J. Qian, “Fremean-based’ Grassmann discriminant analysis,” *Multimedia Systems*, vol. 26, pp. 63–73, 2019.
  - [30] Y. C. Wong, “Differential geometry of Grassmann manifolds,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 57, no. 3, pp. 589–594, 1967.
  - [31] A. Srivastava and E. Klassen, “Bayesian and geometric subspace tracking,” *Advances in Applied Probability*, vol. 36, no. 1, article S0001867800012866, pp. 43–56, 2004.
  - [32] M. Harandi, C. Sanderson, C. H. Shen, and B. C. Lovell, “Dictionary learning and sparse coding on Grassmann manifolds: an extrinsic solution,” in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 3120–3127, Sydney, NSW, Australia, 2013.
  - [33] H. E. Cetingul and R. Vidal, “Intrinsic mean shift for clustering on Stiefel and Grassmann manifolds,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1896–1902, Miami, FL, USA, June 2009.
  - [34] T. Marrinan, B. J. Ross, B. Draper, M. Kirby, and C. Peterson, “Finding the subspace mean or median to fit your need,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1082–1089, Miami Beach, FL, USA, 2014.
  - [35] H. Karcher, “Riemannian center of mass and mollifier smoothing,” *Communications on Pure and Applied Mathematics*, vol. 30, no. 5, pp. 509–541, 1977.
  - [36] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi, “The geometric median on Riemannian manifolds with application to robust atlas estimation,” *NeuroImage*, vol. 45, no. 1, pp. S143–S152, 2009.
  - [37] T. T. Ngo, M. Bellalij, and Y. Saad, “The trace ratio optimization problem for dimensionality reduction,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2950–2971, 2010.
  - [38] F. P. Nie, S. M. Xiang, Y. Q. Jia, C. S. Zhang, and S. C. Yan, “Race ratio criterion for feature selection,” *AAAI*, vol. 2, pp. 671–676, 2008.
  - [39] H. Wang, S. C. Yan, D. Xu, X. O. Tang, and T. Huang, “Trace ratio vs. ratio trace for dimensionality reduction,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Minneapolis, MN, USA, June 2007.

## Research Article

# Evaluation of Cerebral Blood Flow Dynamics in Transient Ischemic Attacks Patients with Fast Cine Phase Contrast Magnetic Resonance Angiography

Yuzhao Wang, Duo Gao, and Huaijun Liu 

Department of Medical Imaging, The Second Hospital of Hebei Medical University, Shijiazhuang 050000, China

Correspondence should be addressed to Huaijun Liu; [huaijunliu2@126.com](mailto:huaijunliu2@126.com)

Received 24 February 2020; Accepted 25 March 2020; Published 13 April 2020

Guest Editor: Kaijian XIA

Copyright © 2020 Yuzhao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fast cine phase contrast magnetic resonance angiography (PC-MRA) has the potential to provide a quantitative measurement method for the diagnosis and treatment of cerebrovascular disease. To evaluate the changes of cerebral blood flow and the characteristics of artery lesion distribution in the patients of transient ischemic attacks (TIA). In all, 98 normal subjects and 106 TIA patients who underwent MRI examination within 72 h after the last symptom onset including the DWI sequence to exclude acute cerebral infarction were enrolled. The blood flow of the cranial total, the area of the internal carotid artery and vertebral artery, the average velocity, and the average blood flow were obtained and compared in normal subjects and TIA group. Analysis of Variance (ANOVA), *t*-test, and Kruskal-Wallis test were used for statistical assessments. The total cerebral blood flow of the TIA group and normal control group was no significant statistical difference ( $P > 0.05$ ). The total blood flow decreased with increasing age, and the TIA group was much lower than the control group. The blood flow of the right internal carotid artery in the TIA group had a significant difference compared with controls ( $P < 0.05$ ). However, the same situation did not happen in both of the left internal carotid artery and vertebral artery. Phase contrast magnetic resonance imaging has the potential to evaluate the change of cerebral blood flow in TIA patients. The decrease in the total blood flow and the symptom onset of TIA is consistent. Phase contrast magnetic resonance imaging could provide guidance to the diagnosis of TIA.

## 1. Introduction

Transient ischemic attacks (TIA) represent a frequent neurovascular condition associated with a risk of stroke recurrence of 5% at 48 h and 10% at 3 months in the absence of specific management [1]. Acute specialized management of TIA patients can significantly reduce the risk of stroke and other related vascular events such as myocardial infarction and vascular death [2–4].

In recent years, some studies have shown that TIA and ischemic stroke shared the same pathophysiological mechanism, but the prognosis may vary depending on the severity. Intracranial atherosclerosis (ICAS) is an important cause of ischemic stroke, especially in black, Hispanic, and Asian populations, where ICAS accounts for up to 50% of all ischemic strokes or TIAs [5–8]. Therefore, evaluation of cerebral blood flow dynamics in TIAs is essential.

The stationary proton will not produce phase displacement in the equilibrium gradient field, while the moving proton will produce phase displacement in the equilibrium gradient field. This is because the flow coding gradient has no effect on the stationary proton, and the stationary proton gets the same signal after two image acquisition. However, the phase displacement of the flowing protons is generated by the accumulation of two different positive and negative gradients between the two gradients, and this phase displacement is proportional to the movement of the flowing protons in the direction of the flow code. We subtract these two images, and we get an image of just the protons flowing. This phase change is the key of phase-sensitive flow imaging technology [9].

PC-MRA method, therefore, is to put MR technology for liquid phase displacement sensitivity combined with heart switch control technology, make the flow of the liquid phase

displacement time combined with rate, thus further get all about the flow of the liquid phase, amplitude, waveform, the relationship between flow rate, and time of quantitative data of MR imaging technology; this technology has realized the MR of liquid flow rate and flow velocity of quantitative analysis. PC-MRA 2D Q-flow sequence is exactly based on this basic principle, so as to achieve MR flow-oriented quantitative research.

Moran [10] firstly proposed the MR quantitative study of blood flow in 1982. He utilized the principle of bipolar gradient phase generated between stationary tissue and fluid to detect the blood flow. Whereafter, Pernicone et al. [11] used phase contrast MRA to determine the anatomy and flow direction of the blood vessels, but they could not quantitatively analyze the blood flow. Later, Enzmann et al. [12] measured the velocity and flow rate of different vessels with phase contrast cine method, and then compared with the results of Transcranial Doppler (TCD) examination, it was found that there were some differences between MR measurement and TCD detection. The results showed that the maximum peak was poorly correlated. Fast cine phase contrast magnetic resonance angiography (PC-MRA) is a magnetic resonance imaging technique that has the ability to characterize velocity-time curves in the human coronary arteries with data acquisition. Fast cine phase contrast magnetic resonance angiography exploits the fact that moving spins accumulate different transverse phase when moving in the direction of a magnetic field gradient. The phase shift is proportional to velocity, and the proportionality constant is easily controlled. As a result, these techniques are ideally suited to quantitative application [13].

Hence, in this study we aimed at investigating the changes of cerebral blood flow and the characteristics of artery lesion distribution in both healthy subjects and subjects with TIA using fast cine PC-MRA.

## 2. Materials and Methods

**2.1. Patients.** This study was approved by the local Ethics Committee. Between May 2015 and June 2016, normal subjects and TIA patients were prospectively enrolled after written informed consent was obtained.

The inclusion criteria for healthy subjects included (1) no history of cerebrovascular disease; (2) no brain parenchyma lesions were detected by conventional MRI scan; (3) no vascular stenosis, occlusion, aneurysm, arterial dissection, or vascular malformation were detected by MRA; (4) no neurological symptom and signs; and (5) good image quality and complete data. The following conditions were excluded: severe hepatic and renal dysfunction; brain tumor, brain hemorrhage, and brain trauma; concurrent infection; mental disorders; and hematological system diseases.

For subjects with TIA, the inclusion criteria included (1) the principles for the diagnosis of TIA refer to the updated Chinese expert consensus on a transient ischemic attack; (2) no signs of acute ischemic stroke were detected by conventional MRI scan; (3) admission in 72 h from onset; and (4) no history of stroke.

The exclusion criteria included (1) no transient neurological deficits caused by hypoglycemia, epilepsy, and migraine; (2) transient global amnesia; (3) a history of stroke in 3 months; (4) brain tumor, brain hemorrhage, and brain trauma; (5) mental disorders; and (6) no cooperative or unable to undergo MR examination.

Overall, 89 healthy subjects (seven females and six males, age range 43–71 years) served as control participants and 106 subjects with TIA (seven females and six males, age range 40–75 years) were included in analysis.

**2.2. MRI Acquisition Protocol.** All subjects underwent brain MRI with a 3.0T MR scanner (3.0T Philips MR Systems Achieva, Best, Netherlands) equipped with a standard head coil. The standardized MR protocol included T1-weighted imaging, T2-weighted imaging, DWI, TOF-MRA, and Q-Flow PC-MRA. T1WI sequence parameters included slice thickness, 6.5 mm; slice gap, 1.3 mm; inversion time (TI), 900 ms; and TR/TE, 2000/20 ms. The parameters of T2WI and T2 FLAIR sequence: slice thickness and gap were consistent with the T1WI sequence; T2WI sequence parameters included TR = 5000 ms, TE = 120 ms; T2-FLAIR sequence parameters included TI, 2300 ms; TR/TE, 10000/120 ms. DWI was collected using single-shot echoplanar imaging with  $b$  values of 0 and 1000 s/mm<sup>2</sup>. repetition time (TR)/echo time (TE) = 2400/104 msec; field of view (FOV), 22 cm; matrix, 128 × 128; section thickness, 5 mm; and intersection gap 0 mm; number of slices, 17–18. 2D Q-Flow sequence was used to quantitatively detect the intracranial cerebral blood flow. The image of vascular localization was obtained based on the 3D TOF-MRA scanning, and the 2D Q-Flow scanning was performed perpendicular to the target vessel. 2D Q-Flow sequence parameters included 2D Q-Flow sequence scanning parameters: number of excitation (NEX) = 2, slice thickness = 4.0 mm, spacing = 0 mm, matrix = 512 × 256, field of view = 240 mm × 240 mm, reverse angle = 20°, bandwidth = 32. A phase change curve was obtained using peripheral gating, phaseless folding, respiratory compensation, and flow compensation techniques. The total of Phase = 16. The flow rate-coded Venc was set to 90 cm/s. Imaging time was approximately 15 to 20 minutes.

### 2.3. Image Analysis

**2.3.1. Image Reconstruction.** 3D TOF-MRA raw images were exported to the PHILIPS Extended MR Workspace 2.6.3.2 workstation. The raw images were reconstructed in three dimensions using the maximum intensity projection (MIP) method. The vascular images of the anterior and posterior cerebral circulation were obtained by postprocessing. Vessel morphology was recorded and the lumen diameter of the vessel layer corresponding to the 2D Q-Flow orientation was measured.

**2.3.2. Calculation of the Degree of Stenosis.** The inner diameter of the internal carotid artery (ICA), anterior cerebral artery, middle cerebral artery, vertebral artery, basilar artery, and posterior cerebral artery was computed, respectively. The lumen at the narrowest place and the normal lumen at the



distal of the stenosis were measured. According to the reduction of lumen diameter and the degree of signal loss, and the North American NASCET guidelines, the degree of stenosis is divided into normal, mild stenosis ( $0\% < \text{the extent of stenosis} < 50\%$ ), moderate stenosis ( $50\% \leq \text{the extent of stenosis} < 75\%$ ), severe stenosis ( $75\% \leq \text{the extent of stenosis} < 100\%$ ), and occlusion (completely signal loss).

Moderate stenosis of vessels or above associated with clinical symptoms was identified as a well-defined intracranial artery lesion. The tortuosity of the vessel was also classified as a mild stenosis group. If there are multiple degrees of stenosis in the same artery, the most severe stenosis is selected to calculate the degree of stenosis.

**2.3.3. Calculation of the Flow in the Lumen.** Phase amplitude images and phase contrast images were obtained by 2D Q-Flow. The two sets of images were 30 images of different phases with the same cardiac cycle, which represents the function of flow velocity and time during the cardiac cycle. The images were exported to the PHILIPS Extended MR Workspace 2.6.3.2 workstation for postprocessing. Two experienced radiologists delineated the cross-sectional area of the region of interest (ROI) on the images. The diameter of the corresponding layer of the 3D TOF-MRA images is taken as a reference so that the cross-sectional area of ROI is the closest to the actual lumen. Then, the time-flow curve in a cardiac cycle would be obtained. The flow rate in the lumen per unit time can be obtained. The calculation formula is based on the flow rate ( $\text{ml/min}$ ) = the flow rate ( $\text{cm/s}$ )  $\times$  cross-sectional area ( $\text{cm}^2$ )  $\times 60$  (s).

Firstly, a fluid model is designed, and it is found that the measured fluid signal strength has a good linear correlation with the velocity change, which indicates that the model can be used for the quantitative analysis of fluid.

The model lumen flow rate was set as 1.0 ml/s, 2.0 ml/s, 3.0 ml/s, 4.0 ml/s, and 5.0 ml/s, and the flow rate was measured according to the basic law of liquid flow  $Q = v r^2$ , given that the inner diameter and flow rate of the lumen and the velocity of liquid flow in the lumen  $v = Q/r^2$  can be calculated. The measured velocity was positively correlated with the actual velocity (correlation coefficient = 0.98,  $P < 0.0001$ ). Therefore, a cross-sectional area is a parameter to measure the flow rate.

**2.3.4. Statistical Analysis.** SPSS Statistics (v. 22, IMB, Armonk, NY) was used for statistical analyses. Data distributions were tested with the Shapiro-Wilk W test. All data were expressed as mean  $\pm$  standard deviation, except where noted. Mean flow velocity, peak flow rate, and flow rate of the bilateral internal carotid artery, A1 segment in anterior cerebral artery, M1 segment in middle cerebral artery, vertebral artery, basilar artery, and P1 segment in the posterior cerebral artery were analyzed with Student's  $t$  test. The flow rate and flow of intracranial artery between different genders were compared by the two sample  $t$  test. If the data for the two groups were nonnormally distributed or the variances were not equal, the nonparametric Kruskal-Wallis test was used. The statistical significance level was set at  $P < 0.05$  for all analyses.

TABLE 1: Baseline characteristics of the TIA group and control group.

	TIA group ( $n = 106$ )	Control group ( $n = 50$ )	$T/\chi^2$	$P$
Age	$58.05 \pm 19.10$	$60.14 \pm 12.20$	25.61	0.44
Male	56 (0.53)	26 (0.52)	0.095	0.76
Hypertension	20 (0.19)	4 (0.08)	0.352	0.553
Diabetes mellitus	20 (0.19)	3 (0.06)	13.344	0.03
Smoke	54 (0.51)	14 (0.50)	1.835	0.176
Atrial fibrillation	19 (0.18)	3 (0.06)	20.448	0.02

TABLE 2: Narrowness of varying degrees between the TIA group and control group.

	TIA group (total 106)	Control group (total 50)	$P$
Nonstenosis	37	21	0.23
Mild stenosis	40	22	0.22
Moderate stenosis	20	7	0.15
Severe stenosis	9	0	0.03

### 3. Results and Discussion

106 patients (56 male, 50 female) with TIA, ranging in age from 21 to 85 years ( $58.05 \pm 19.10$ ), were evaluated in MRI routine examination including MRI, DWI, and MRA within 72 hours after the last symptom. The control group included 50 qualified volunteers, including 26 males and 24 females in age from 28 to 71. Their demographic characteristics including mean  $\pm$  standard deviation, age, body mass index (BMI), hypertension, diabetes mellitus, smoke, and Atrial fibrillation are listed in Table 1.

**3.1. The Degree of Stenosis and Distribution Characteristics in Patients with TIA.** The narrowness of varying degrees between the TIA group and the control group is shown in Table 2. There is no statistically significant difference between the incidence of mild stenosis in the TIA group and the control group. However, there was a statistically significant difference between the incidence of severe stenosis and the normal control group in the TIA group ( $P < 0.05$ ).

In the 106 TIA subjects, 69 had different degrees of vascular stenosis, including 19 with anterior cerebral artery stenosis, 21 with cerebral artery stenosis, 28 with internal carotid artery stenosis, and 17 with posterior cerebral artery stenosis; 23 patients with vertebral artery stenosis; and 12 patients with basilar artery stenosis. There are two sample demonstrated moderate middle cerebral artery stenosis and BA mild inhomogeneous stenosis using PC-MRA and TOF-MRA (Figures 1–3). There were only 26 cases involving one vessel stenosis, 29 cases with two vessels, and 14 patients with three or more vessel stenosis. There were no significant differences



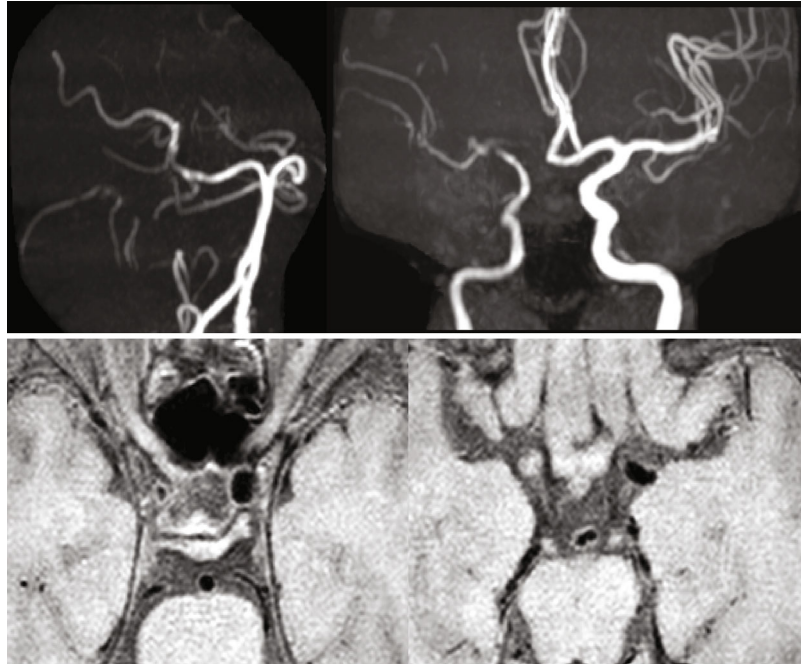


FIGURE 1: PC-MRA and TOF-MRA for different cerebrovascular stenosis.

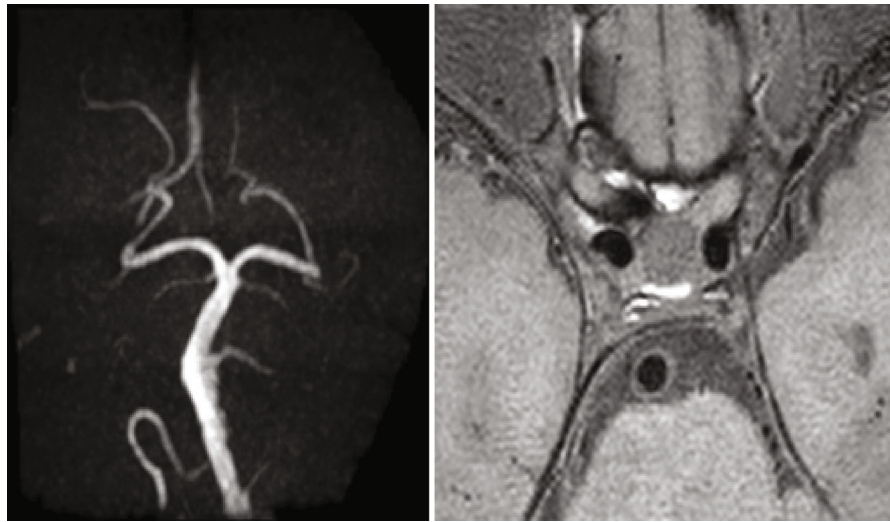


FIGURE 2: PC-MRA and TOF-MRA for BA mild inhomogeneous stenosis.

in different vascular stenosis groups ( $P > 0.05$ ). There was also no significant difference between the cumulative one and two-vessel groups ( $P > 0.05$ ), and there was a statistical difference between the one or two-vessel groups with the cumulative three or more vessel groups ( $P < 0.05$ ). There was no significant difference in the degree of vascular stenosis between different gender groups ( $P > 0.05$ ).

**3.2. The Characteristics of Cerebral Flow in Patients with TIA.** The main blood supply of the head is bilateral ICA and vertebral artery and basilar artery (V-BA). In this study, we consider the sum of bilateral ICA and V-BA blood flow as

the whole cerebral flow to compare the TIA group and control group in total cerebral blood flow (Table 3). The results showed that there was no significant difference in the whole cerebral blood flow between the TIA group and the normal control group ( $P > 0.05$ ).

The total blood flow decreased with age. The total blood flow in the TIA group was slightly lower than that in the control group, but there was no statistically significant difference ( $P > 0.05$ ). The right internal carotid artery flow in the TIA group was significantly different from the control group ( $P < 0.05$ ). The right internal carotid artery flow in the TIA group decreased compared with the control group. There

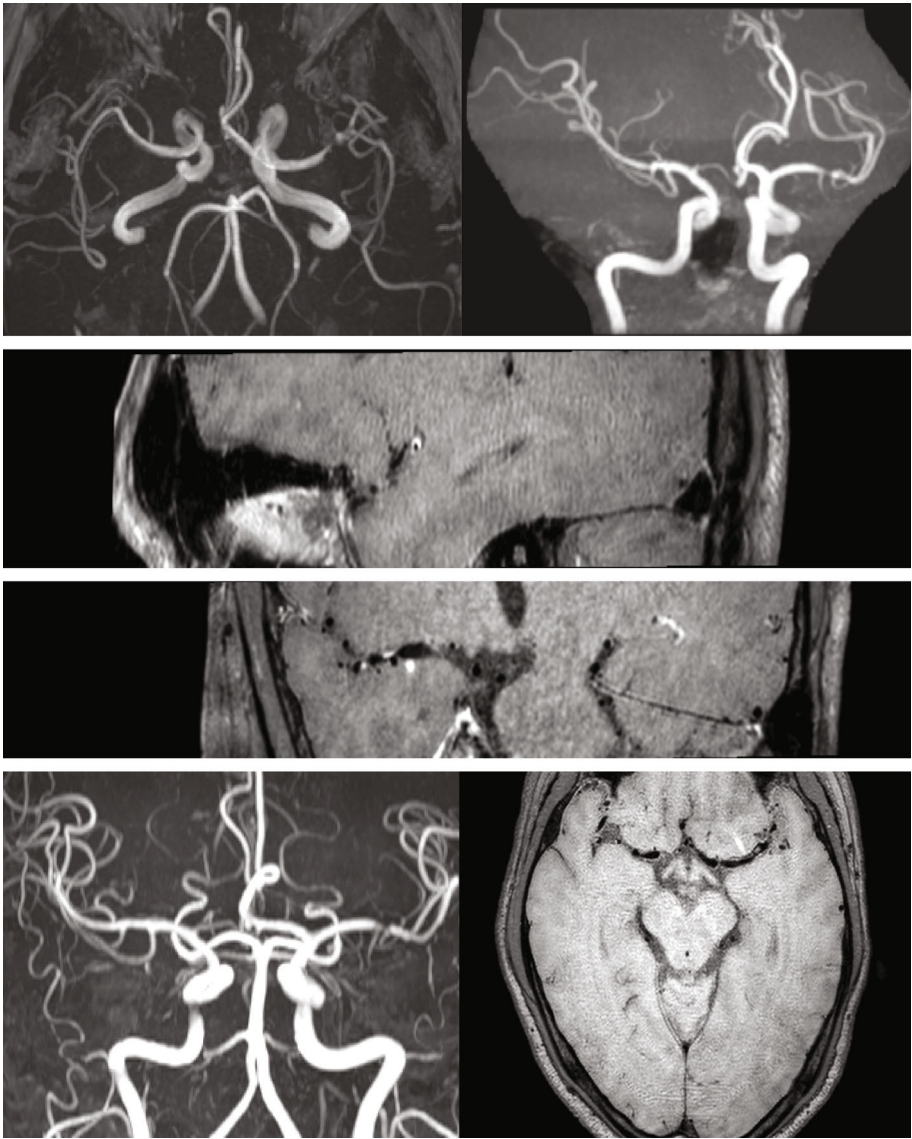


FIGURE 3: PC-MRA and TOF-MRA for MCA stenosis.

TABLE 3: The blood flow of the internal carotid artery (ICA) and vertebral artery (VA) between the TIA group and control group.

	Flow(ml/min)		P
	TIA group	Control group	
ICA(L)	238 ± 63	242 ± 71	0.38
ICA(R)	261 ± 71	271 ± 79	0.45
VA(L)	67 ± 39	69 ± 45	0.54
VA(R)	59 ± 37	63 ± 39	0.26

was no significant difference in the left internal carotid artery and bilateral vertebral artery flow between the TIA group and the control group (Table 4).

3.3. The Relationship between the Occurrence of TIA and Atherosclerotic Lesions. Among the 106 patients with TIA,

TABLE 4: The blood flow of the internal carotid artery (ICA) and vertebral artery (VA) among different TIA group and control group.

	Flow(ml/min)			P
	TIA group (ICA)	TIA group (V-BA)	Control group	
ICA(L)	227 ± 67	221 ± 74	242 ± 71	0.09
ICA(R)	221 ± 76	272 ± 81	271 ± 79	0.038
VA(L)	76 ± 36	59 ± 47	69 ± 45	0.07
VA(R)	74 ± 37	61 ± 44	63 ± 39	0.12

62 patients had different degrees of atherosclerotic lesions, including 26 cases involving internal carotid artery, 22 cases involving only intracranial arteries, and 14 cases involving both internal carotid artery and intracranial artery. There were 21 cases of atherosclerotic lesions in 50 cases of the

control group. 9 cases were involved in the internal carotid artery, only 8 cases involving the intracranial artery, and 4 cases involving both the internal carotid artery and intracranial artery. There was no statistically significant difference in the distribution of atherosclerotic lesions in the cranial neck between the TIA patients and the control group ( $P > 0.05$ ).

**3.4. The Relationship between the Occurrence of TIA and Vascular Morphology.** The proportion of vascular dysplasia or vascular variability in the TIA group (38.7%) is slightly higher than the control group (33.9%), but there was no significant statistical difference ( $P > 0.05$ ).

## 4. Conclusions

Most TIA patients have varying degrees of stenosis or occlusion of the supply artery. The risk of ischemic stroke in these TIA patients increased with some risk factors such as the degree, extent, or duration of vascular stenosis. Carotid, vertebral and intracranial atherosclerosis, wall thickening, unstable plaque formation, and others can lead to the stenosis of the vascular lumen. In this study, 58% of TIA patients were accompanied by atherosclerotic lesions of cranial and cervical supply arteries, mostly involving the bifurcation of the carotid artery and the internal carotid artery. In the TIA group, the blood supplement of ICA and V-BA that accompanied atherosclerotic lesions was significantly increased compared with the normal control group. Therefore, poor vascular development and atherosclerosis are likely to be one of the risk factors for TIA.

In this study, the 2D Q-flow PC-MRA method was used to compare the cerebral blood flow of TIA patients and the normal control group. We analyzed that the symptoms of some TIA patients may be caused by microembolic embolization of the small distal supply artery, but there were no significant changes in total cerebral blood flow. Some studies have found that there is no significant difference between whole brain blood flow in patients with atherosclerosis and normal healthy group, which is consistent with the results of this study. Since the MR examination of all TIA patients in this study was conducted after the onset of symptoms, it could not fully show the vascular conditions at the onset of TIA. Therefore, it has certain limitations.

In conclusion, atherosclerotic disease and changes of vascular dysplasia or morphologic are likely to be the risk factors for TIA. Magnetic resonance phase contrast angiography provides more evidence for the diagnosis of transient ischemic attack, and it has a broad clinical application prospect.

## Data Availability

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

## Conflicts of Interest

The authors declare no competing interests.

## Acknowledgments

We would like to thank everyone in the medical imaging departments in the Second Hospital of Hebei Medical University for their support.

## References

- [1] C. M. Wu, K. McLaughlin, D. L. Lorenzetti, M. D. Hill, B. J. Manns, and W. A. Ghali, "Early risk of stroke after transient ischemic attack: a systematic review and meta-analysis," *Archives of Internal Medicine*, vol. 167, no. 22, pp. 2417–2422, 2007.
- [2] P. M. Rothwell, M. F. Giles, A. Chandratheva et al., "Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): a prospective population-based sequential comparison," *The Lancet*, vol. 370, no. 9596, pp. 1432–1442, 2007.
- [3] P. C. Lavallée, E. Meseguer, H. Abboud et al., "A transient ischaemic attack clinic with round-the-clock access (SOS-TIA): feasibility and effects," *The Lancet Neurology*, vol. 6, no. 11, pp. 953–960, 2007.
- [4] C. W. Cereda and J. M. Olivot, "Emergency department (ED) triage for transient ischemic attack (TIA)," *Current Atherosclerosis Reports*, vol. 20, no. 11, pp. 55–65, 2018.
- [5] J. F. Arenillas, "Intracranial atherosclerosis: current concepts," *Stroke*, vol. 42, Supplement 1, pp. S20–S23, 2011.
- [6] C. A. Holmstedt, T. N. Turan, and M. I. Chimowitz, "Atherosclerotic intracranial arterial stenosis: risk factors, diagnosis, and treatment," *The Lancet Neurology*, vol. 12, no. 11, pp. 1106–1114, 2013.
- [7] A. I. Qureshi and L. R. Caplan, "Intracranial atherosclerosis," *The Lancet*, vol. 383, no. 9921, pp. 984–998, 2014.
- [8] P. Gorelick, K. S. Wong, and L. Liu, "Epidemiology," *Frontiers of Neurology and Neuroscience*, vol. 40, no. 1, pp. 34–46, 2016.
- [9] M. G. Knuttinen, J. Karow, W. Mar, M. Golden, and K. L. Xie, "Blood pool contrast-enhanced magnetic resonance angiography with correlation to digital subtraction angiography: a pictorial review," *Journal of Clinical Imaging Science*, vol. 4, p. 63, 2014.
- [10] P. R. Moran, "A flow velocity zeugmatographic interlace for NMR imaging in humans," *Magnetic Resonance Imaging*, vol. 1, no. 4, pp. 197–203, 1982.
- [11] J. R. Pernicone, J. E. Siebert, T. A. Laird, T. L. Rosenbaum, and E. J. Potchen, "Determination of blood flow direction using velocity-phase image display with 3-D phase-contrast MR angiography," *AJNR American Journal of Neuroradiology*, vol. 5, no. 13, pp. 1435–1438, 1992.
- [12] D. R. Enzmann, M. R. Ross, M. P. Marks, and N. J. Pelc, "Blood flow in major cerebral arteries measured by phase-contrast cine MR," *AJNR American Journal of Neuroradiology*, vol. 1, no. 15, pp. 123–129, 1994.
- [13] M. P. Marks, N. J. Pelc, M. R. Ross, and D. R. Enzmann, "Determination of cerebral blood flow with a phase-contrast cine MR imaging technique: evaluation of normal subjects and patients with arteriovenous malformations," *Radiology*, vol. 182, no. 2, pp. 467–476, 1992.

## Research Article

# A Modified Skip-Gram Algorithm for Extracting Drug-Drug Interactions from AERS Reports

Li Wang,<sup>1,2</sup> Wenjie Pan,<sup>1</sup> QingHua Wang,<sup>1</sup> Heming Bai,<sup>2</sup> Wei Liu,<sup>2</sup> Lei Jiang,<sup>3</sup>  
and Yuanpeng Zhang<sup>1,2</sup> 

<sup>1</sup>Department of Medical Informatics, Medical School, Nantong University, Nantong 226001, China

<sup>2</sup>Research Center for Intelligence Information Technology, Nantong University, Nantong 226001, China

<sup>3</sup>Department of Rheumatology and Immunology, Changzheng Hospital, The Second Military Medical University, Shanghai 200433, China

Correspondence should be addressed to Yuanpeng Zhang; [maxbirdzhang@ntu.edu.cn](mailto:maxbirdzhang@ntu.edu.cn)

Received 19 December 2019; Revised 22 February 2020; Accepted 27 February 2020; Published 12 April 2020

Guest Editor: Kaijian XIA

Copyright © 2020 Li Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Drug-drug interactions (DDIs) are one of the indispensable factors leading to adverse event reactions. Considering the unique structure of AERS (Food and Drug Administration Adverse Event Reporting System (FDA AERS)) reports, we changed the scope of the window value in the original skip-gram algorithm, then propose a language concept representation model and extract features of drug name and reaction information from large-scale AERS reports. The validation of our scheme was tested and verified by comparing with vectors originated from the cooccurrence matrix in tenfold cross-validation. In the verification of description enrichment of the DrugBank DDI database, accuracy was calculated for measurement. The average area under the receiver operating characteristic curve of logistic regression classifiers based on the proposed language model is 6% higher than that of the cooccurrence matrix. At the same time, the average accuracy in five severe adverse event classes is 88%. These results indicate that our language model can be useful for extracting drug and reaction features from large-scale AERS reports.

## 1. Introduction

Drug-drug interactions (DDIs) accounted for over 30% of all adverse drug events [1]. More serious fact is that large quantity of DDIs manifested after a long period of exposure. As a result, AERS reports have been served as the cornerstone for detecting unanticipated interactions. The development of computational prediction and assessment of DDIs become attractive to the US FDA and pharmaceutical companies [2]. Harpaz et al. developed a taxonomy that characterized the associations and predicted several potential multi-item drug adverse effects [3]. They revealed that duplicate reports caused spurious associations. Tatonetti et al. constructed a drug-reaction frequency matrix and used Fisher's exact test for feature extraction from frequency matrices for DDI prediction [4]. Logistic regression was used for classification. Predicted DDIs were significantly enriched for known effects. Cheng and Zhao integrated drug phenotypic, therapeutic,

chemical, and genomic properties to predict DDIs [5]. These four types of drug-drug similarities were calculated as features of each drug-drug pair for prediction. Five machine learning algorithms were implemented, and they found that integration of multidata sources can improve the performance of DDI prediction. Cami et al. proposed a Predictive Pharmacointeraction Network [6]. They exploited the network structure of all known DDIs, combined with various taxonomic and intrinsic properties of drugs to predict unknown DDIs. While these methods performed well, their limitations are obvious either. From the above, similarity-based methods rely on various profiles including drug molecular structure profiles, drug-drug interaction profiles, and pharmacophoric profiles [7, 8]. First of all, barely any of the previous work took a systematic data preprocessing method before taking advantage of AERS reports; a standard language description framework should be used to organize all the reports. Second, in the face of the large amount of free



text reports, to solve the problem that the integration of profiles cost large amount of manually check and selection biases, a language concept representation model is urgently needed. Third, the existing approaches do not seem to extract features from AERS reports efficiently and to test the quality of the new-mined DDI concept by our MSG model, which also are needed to be distinguished in the DrugBank database.

The skip-gram algorithm was one of the language models set in the open-source word2vec [9, 10]. This algorithm was used to render distributional representation of words from large-scale unmarked text. The skip-gram algorithm has been widely acknowledged and successfully applied to many natural language processing tasks, such as text clustering, entity completion in incomplete knowledge bases or ontologies, and text retrieval [11–13]. At the same time, there are few researchers focusing on applying this language model to mining the pharmacovigilance information from large-scale reports in free text format.

The main contribution of this work can be listed as follows:

- (1) We proposed a new language concept representation model by changing the scope of the window value in the original skip-gram algorithm
- (2) Compared to the previous traditional language model, the new model can extract features of drug name and reaction information from large-scale AERS reports more efficiently
- (3) The new drug-drug interaction datasets can be mined through the use of the proposed MSG language concept representation model

## 2. Method

As shown in Figure 1, in the whole research scheme, we proposed a modified skip-gram (MSG) algorithm for drug name and reaction description feature extraction from FDA AERS reports, and the description of DDIs in DrugBank was enriched.

In Step 1, original free text AERS reports are transformed into structured tables (Figure 1(a)). Our study refers to the framework of OHDSI (Observational Health Data Sciences and Informatics) and puts Banda et al.’s research into practice, completing the cleaning and standardization of AERS reports [14, 15]; all the structured tables are stored in a PostgreSQL database. In addition, we extracted DrugBank DDI and toxicity data into text files as shown in Figure 1(b).

In Step 2, the MSG algorithm was applied to calculate the embeddings of drug name and reaction descriptions from AERS reports and DrugBank DDI. The vectors with a dimension of 100 were represented as features of drug name and reaction description. All the names of drugs and reactions are converted from string to the numbers as shown in Figure 1(c). For example, one drug name is represented as a drug concept ID 1327356 with 100 dimension separated numbers.

In Step 3, a logistic regression classifier was used to validate the above embedding values. To compare the quality of embeddings generated from MSG, CM-TF-IDF, another traditional language representation model, was also tested. We chose to assess the performance of the classifier in comparing the area under the curve (AUROC) of AERS reports and DrugBank DDI with a cross-validation approach.

Finally, the descriptions of DDIs in five severe adverse event classes were enriched into the current adverse event results.

**2.1. Data Collection and Preprocess.** We collected AERS reports from the FDA’s website between 2004 and 2014. In addition, SIDER was used as the gold standard for positive reference samples [16]. DDI data and drug toxicity data were extracted from the DrugBank database [17].

Although it is a free and publicly available resource, the FDA AERS data still presents multiple hurdles in consolidating all relevant data. To avoid producing unreliable and irreproducible results, widely accepted data preprocessing methods were referred to and put into practice. Thanks to the efforts of large communities such as the Observational Health Data Sciences and Informatics (OHDSI) [14], we can focus more on model building than on lots of time-wasted efforts such as cleaning and standardizing the AERS reports. For details of preprocessing FDA AERS reports, we referred to Banda et al.’s research [15]. First, AERS reports in Extensible Markup Language (XML) format were extracted into seven individual tables; these tables were loaded onto PostgreSQL. Second, a demo table was created for missing value imputation and case deduplication. Missing value imputation was performed on four demographic fields (age, sex, country, and event date). As a case may exist in the legacy AERS dataset or in the new FAERS dataset, different unique row keys were managed in a case deduplication step. Finally, regular expression was taken as the main method for mapping drug and reaction concept into the OHDSI standard vocabulary concept identifier (consisting of RxNorm CUIs and MedDRA standard codes).

After the preprocessing of AERS reports, in total, 4,493,179 reports are achieved, within which 713,441 reports listed exactly two drugs. To ensure reasonable estimates and statistical significance, at least ten AERS reports are required to support one given drug concept [18]. We selected ten as the threshold to filter out drug concepts in AERS reports. As a result, 675 drug concepts are filtered out and existed in DrugBank DDI dataset drug concepts either. These 675 drugs are included in Reference Drug Lists (RDL).

**2.2. Modified Skip-Gram (MSG) Model.** The original skip-gram algorithm was modified for drug name and reaction description feature extraction from FDA AERS reports and DrugBank DDIs. Based on distributional hypothesis theory [9], a word can be characterized into an embedding value by contexts, which are the surrounding words around its position in the sentence. These embeddings encode the semantic meanings of the target word into a low-dimensional vector. In this research, all the drugs with reaction words are encoded into a low-dimensional vector. Our



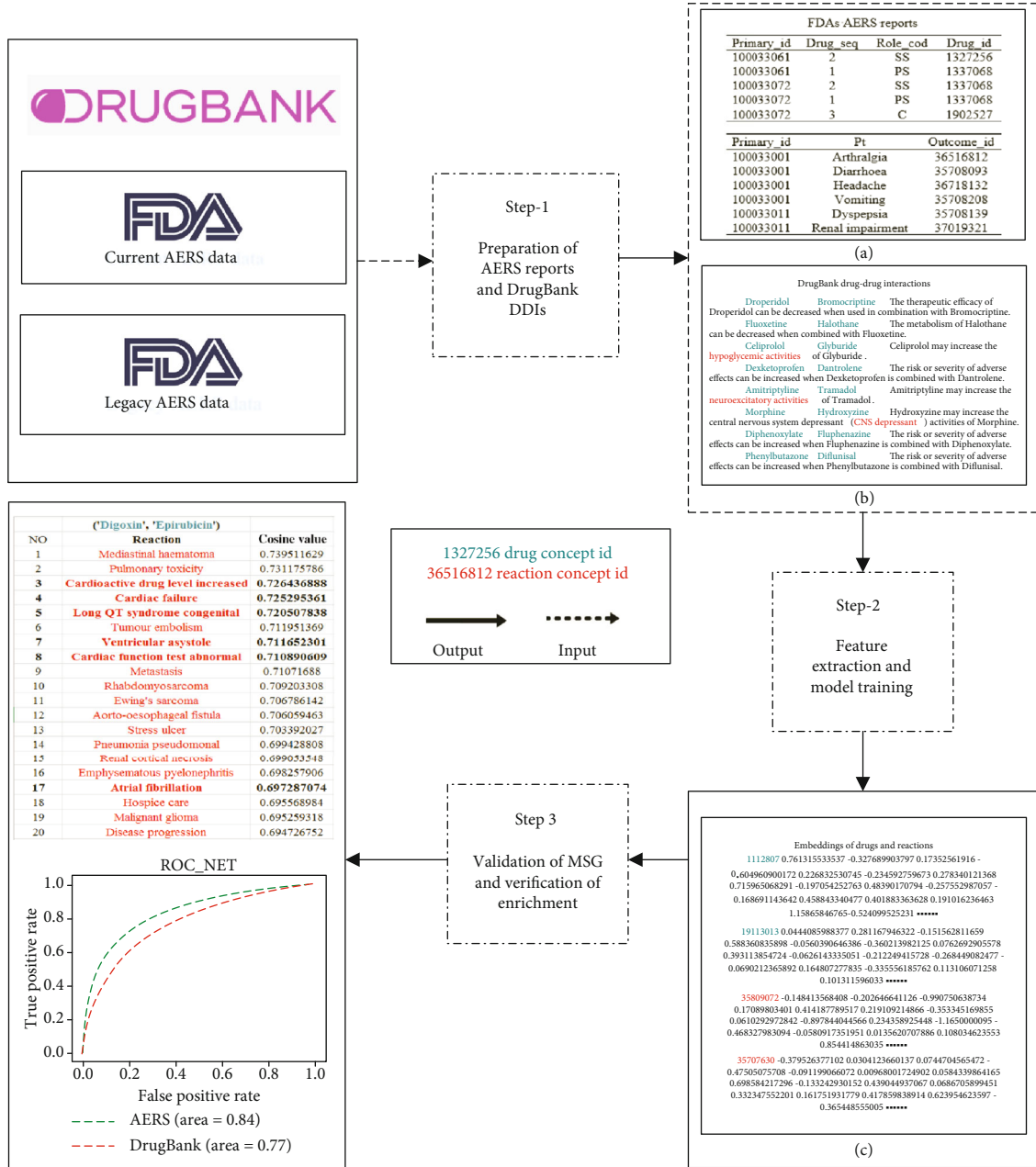


FIGURE 1: The scheme of DDI extraction based on the MSG algorithm.

modified skip-gram model was trained by the hierarchical softmax procedure presented in Mikolov et al.'s research [10]. The objective of the skip-gram model is to maximize the log probability:

$$\sum_{(w,c) \in D} \sum_{j \in c} \log P(w | w_j). \quad (1)$$

In Equation (1),  $c$  is the limited set of drugs and reactions in one AERS report. When  $w$  denotes the drug,  $w_j$  denotes the reaction in the report and vice versa. In the equation

above,  $P(w | w_j)$  can be detailed as follows:

$$P(w | w_j) = \frac{\exp(e'(w)^T e(w_j))}{\sum_{w' \in V} \exp(e'(w')^T e(w_j))}. \quad (2)$$

In Equation (2),  $e'(w)$  is the embedding of drug  $w$  (reaction either),  $w'$  is one of the words in the vocabulary  $V$  consisting of drug name and reaction descriptions.

The key difference between the original skip-gram and modified skip-gram is the way we define "context." In the

original skip-gram, the context is  $2n$  words around the current target word. The  $2n$  words are composed of  $n$  words forward—the current target word—and  $n$  words backward. In default,  $n$  is set in five. According to the particularity orders of words in drug-drug interaction reports, the context of each drug name word is every reaction description word appearing in the sentences of each AERS report in our modified skip-gram. When it comes to reaction description words, the context words  $c$  are changed into every drug name word in the corresponding sentences. As shown in Figure 2, the contexts of Drug<sub>1</sub> are Reaction<sub>1</sub>, Reaction<sub>2</sub>, and Reaction<sub>3</sub>.

**2.3. Cooccurrence Matrix Based on Term Frequency-Inverse Document Frequency (CM-TF-IDF).** Term frequency-inverse document frequency (TF-IDF) is well known as a statistical method for evaluating the importance of one word in the corpus [19]. The importance of the word is increased in direct proportion to how many times it appears in the file and at the same time is declined in inverse proportion to how many times it appears in the whole corpus. In Equation (3),  $n_{i,j}$  is the time word  $t_i$  appearing in the file  $d_j$  and  $\sum_k n_{k,j}$  is the sum of frequencies of all words appearing in the file  $d_j$ .  $|D|$  denotes the total number of documents, and  $|\{j : t_i \in d_j\}|$  is the number of documents which contain the word  $t_i$  in the corpus.

$$\text{tfidf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|D|}{|\{j : t_i \in d_j\}|}. \quad (3)$$

As shown in Figure 3, we constructed a drug name/-reaction description report cooccurrence matrix based on TF-IDF for feature vectorization of drugs and reactions. For example, if Drug<sub>*i*</sub> was recorded in Report<sub>*j*</sub>, the element in the matrix is the tfidf of Drug<sub>*i*</sub>; otherwise, the element is zero.

**2.4. Logistic Regression.** According to the MSG algorithm, drug encoded its reaction information into a low-dimensional vector after the MSG training. CM-TF-IDF also generated drug and reaction vectors from the cooccurrence matrix. These low-dimensional vectors are rendered as features for identifying whether or not the drug pairs are associated with the adverse event class. As logistic regression has been widely used in pharmacovigilance and achieved good performance, it was applied in our research [4–6].

Referring to distinct severe adverse event classes presented by Tatonetti et al. [4], five clinically significant adverse event classes are taken into consideration for binary classification and DDI enrichment: Renal Impairment (REI), Hepatotoxic (HTT), Abnormal Blood Pressure (ABP), Cardiotoxicity (CDT), and Neurotoxic (NET). The logistic regression model requires positive and negative labels which indicate whether or not the pair of drugs is associated with the adverse event class. Because there is no well-recognized gold standard for drug-drug interaction, we crosssearched three datasets (DrugBank\_Toxicity, DrugBank\_DDI, and

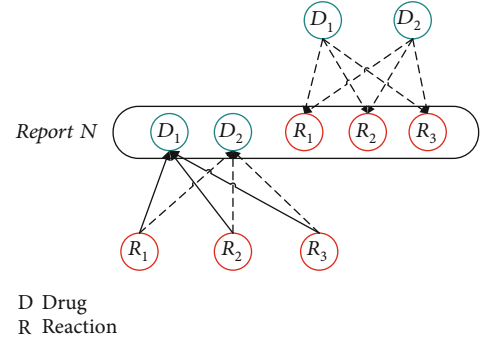


FIGURE 2: Dynamic scope of the window of the modified skip-gram model.

SIDER) and compiled three strategies as follows to define the positive reference samples, see Figure 4.

In the first strategy, if at least one of the drugs in one pair existed in SIDER's specific drug lists where drugs are associated with the adverse event, we labeled this pair of drugs as positive.

In the second strategy, if at least one of the drugs in one pair manifests as an adverse event-associated toxicity in DrugBank\_Toxicity, we labeled this drug pair as positive.

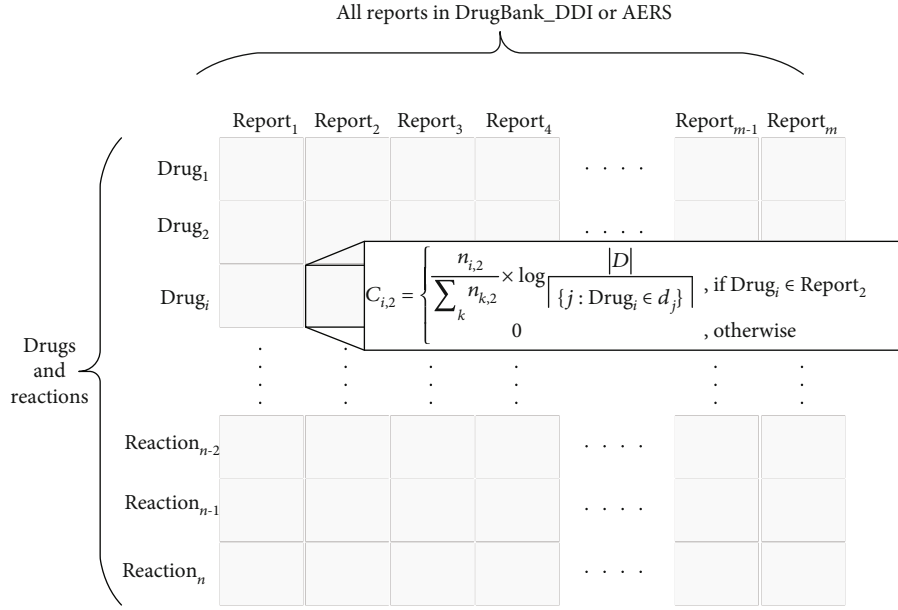
In the third strategy, according to DrugBank\_DDI, if the pair is known to interact which results in the adverse event, the pair is labeled as positive.

### 3. Evaluation and Experiment Results

To assess the performance of the scheme based on MSG, we compared vectors generated from MSG with CM-TF-IDF. The receiver operating characteristic (ROC) is used for the evaluation of binary classifiers [20]. To obtain robust estimates, we performed 10-fold cross-validation; the whole dataset was divided into ten cross-validation splits. During each cross-validation step, a set of nine cross-validation splits was used for model training while the tenth sample set was applied as the test set.

No matter how the embedding was generated, all embedding models are constructed based on the distributional hypothesis. That is to say, if two words have similar context, their value of embeddings is close in the low-dimensional space. Furthermore, the value of drug pair embeddings is theoretically close with its interactions in the low-dimensional space. As a result, we extended this idea to the enrichment of DDIs in DrugBank. Cosine between drug pair embeddings and reaction embeddings was calculated as the reference for ranking.

Although the drug and reaction embeddings were generated after the MSG training, there are no explicit drug pair embeddings. As shown in Equation (3), we empirically constructed drug pair embeddings by addition. For details about variables in Equation (3),  $e_{\text{Drug1}} = (a_1, a_2, \dots, a_n)$  and  $e_{\text{Drug2}} = (b_1, b_2, \dots, b_n)$ .  $a_i$  and  $b_i$  are the values of each  $n$  dimension of drug embedding. Cosine between drug pair embedding and reaction embedding was calculated



- a. "<Lepirudin>\*, <Adapalene>, <Adapalene> may increase the anticoagulantactivities of <Lepirudin>"
- b. "<Adapalene>, <Lepirudin>, <Adapalene> may increase the anticoagulantactivities of <Lepirudin>"
- c. 1112807 705103 | 35607483 36718418 35205025 35809079 36718287  
 1151789 1151789 | 36516812 36516959  
 722424 43526424 | 35708128 35708093  
 715233 735843 | 36718111 36718112 35708100 36416706 35707557  
 1112807 704943 | 35607483 35205025 35809083 36718301  
 1112807 1115008 | 35104074 35707849 35707871
- d. "<Mitomycin>, <Cyclophosphamide>, <Cyclophosphamide> may increase the cardiotoxic activities of <Mitomycin>."
- e. "<Interferon Alfa-2a (Recombinant)>, <Interferon Alfa-2a (Recombinant)> may cause serious adverse effects such as anemia; autoimmune diseases, including vasculitis, arthritis, hemolytic anemia, and erythematosus syndrome; cardiotoxicity; hepatotoxicity; hyperthyroidism or hypothyroidism; transient ischemic attacks; leukopenia; neurotoxicity; peripheral neuropathy; and thrombocytopenia."
- \*Words involved in <> are drug names

FIGURE 4: Examples of DDIs in DrugBank and report in DrugBank\_Toxicity.

according to Equation (4). In Equation (4),  $\mathbf{e}_{\text{Reaction}} = (r_1, r_2, \dots, r_n)$ .  $r_i$  is the value of each  $n$  dimension of reaction embedding.

$$\mathbf{e}_{(\text{Drug1}, \text{Drug2})} = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n), \quad (4)$$

$$\cos \left( \mathbf{e}_{(\text{Drug1}, \text{Drug2})}, \mathbf{e}_{\text{Reaction}} \right) = \frac{\sum_1^n [(a_i + b_i) \times r_i]}{\sqrt{\sum_1^n (a_i + b_i)^2} \times \sqrt{\sum_1^n r_i^2}}. \quad (5)$$

In summary, we sorted cosine of candidate reactions and drug pairs. Finally, top 20 candidate reactions were used to enrich descriptions of DDIs in DrugBank.

In total, 713,441 reports listed only two drugs in FDA AERS reports and 561,180 DDIs in the DrugBank database. We only included the record where drug pairs are listed in RDL. It is worth noting that deduplication of DDIs in DrugBank is also important. As shown in Figures 3(a) and 3(b), there are two DDIs from DrugBank. These two DDIs actually represent the same knowledge, so only one of them was kept for our research. As a result, 218,866 AERS reports and 46,203 DrugBank DDIs were included in our analysis. As shown in Figure 3(c), 218,866 AERS reports were exported from PostgreSQL into plain text format for MSG training. On the left side of the symbol "|" are drug concept ID and right side of the symbol "|" are reaction concept ID. Four crucial parameters of the MSG model are shown in Table 1. "Min count for drugs or reactions" was set to 10 as described in Section 3. "Starting alpha" and "Dimensionality of word

TABLE 1: Parameters of the modified skip-gram model.

	Dimensionality of word embeddings	Starting alpha	Min count for drugs or reactions	Gradient calculation
Parameters	100	0.025	10	Hierarchical softmax

embeddings” were set to default as 0.025 and 100, separately. “Gradient calculation” was set to Hierarchical softmax for performance improvement.

As mentioned in Section 3, five logistic regression models required five sets of samples which consist of positive and negative labels. The detailed distribution of positive samples in three datasets (DrugBank\_DDI, DrugBank\_Toxicity, and SIDER) is listed in Table 2. In column DrugBank\_Toxicity and SIDER, the number is the positive samples of drugs. In column DrugBank\_DDI, the number is the positive samples of drug pairs. For example, as shown in Figure 3(d), the DDIs have the keyword “cardiotoxic”. As a result, the drug pair <Mitomycin, Cyclophosphamide> was one of the 544 positive samples (Table 2) in Cardiotoxicity (CDT) adverse event class. In DrugBank\_Toxicity dataset, as shown in Figure 3(e), the report of drug <Interferon Alfa-2a (Recombinant)> has the keyword “cardiotoxicity”; we included the drug pair as a positive sample in Cardiotoxicity (CDT) adverse event class if the drug pair has drug <Interferon Alfa-2a (Recombinant)>. In the SIDER dataset, we marked the drug pair as a positive sample in Cardiotoxicity (CDT) adverse event class if the drug pair has the drug listed in 448 manually checked drugs (Table 2).

**3.1. Validation of Logistic Regression Models Based on MSG and CM-TF-IDF.** We trained and validated logistic regression models for Renal Impairment (REI), Hepatotoxic (HTT), Abnormal Blood Pressure (ABP), Cardiotoxicity (CDT), and Neurotoxic (NET). When it comes to the vectors generated from MSG, each one of the five clinical significant adverse event classes has two logistic regression models based on AERS reports and the DrugBank DDI dataset individually. When it comes to the vectors generated from CM-TF-IDF, ten logistic regression models are also trained like MSG. In order to avoid dimension disaster of CM-TF-IDF, principal component analysis (PCA) was used for feature dimensionality reduction of CM-TF-IDF. The ROC curve of five adverse event classes based on MSG and CM-TF-IDF is shown in Figures 5 and 6; AUROC of five adverse event classes is shown in Figure 7. As shown in Figure 7, five logistic regression models based on AERS reports achieved a higher value of AUROC than five logistic regression models based on DrugBank DDIs. All AUROC based on MSG in Figure 7 are higher than those based on CM-TF-IDF, which means that our modified skip-gram model can extract features from AERS reports and DrugBank DDI dataset more effectively than the traditional statistical method CM-TF-IDF. At the same time, logistic regression has a good performance of classification in these five adverse event classes as we expected.

TABLE 2: Positive reference samples of five event classes.

Event class	DrugBank_ DDI	DrugBank_ Toxicity	SIDER
Renal Impairment (REI)	117	47	270
Hepatotoxic (HTT)	11	29	265
Abnormal Blood Pressure (ABP)	757	132	275
Cardiotoxicity (CDT)	544	51	448
Neurotoxic (NET)	221	158	298

**3.2. Enrichment of DDIs in DrugBank.** We calculated the cosine of 1,650 DrugBank DDIs for description enrichment and taken MedDRA for verification of description enrichment in five adverse event classes: Renal Impairment (REI), Hepatotoxic (HTT), Abnormal Blood Pressure (ABP), Cardiotoxicity (CDT), and Neurotoxic (NET).

As we know, there are five levels in the MedDRA hierarchy, arranged from specific to general: {System Organ Class (SOC)}, {High level Group Terms (HLGT)}, {High Level Terms (HLT)}, {Preferred Term (PT)}, and {Lowest Level Term (LLT)} [19]. In order to verify the enrichment of DDIs in DrugBank, twenty-seven {System Organ Class (SOC)} are taken into our consideration. Taken Neurotoxic (NET) as example, when we verified the enrichment of DDIs in Neurotoxic (NET), {Nervous system disorders} in {System Organ Class (SOC)} is set as gold standard for the right reactions in Neurotoxic (NET). If at least one of the reactions in top 20 of the drug pair is under the {Nervous system disorders} category, then we define the description enrichment of the drug pair in Neurotoxic (NET) is valid. For example, the enrichment of drug pair <Digoxin, Epirubicin> in Cardiotoxicity (CDT) class is shown in Table 3. Six bold font reactions are verified under the {Cardiac disorders} in System Organ Class (SOC), so the description enrichment of drug pair <Digoxin, Epirubicin> is valid. Table 4 shows the details of DDI enrichment of five classes. In total, 1,456 description enrichments are verified valid, and the average accuracy is 0.882424, which means the description of DDIs in DrugBank is enriched efficiently by using MSG model.

## 4. Discussion

In order to verify and demonstrate the advantage of our presented new scheme, we repeated the whole experiments using the cooccurrence matrix based on tfidf model to generate drug and adverse feature vectors. From the results of ten logistic regression models (as shown in Figure 7 the results show that five logistic regression models based on AERS reports all achieved higher value of AUROC than five logistic regression models based on DrugBank DDIs. In FDA AERS datasets, the vectors generated by the MSG can give better performance in feature extraction than by the tfidf-based cooccurrence matrix model. The main reasons behind the above results are as follows: (1) the cooccurrence matrix based on tfidf model can cause dimensionality disaster when

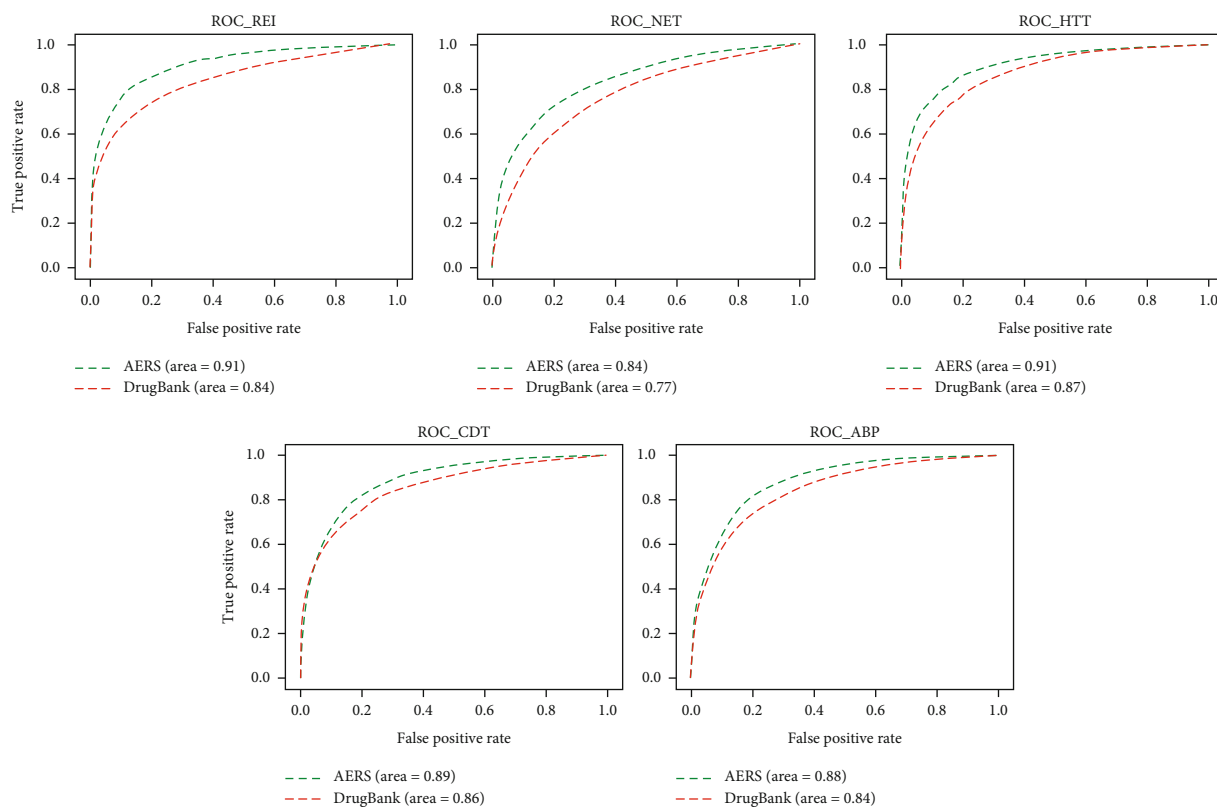


FIGURE 5: ROC of ten logistic regression models based on MSG.

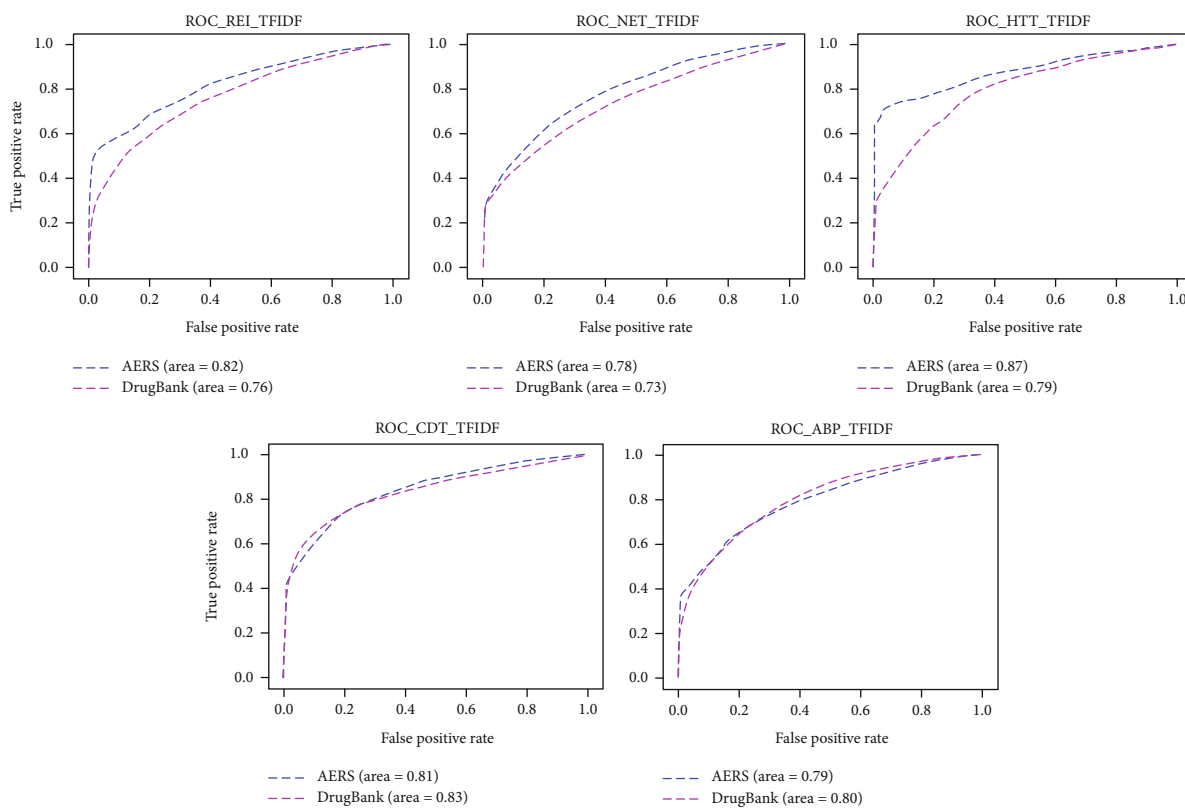


FIGURE 6: ROC of ten logistic regression models based on CM-TF-IDF.



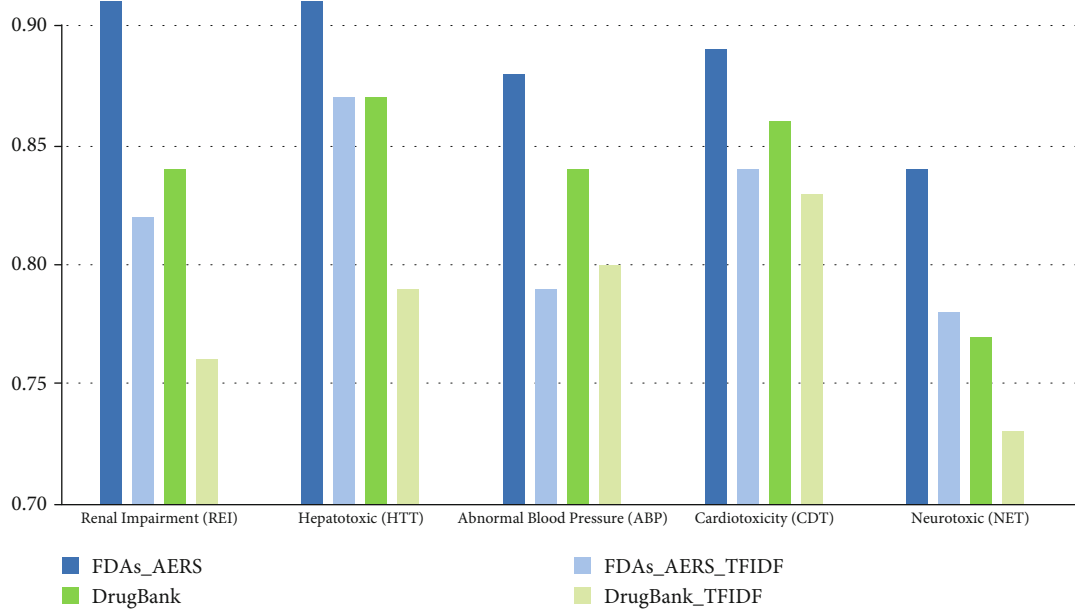


FIGURE 7: AUROC of twenty logistic regression models.

TABLE 3: Enrichment of drug pair &lt;Digoxin, Epirubicin&gt; in DrugBank.

No.	Reaction	Cosine
1	Mediastinal haematoma	0.739511629
2	Pulmonary toxicity	0.731175786
3	<b>Cardioactive drug level increased</b>	<b>0.726436888</b>
4	<b>Cardiac failure</b>	<b>0.725295361</b>
5	<b>Long QT syndrome congenital</b>	<b>0.720507838</b>
6	Tumour embolism	0.711951369
7	<b>Ventricular asystole</b>	<b>0.711652301</b>
8	<b>Cardiac function test abnormal</b>	<b>0.710890609</b>
9	Metastasis	0.71071688
10	Rhabdomyosarcoma	0.709203308
11	Ewing's sarcoma	0.706786142
12	Aorto-oesophageal fistula	0.706059463
13	Stress ulcer	0.703392027
14	Pneumonia pseudomonal	0.699428808
15	Renal cortical necrosis	0.699053548
16	Emphysematous pyelonephritis	0.698257906
17	<b>Atrial fibrillation</b>	<b>0.697287074</b>
18	Hospice care	0.695568984
19	Malignant glioma	0.695259318
20	Disease progression	0.694726752

the data size is large; some features are bound to be lost when using the principal component analysis (PCA). The MSG model defines the dimension of the space vector at initialization, which avoids the work of secondary feature engineering and avoids the loss of feature information. (2) The MSG model constructs a Huffman tree based on word frequency during initialization, and the activation function of each node is softmax, which greatly shortens the time for updating

TABLE 4: Details of drug pair DDI enrichment in DrugBank.

Event class	Number of DDIs	Number of valid enrichments	Accuracy
Renal Impairment (REI)	117	99	0.846154
Hepatotoxic (HTT)	11	9	0.818182
Abnormal Blood Pressure (ABP)	757	660	0.871863
Cardiotoxicity (CDT)	544	494	0.908088
Neurotoxic (NET)	221	194	0.877828
Total	1650	1456	0.882424

weights and vectors in the whole learning process. Because of these reasons, the MSG model can be applied to large-scale datasets compared with the traditional tfidf-based co-occurrence matrix and also can quickly perform feature learning. At the same time, we also found that the MSG model can perform well for the noisy dataset. When MSG model is applied in DrugBank data, all the noisy data are not specifically cleaned after the alignment of drug and adverse reaction strings. However, from the five AUROC values (as shown in Figure 7 DrugB\_MSG), the average AUROC values of the five major adverse reaction groups are around 0.8, which shows that the MSG model can also effectively generate feature vectors from the noise dataset.

## 5. Conclusions

In this work, we proposed an efficient method of feature vector extraction and calculation from FDA AERS and DrugBank texts based on the modified skip-gram model. Feature vectors are taken to expand drug-drug interaction datasets of the DrugBank database. All the accuracy values are higher than 80% (as shown in Table 4) and show that these new

features are valuable in five severe adverse event classes. The contribution of clinicians may accelerate the process of MSG model application in the clinical field.

In the future, on the one hand, we will continue to optimize the accuracy of the word vector and try to integrate the attention mechanism into the language representation algorithm, and on the other hand, we are going to apply the detection of adverse drug reactions to the actual electronic medical record medication prescription system, so as to promptly remind doctors and patients when using drugs.

## Data Availability

We do not want to share our data due to our future works.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Ministry of Science and Technology Key Research and Development Program of China (No. 2018YFC0116902) and National Science Foundation of China (No. 81873915).

## References

- [1] J. Strandell, A. Bate, M. Lindquist, I. R. Edwards, and The Swedish, Finnish, INteraction X-referencing drug-drug interaction database (the SFINX group)3, "Drug-drug interactions - a preventable patient safety issue?," *British Journal of Clinical Pharmacology*, vol. 65, no. 1, pp. 144–146, 2010.
- [2] B. Percha and R. B. Altman, "Informatics confronts drug–drug interactions," *Trends in Pharmacological Sciences*, vol. 34, no. 3, pp. 178–184, 2013.
- [3] R. Harpaz, K. Haerian, H. S. Chase, and C. Friedman, "Statistical mining of potential drug interaction adverse effects in FDA's spontaneous reporting system," *AMIA Annual Symposium proceedings AMIA Symposium*, vol. 7, pp. 281–285, 2010.
- [4] N. P. Tatonetti, G. H. Fernald, and R. B. Altman, "A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports," *Journal of the American Medical Informatics Association*, vol. 19, no. 1, pp. 79–85, 2012.
- [5] F. Cheng and Z. Zhao, "Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. e278–e286, 2014.
- [6] A. Cami, S. Manzi, A. Arnold, and B. Y. Reis, "Pharmacointeraction network models predict unknown drug-drug interactions," *PLoS One*, vol. 8, no. 4, article e61468, 2013.
- [7] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions," *Science translational medicine*, vol. 4, no. 125, article 125ra31, 2012.
- [8] P. Willett, "Similarity-based virtual screening using 2D fingerprints," *Drug Discovery Today*, vol. 11, no. 23–24, pp. 1046–1053, 2006.
- [9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," <https://arxiv.org/abs/1301.3781>.
- [10] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111–3119, 2013.
- [11] D. Chen, R. Socher, C. D. Manning, and A. Y. Ng, "Learning new facts from knowledge bases with neural tensor networks and semantic word vectors," <https://arxiv.org/abs/1301.3618>.
- [12] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep Boltzmann machines," *NIPS Proceedings*, vol. 15, no. 8, pp. 1967–2006, 2012.
- [13] J. Weston, A. Bordes, O. Yakhnenko, and N. Usunier, "Connecting language and knowledge bases with embedding models for relation extraction," <https://arxiv.org/abs/1307.7973>.
- [14] G. Hripcsak, J. D. Duke, N. H. Shah et al., "Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers," *Studies in Health Technology and Informatics*, vol. 216, pp. 574–578, 2015.
- [15] J. M. Banda, L. Evans, R. S. Vanguri, N. P. Tatonetti, P. B. Ryan, and N. H. Shah, "A curated and standardized adverse drug event resource to accelerate drug safety research," *Scientific Data*, vol. 3, no. 1, article 160026, 2016.
- [16] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The SIDER database of drugs and side effects," *Nucleic Acids Research*, vol. 44, no. D1, pp. D1075–D1079, 2016.
- [17] V. Law, C. Knox, Y. Djoumbou et al., "DrugBank 4.0: shedding new light on drug metabolism," *Nucleic Acids Research*, vol. 42, pp. 1091–1097, 2014.
- [18] S. Iyer, "Mining clinical text for signals of adverse drug-drug interactions," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. 353–362, 2014.
- [19] R. Fescharek, J. Kübler, U. Elsasser, M. Frank, and P. Güthlein, "Medical dictionary for regulatory activities (MedDRA)," *International Journal of Pharmaceutical Medicine*, vol. 18, no. 5, pp. 259–269, 2004.
- [20] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, 2000.

## Research Article

# Evaluating the Acute Effect of Stereoscopic Recovery by Dichoptic Stimulation Using Electroencephalogram

Wei Shi <sup>1</sup>, Luyang He <sup>2</sup>, Bin Lv <sup>2</sup>, Li Li <sup>1</sup> and Tongning Wu <sup>2</sup>

<sup>1</sup>Department of Ophthalmology, Beijing Children's Hospital, Capital Medical University, Beijing, China

<sup>2</sup>China Academy of Information and Communications Technology, Beijing, China

Correspondence should be addressed to Tongning Wu; [wutongning@caict.ac.cn](mailto:wutongning@caict.ac.cn)

Received 29 February 2020; Revised 19 March 2020; Accepted 23 March 2020; Published 13 April 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Wei 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.

Amblyopia is a common developmental disorder in adolescents and children. Stereoscopic loss is a symptom of amblyopia that can seriously affect the quality of patient's life. Recent studies have shown that the push-pull perceptual learning protocol had a positive effect on stereoscopic recovery. In this study, we developed a stereoscopic training method using a polarized visualization system according to the push-pull protocol. Dichoptic stimulation for 36 anisometropic and amblyopic subjects and 33 children with normal visual acuity (VA) has been conducted. Electroencephalogram (EEG) was used to evaluate the neurophysiological changes before, during, and after stimulation. For the anisometropic and amblyopic subjects, the statistical analysis demonstrated significant differences ( $p < 0.01$ ) in the beta rhythm at the middle temporal and occipital lobes, while the EEG from the normal VA subjects indicated no significant changes when comparing the results before and after training. We concluded that the dichoptic training in our study can activate the middle temporal visual area and visual cortex. The EEG changes can be used to evaluate the training effects. This study also found that the beta band EEG acquired during visual stimulation at the dorsal visual stream can be potentially used for predicting acute training effect. The results facilitated the optimization of the individual training plan.

## 1. Introduction

Amblyopia is a common developmental visual disorder [1], which reduces the visual acuity (VA) of one or both eyes without obvious defects in the visual pathway, and cannot be solved immediately using eyeglasses or contact lenses [2]. Patients with amblyopia usually present several visual functional deficiencies, such as refractive errors, low sensitivity to contrast or dynamic objects, and limited stereopsis [3, 4], which is estimated to affect 1-5% of the adult population and even more for children [2, 4]. Stereoscopic vision is one of the most advanced visual functions, which provides a sense of depth in the environment and helps develop basic skills, such as grasping, catching, and walking around obstacles at a high speed [5]. Hence, poor stereoscopic depth perception can seriously affect the quality of patient's life. Therefore, it has clinical importance to develop treatment methods for the recovery of stereopsis.

The conventional methods for stereoscopic recovery included monocular occlusion therapy and therapeutic drugs such as atropine [2, 6, 7]. These methods cover or blur the sound eye to force the use of the amblyopic one [8], with the hypothesis that shielding the function of the strong eye facilitated the development of the weak eye. As consequence, the disparity of VA would be levelled. These treatments are often exclusive to children, and adults with stereoscopic deficits are considered untreatable [2]. The use of monacles has been reported to help 50% to 85% of amblyopia children achieve normal VA [6], but with a significant recurrence rate as high as 27% [9]. Dichoptic treatment is another effective approach to solve stereoscopic vision problems [2]. This treatment presents different images to both eyes simultaneously, aiming to reduce intraocular suppression, which is the primary cause for multiple visual deficits [10]. Based on this concept, a series of dichoptic perceptual learning methods became available, including dichoptic videos [11,

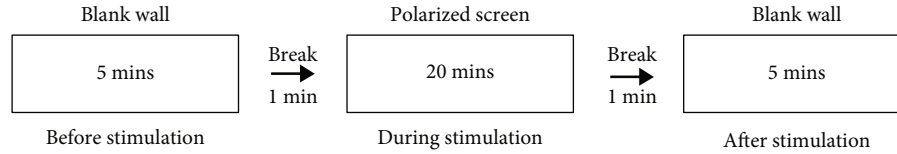


FIGURE 1: Experimental flowchart. This experiment procedure consisted of three steps: before, during, and after stimulation. In each step, the subjects were told to look at blank wall or polarized screen while recording EEG. Between each step, there is a 1-minute break.

12], three-dimensional (3D) video games [13, 14], and virtual reality systems [15, 16]. These studies suggested that VA and stereo sensitivity can be improved through 5 to 30 hours of perceptual learning sessions for children and adults [14, 15]. Among the dichoptic perceptual learning methods, Ooi et al. [17] proposed a push-pull perceptual learning protocol. This protocol can effectively reduce sensory eye dominance and enhance binocular balance [17, 18]. It was designed to stimulate the weak eye (push), while suppressing the perception of the strong eye (pull) in order to recalibrate the binocular balance of excitatory and inhibitory interactions. In contrast, the conventional push-only protocols solely stimulated the weak eye without inhibiting the strong one. This treatment shifted the balance between two eyes towards the weak one and thus improved stereopsis. Numerous studies indicated that this protocol would be promising for stereoscopic recovery [2, 18, 19].

The widely used methods for evaluating stereoscopic recovery were based on clinical examinations, e.g., synoptophore examination, Randot circles test, and Titmus test [20]. These tests focused on assessing stereo sensitivity through behavioural reactions, which may depend on subjective description [2]. Moving toward an objective evaluation for stereoscopic recovery, recent studies have implemented neurophysiologic methods such as functional magnetic resonance imaging (fMRI) [21–25] and functional near-infrared spectroscopy (fNIRS) [26]. The results demonstrated the capability in evaluating stereoscopic recovery. However, those methods required expensive equipment and inconvenient operations, especially unsuitable for children. Besides, the abovementioned methods were incapable of monitoring real-time neurophysiological changes during training. To solve the problem, electroencephalogram (EEG) has also been used in the relevant study. For example, Deng et al. assessed binocular processing differences following perceptual learning in adult anisometropic patients using steady-state visual evoked potential [27].

In our study, we developed a dichoptic stimulation system based on the push-pull protocol. The applicability of EEG to assess the training effect was investigated. The possibility of using EEG as a biomarker to predict the acute training effect has been discussed as well.

## 2. Materials and Methods

**2.1. Subjects.** In the present study, 69 children (4 to 17 years old, mean  $\pm$  standard deviation:  $7.1 \pm 2.7$  years old) were recruited from the Department of Ophthalmology of Beijing Children's Hospital, Capital Medical University from April 2017 to August 2018. Thirty-six children were diagnosed

with anisometropia and amblyopia, and the other 33 were children with normal VA. All subjects and their parents were informed about the experimental protocol, and written informed consent was obtained. Our experiments were approved by the ethics committee of Beijing Children's Hospital.

**2.2. EEG Data Acquisition and Associated Procedures.** EEG data were recorded using a NeuroScan SynAmps2 64-channel EEG amplifier (Compumedics, Victoria, Australia) with a 64-channel elastic cap (Quik-Cap, NeuroScan) at a sampling rate of 1000 Hz. The impedance of all electrodes was adjusted to less than 10 k $\Omega$ , and the mean values of M1 and M2 (according to the international 10-20 EEG system) were set as the reference. Before training, all subjects were well rested and stayed alert while sitting in a comfortable chair. At the beginning of training, subjects were told to look at a blank wall, and EEG signals were recorded for 5 minutes. During the stimulation, subjects placed their head in a chinrest to ensure screen alignment (the centre of the two eyes was aligned to the centre of the screen) and observed the stimuli for 20 minutes. EEG signals were recorded for 20 minutes simultaneously. We then recorded the EEG signal for 5 minutes after the visual stimulus, while the subjects looked at the blank wall again. The experimental flow chart is shown in Figure 1.

**2.3. Dichoptic Stimulation.** The dichoptic stimulation was performed in a clinical assessment room with constant luminance. The experiment was conducted using a PC and an LG D2343P polarized 3D monitor (LG Electronics, Seoul, South Korea) with a resolution of  $1920 \times 1080$  and a refreshing rate of 120 Hz. All subjects wore polarized glasses and their corrective lenses. The visual stimuli were developed using the push-pull perceptual learning protocol and presented at an observation distance of 80 cm, with a background brightness of 35 candelas *per* square meter ( $\text{cd}/\text{m}^2$ ). During the stimulation, the rivaling half image to the amblyopic eye was perceived (push), while the half image to the strong eye was perceptually suppressed (pull) (see Figure 2). The image size for the strong eye was  $200 \times 200$  pixels, and the image size for the weak eye had four levels:  $200 \times 200$ ,  $400 \times 400$ ,  $600 \times 600$ , and  $800 \times 800$  pixels. The contrast ratio for weak eye was set as 100% during the stimulation, while 5% to 50% for the strong eye. The selection of the image size and contrast ratio depended on the extent of the VA disparity. For normal VA subjects, the image size for both eyes was set at  $200 \times 200$  pixels, and contrast ratio was set at 100%. Figure 3 shows an experimental scenario in which a subject



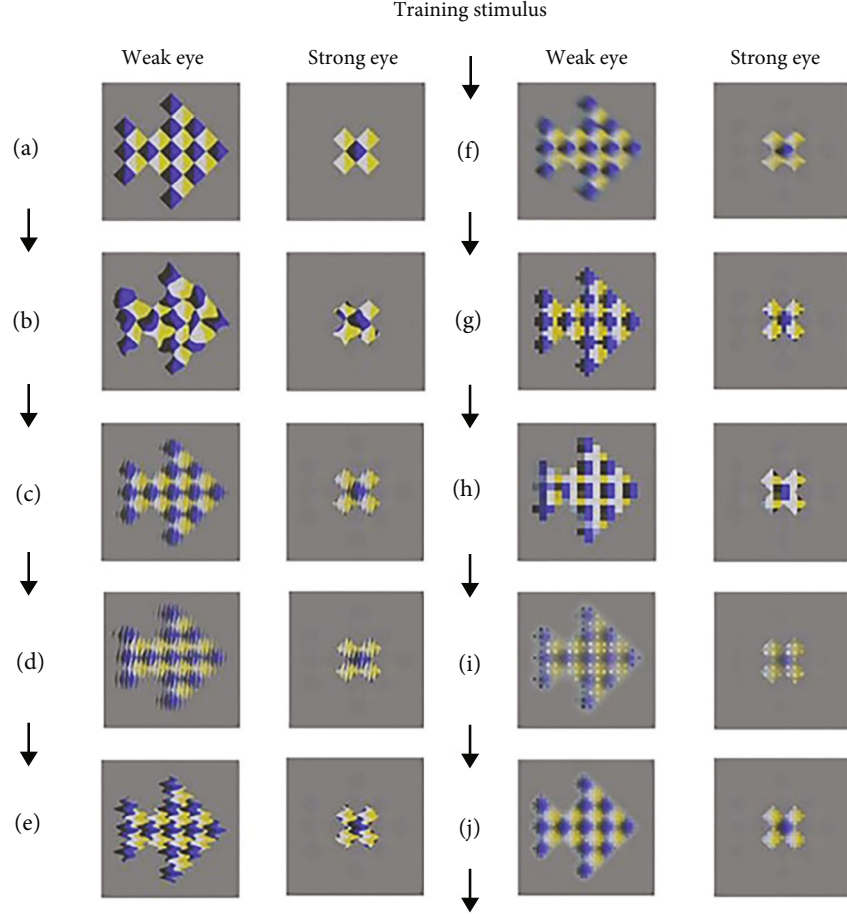


FIGURE 2: Push-pull training scenario used in the study. The initial images perceived by the two eyes are shown in (a). During the stimulation process, the images seen by both eyes were synchronously processed and switched every 500 milliseconds and looped from (a) to (j) over 5 seconds, allowing the half image viewed by the weak amblyopic eye to be perceived and the one viewed by the strong eye to be suppressed.

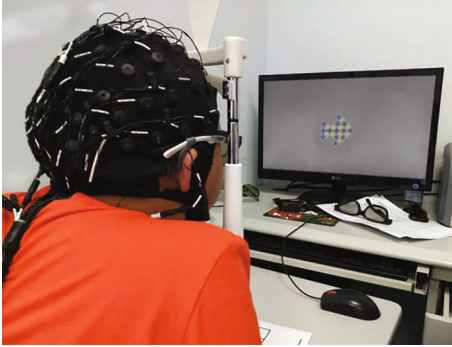


FIGURE 3: Experimental scenario. One subject was looking at a polarized screen while recording EEG.

was looking at the stimulus image on the screen, while the EEG was being recorded.

**2.4. EEG Data Analysis and Statistical Analysis.** The raw EEG signals were processed offline using Curry 7.0 (Compumedics, Victoria, Australia) and MATLAB (MathWorks, Natick, MA). The preprocessing of the raw EEG data included the use of band-pass filters (0.1-30 Hz), rereferen-

cing (the mastoids are chosen as reference electrodes), and independent component analysis (ICA) for artefact removal (including eye movements, temporal muscle activity, and linear noise). The procedures were realised by EEGLAB [28]. We calculated the power spectral density (PSD) of each electrode for three different states, i.e., before, during, and after stimulation, using Welch method [29]. Moreover, four frequency bands were analyzed in the study, i.e., delta (0.1 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 14 Hz), and beta (14 to 30 Hz). To reduce the impact of individual differences, the absolute PSD in each frequency band was normalized by dividing by the total PSD to obtain the relative PSD, which was used for statistical analysis.

A two-way analysis of variance (ANOVA) was performed (*per* channel and *per* frequency band) to evaluate the resulting PSD changes with the factors of VA disparity and stimulation state. VA was converted to the logarithm of the minimum angle of resolution (LogMAR). Accordingly, the subjects were divided into three levels: 0 to 0.2 (2 lines in the visual chart of LogMAR) for mild VA disparity, 0.2 to 0.4 (2 to 4 lines) for moderate VA disparity, and above 0.4 for severe VA disparity (exceeding 4 lines). There were 10 subjects with mild VA disparity, 13 subjects with moderate VA disparity, and 13 with severe VA disparity. The three



stimulation states were before, during, and after stimulation. In order to assess the effect relating to different ages, we also divided 69 subjects into two groups, i.e., 16 patients plus 18 normal VA subjects in the group below 7 years old and 20 patients plus 15 normal VA subjects in the group exceeding 7 years old. We performed independent  $t$  tests for the normal VA subjects and the patients, respectively, to assess the between-group difference.

Moreover, it would be of great interest to investigate the correlation between the EEG change during stimulation and the treatment effect, because it can provide insight on predicting the individual recovery effect which enabled the optimization of the training contents. The statistical experiments were designed for the 36 children with anisometropia and amblyopia. Correlation between recovery effect (calculated by subtracting the poststimulation VA disparity from the prestimulation VA disparity) and the change of relative PSD was investigated. The analyses were conducted per channel and per band.

The statistical analyses were performed using SPSS 25.0 (IBM; Armonk, NY).

### 3. Results

Figure 4 illustrates the EEG topographic maps averaged across all 36 amblyopia subjects and 33 normal VA subjects in four frequency bands. Two-way ANOVA yielded significant differences for the interaction effect of VA disparity and stimulation state in the beta pattern, TP7,  $p = 0.004$ ,  $F(4, 198) = 4.016$ . There were no other channels or frequency bands with significant differences for the interaction effect. For the main effect of VA disparity, we found significant differences in the delta band (C6, T8, C1, CZ, CP1, CPZ,  $p < 0.01$ ) and the theta band (F7, F5, F6, F8, TP7, T8,  $p < 0.01$ ). For the main effect of stimulation state, significant differences were found in the beta (C1, C6, P1, P6, T7, P8, TP7, TP8, O1, O2, OZ,  $p < 0.01$ ) and delta bands (FP1, FPZ, FP2, AF3, AF4, F7, F5, F6, F8, TP7, T8,  $p < 0.01$ ). We used the least significant difference (LSD) for the post hoc test. In terms of after vs. before stimulation, the post hoc test yielded significant differences in the beta bands across six channels: T7 ( $p = 0.009$ ), P8 ( $p = 0.007$ ), TP8 ( $p = 0.008$ ), O1 ( $p = 0.004$ ), OZ ( $p = 0.001$ ), and O2 ( $p = 0.007$ ). There was no significant change after vs. before stimulation in the delta frequency bands. We also found significant differences in the beta and delta rhythms after vs. during stimulation (beta: all channels; delta: FP1, FPZ, FP2, AF3, AF4, F7, F5, F6,  $p < 0.01$ ), including occipital areas and the temporal lobes.

To compare the EEG differences between normal subjects and subjects with anisometropia, we performed  $t$  tests for before and after the stimulation on the relative PSDs of the four frequency bands of the 33 normal VA subjects and the 36 anisometric subjects (per channel). The  $t$  test results from the normal VA subjects revealed no significant differences across all channels and all frequency bands, but the results from the anisometric patients revealed significant differences in the beta frequency PSDs on 25 channels (FT7, FC5, FC3, T7, C5, C3, C4, C6, T8, TP7, CP5, CP3,

CP4, CP6, TP8, P5, P3, P1, P4, P6, P8, O1, OZ, O2). The anisometric subjects had no significant differences in any of the channels in the alpha, theta, or delta bands. The  $t$  tests for between-group difference showed no significant changes from the normal VA subjects and the patients in the two age groups.

The correlation analysis showed that high correlation ( $r > 0.8$ ) for two channels at occipital region (O1 and OZ). Five channels at bilateral middle temporal regions (FT7, T8, T7, P8, and TP8) showed correlation coefficients between 0.5 and 0.8. The detected significant correlation was found at beta band. The channels are shown in Figure 5.

### 4. Discussion

After visual training, amblyopic patients showed significantly changed beta band PSD at occipital and middle temporal regions. In contrast, no difference was detected between before and after training from normal VA subjects. The finding revealed that the stimulation based on the push-pull protocol had limited effects on normal VA subjects. The comparison showed no difference between the two age groups, which may indicate that the applied stimulus pattern was not age-dependent.

Amblyopia is believed to result from a dysfunction of the cortical developmental processes [23, 24], although the exact extent of this visual cortical deficit is currently unknown. As a result, functional neuroimaging techniques have been used to study the changes at the neurophysiological level after treatment [25]. For example, Yang et al. studied blood oxygen level-dependent activity in amblyopia patients before and after levodopa treatment [24]. Similarly, Gupta et al. studied hemodynamic changes in individuals with strabismic amblyopia using fMRI and diffusion tensor imaging before and after eye patching [25]. In addition, Iwata et al. used fNIRS to assess the differences in oxygenated haemoglobin concentration changes between binocular and monocular treatment [26]. Their results revealed that during or after stimulation, the occipital region demonstrated neurophysiological differences, such as an elevated number of active voxels or blood oxygen levels, resulting in improved stereopsis after stimulations. Numerous studies have shown that during visual stimulation, beta band or gamma band power increases in occipital-parietal brain regions [27, 28]. The increase of high-frequency power was associated with enhanced neuronal discharge activity [30], which was likely linked to the increase of blood-oxygen concentration. The change in the delta band may be associated with fatigue. In all, the EEG results were consistent with reports using fNIRS and fMRI. In view of its mobility and convenience, EEG was advantageous in evaluating the acute effect of stereoscopic training.

There was a significant difference in beta power spectral density in the primary visual cortex, and there was also a significant difference in the middle temporal lobe region, where MT was located [31]. MT was a specific region of the dorsal extrastriate processing stream, which was highly motion sensitive and can integrate local motion signals into a coherent motion representation [32, 33]. In fact, converging evidence from human psychophysics and animal neurophysiology

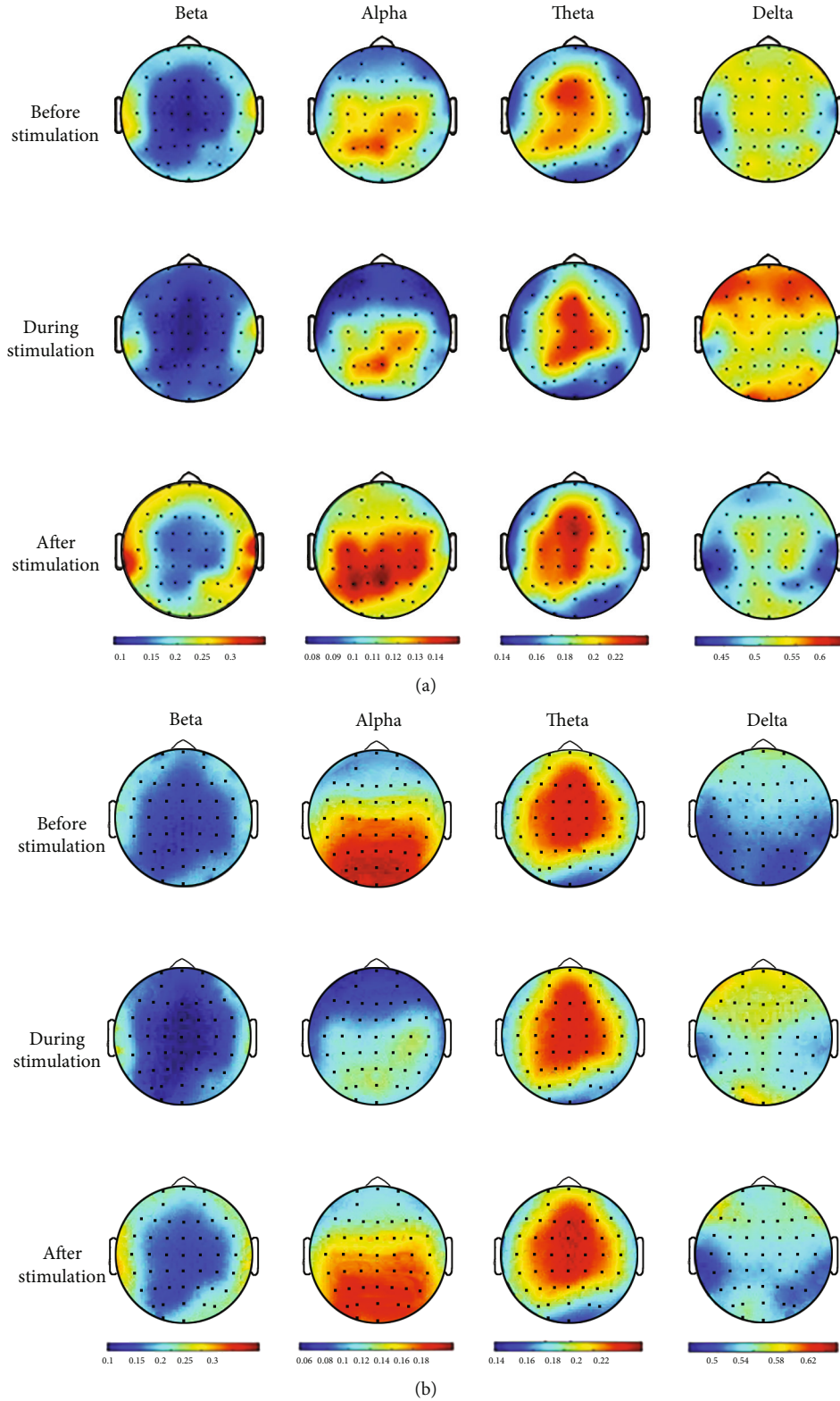


FIGURE 4: Mean EEG topographic maps over 36 anisometric subjects (a) and 33 normal VA subjects (b). Each column represents the distribution of the averaged relative PSDs of the EEGs for the four frequency bands, while each row represents different stimulation states (before, during, and after stimulation).

has indicated that stereopsis was associated with abnormal MT function. Our study showed that after visual stimulation, the temporal lobe area of the anisometric subjects was

active (compared to before stimulation), while that of the normal subjects was not, suggesting that MT played an important role in the formation of anisometropia. Dichoptic

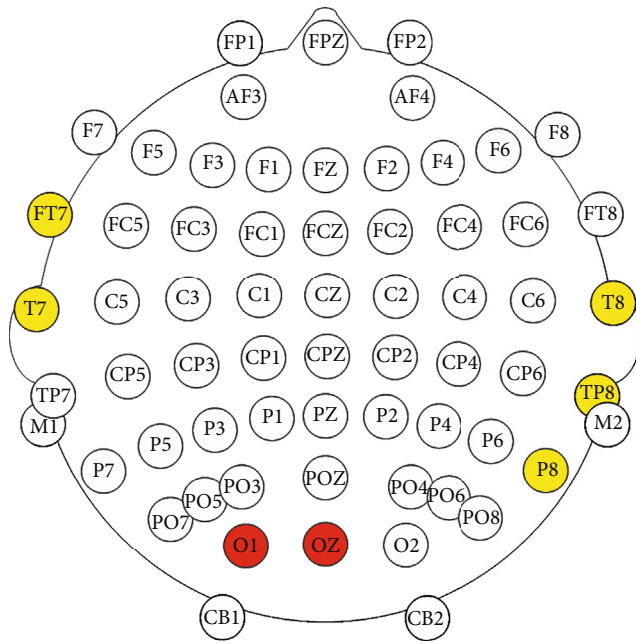


FIGURE 5: EEG channels showed significant correlation between relative PSD change and training effect. Red colour indicates the channels with correlation coefficient beyond 0.8, while yellow colour indicates the channels with correlation coefficient between 0.5 and 0.8.

stimulation in our study was able to stimulate MT as well as the visual cortex, thus possibly creating a higher chance of restoring the stereo sensitivity of anisometropia.

Analysis focusing on patients with anisometropia and amblyopia clearly demonstrated that the relative PSD change during training and the training effect correlated significantly along the dorsal extrastriate processing stream. The most intense correlation was found at area V1, which was essential to the conscious processing of visual stimuli [34]. In addition, damage to V1 may lead to disruption or even loss specific aspects of vision (e.g., depth perception [35]). The strong correlation between the PSD change and the posttraining effect along the dorsal visual stream revealed that these regions may closely associate with the acute treatment effect. It also indicated the possibility to use the EEG data from this area as an indicator for choosing the desired training contents. As such, it has the potential to act as a biomarker for predicting the stimulating effect and for adjusting the individual training plan (the stimulation length or the content). The features derived from the EEG on these channels may also be employed as a classifier to identify the appropriate stimulation method for the patients, which would be investigated in our further studies when more subjects were recruited.

The limitation of the study was the limit number of the subjects under evaluation. In addition, we did not track their long-term training effect due to the mobility of the subjects.

## 5. Conclusions

In our study, we focused on the stereoscopic recovery of amblyopic patients. We designed a dichoptic stimulation system adopted from push-pull perceptual learning using polar-

ized glasses, performed training on 33 normal VA children and 36 anisometropic children, and evaluated its acute effects through EEG. After a 20-min dichoptic stimulation, we detected enhanced beta rhythm PSD after stimulation in the visual cortex and MT of the anisometropic subjects. Our study provided evidence indicating that dichoptic training was able to stimulate MT and the primary visual cortex, and EEG acquired in those regions has potential applications in evaluating the acute effect of stereoscopic training. The study also discussed the possibility of using the EEG signal as biomarker to predict the acute training effect. In future studies, we will enroll more subjects and investigate the use of this EEG feature to optimizing the individual training plan.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

This work was supported by the Beijing Scientific Research and Development Program, [PX2017003], National Science and Technology Major Project, [2018ZX10301201], Wu Jieping Medical Foundation, [320.6750.17179], and National Natural Science Foundation Project, [61971445 and 61671158].

## References

- [1] B. T. Barrett, A. Bradley, and T. R. Candy, "The relationship between anisometropia and amblyopia," *Progress in Retinal and Eye Research*, vol. 36, pp. 120–158, 2013.
- [2] D. M. Levi, D. C. Knill, and D. Bavelier, "Stereopsis and amblyopia: a mini-review," *Vision Research*, vol. 114, pp. 17–30, 2015.
- [3] K. Simons, "Amblyopia characterization, treatment, and prophylaxis," *Survey of Ophthalmology*, vol. 50, no. 2, pp. 123–166, 2005.
- [4] R. F. Hess, B. Mansouri, S. C. Dakin, and H. A. Allen, "Integration of local motion is normal in amblyopia," *Journal of the Optical Society of America A*, vol. 23, no. 5, p. 986, 2006.
- [5] M. Hayhoe, B. Gillam, K. Chajka, and E. Vecellio, "The role of binocular vision in walking," *Visual Neuroscience*, vol. 26, no. 1, pp. 73–80, 2009.
- [6] E. E. Birch, "Amblyopia and binocular vision," *Progress in Retinal and Eye Research*, vol. 33, no. 1, pp. 67–84, 2013.
- [7] D. P. Spiegel, J. Li, R. F. Hess et al., "Transcranial direct current stimulation enhances recovery of stereopsis in adults with amblyopia," *Neurotherapeutics*, vol. 10, no. 4, pp. 831–839, 2013.
- [8] T. Li and K. Shotton, "Cochrane Review: Conventional occlusion versus pharmacologic penalization for amblyopia," *Evidence-Based Child Health: A Cochrane Review Journal*, vol. 5, no. 4, pp. 1873–1909, 2010.



- [9] R. Bhola, R. V. Keech, P. Kutschke, W. Pfeifer, and W. E. Scott, "Recurrence of amblyopia after occlusion therapy," *Ophthalmology*, vol. 113, no. 11, pp. 2097–2100, 2006.
- [10] J. Li, B. Thompson, C. S. Y. Lam et al., "The role of suppression in amblyopia," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 7, pp. 4169–4176, 2011.
- [11] R. F. Hess, B. Mansouri, and B. Thompson, "A binocular approach to treating amblyopia: antisuppression therapy," *Optometry and Vision Science*, vol. 87, no. 9, pp. 697–704, 2010.
- [12] J. Ding and D. M. Levi, "Recovery of stereopsis through perceptual learning in human adults with abnormal binocular vision," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 37, pp. E733–E741, 2011.
- [13] R. F. Hess, B. Thompson, and D. H. Baker, "Binocular vision in amblyopia: structure, suppression and plasticity," *Ophthalmic and Physiological Optics*, vol. 34, no. 2, pp. 146–162, 2014.
- [14] S. L. Li, R. M. Jost, S. E. Morale et al., "A binocular iPad treatment for amblyopic children," *Eye*, vol. 28, no. 10, pp. 1246–1253, 2014.
- [15] D. Levi, I. Vedamurthy, D. Knill, J. Ding, O.-S. Kwon, and D. Bavelier, "Recovering stereo vision by squashing virtual bugs in a VR environment," *Journal of Vision*, vol. 16, no. 12, p. 201, 2016.
- [16] P. Žiak, A. Holm, J. Halička, P. Mojžiš, and D. P. Piñero, "Amblyopia treatment of adults with dichoptic training using the virtual reality oculus rift head mounted display: preliminary results," *BMC Ophthalmology*, vol. 17, no. 1, pp. 105–108, 2017.
- [17] T. L. Ooi, Y. R. Su, D. M. Natale, and Z. J. He, "A push-pull treatment for strengthening the 'lazy eye' in amblyopia," *Current Biology*, vol. 23, no. 8, pp. R309–R310, 2013.
- [18] J. P. Xu, Z. J. He, and T. L. Ooi, "Effectively reducing sensory eye dominance with a push-pull perceptual learning protocol," *Current Biology*, vol. 20, no. 20, pp. 1864–1868, 2010.
- [19] J. P. Xu, Z. J. He, and T. L. Ooi, "Push-pull training reduces foveal sensory eye dominance within the early visual channels," *Vision Research*, vol. 61, no. 61, pp. 48–59, 2012.
- [20] P. J. Knox, A. J. Simmers, L. S. Gray, and M. Cleary, "An exploratory study: prolonged periods of binocular stimulation can provide an effective treatment for childhood amblyopia," *Investigative Ophthalmology & Visual Science*, vol. 53, no. 2, pp. 817–824, 2012.
- [21] K. Shibata, Y. Sasaki, M. Kawato, and T. Watanabe, "Neuroimaging evidence for 2 types of plasticity in association with visual perceptual learning," *Cerebral Cortex*, vol. 26, no. 9, pp. 3681–3689, 2016.
- [22] C. S. Ho and D. E. Giaschi, "Low- and high-level motion perception deficits in anisometropic and strabismic amblyopia: evidence from fMRI," *Vision Research*, vol. 49, no. 24, pp. 2891–2901, 2009.
- [23] Z. Tan, Y. Li, D. Zang et al., "Altered regional homogeneity in epileptic patients with infantile spasm: a resting-state fMRI study," *Journal of X-Ray Science and Technology*, vol. 24, no. 2, pp. 285–295, 2016.
- [24] C. I. Yang, M. L. Yang, J. C. Huang et al., "Functional MRI of amblyopia before and after levodopa," *Neuroscience Letters*, vol. 339, no. 1, pp. 49–52, 2003.
- [25] S. Gupta, S. S. Kumaran, R. Saxena, S. Gudwani, V. Menon, and P. Sharma, "BOLD fMRI and DTI in strabismic amblyopes following occlusion therapy," *International Ophthalmology*, vol. 36, no. 4, pp. 557–568, 2016.
- [26] Y. Iwata, T. Handa, H. Ishikawa, N. Shoji, and K. Shimizu, "Efficacy of an amblyopia treatment program with both eyes open: a functional near-infrared spectroscopy study," *American Orthoptic Journal*, vol. 66, no. 1, pp. 87–91, 2016.
- [27] S. Deng, L. Gu, Z. L. Lu et al., "EEG Recordings during Binocular Rivalry Reveals Changes of Binocular Interaction Following Perceptual Learning in Adult Amblyopia," *Investigative Ophthalmology & Visual Science*, vol. 59, 2018.
- [28] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.
- [29] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified Periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
- [30] S. Nakamura, N. Sadato, T. Oohashi, E. Nishina, Y. Fuwamoto, and Y. Yonekura, "Analysis of music-brain interaction with simultaneous measurement of regional cerebral blood flow and electroencephalogram beta rhythm in human subjects," *Neuroscience Letters*, vol. 275, no. 3, pp. 222–226, 1999.
- [31] L. G. Ungerleider and J. V. Haxby, "'What' and 'where' in the human brain," *Current Opinion in Neurobiology*, vol. 4, no. 2, pp. 157–165, 1994.
- [32] B. Thompson, M. Y. Villeneuve, C. Casanova, and R. F. Hess, "Abnormal cortical processing of pattern motion in amblyopia: evidence from fMRI," *NeuroImage*, vol. 60, no. 2, pp. 1307–1315, 2012.
- [33] A. J. Simmers, T. Ledgeway, B. Mansouri, C. V. Hutchinson, and R. F. Hess, "The extent of the dorsal extra-striate deficit in amblyopia," *Vision Research*, vol. 46, no. 16, pp. 2571–2580, 2006.
- [34] D. Leopold, "Primary visual cortex: awareness and blindsight," *Annual Review of Neuroscience*, vol. 35, pp. 91–109, 2012.
- [35] S. Henriksen, S. Tanabe, and B. Cumming, "Disparity processing in primary visual cortex," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1697, article 20150255, 2016.

## Research Article

# A Novel Deep Neural Network for Robust Detection of Seizures Using EEG Signals

Wei Zhao <sup>1</sup>, Wenbing Zhao <sup>2</sup>, Wenfeng Wang <sup>3</sup>, Xiaolu Jiang <sup>1</sup>, Xiaodong Zhang <sup>4</sup>,  
Yonghong Peng <sup>5</sup>, Baocan Zhang<sup>1</sup> and Guokai Zhang <sup>6</sup>

<sup>1</sup>Chengyi University College, Jimei University, Xiamen 361021, China

<sup>2</sup>Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, Ohio 44115, USA

<sup>3</sup>School of Electronic and Electrical Engineering, Shanghai Institute of Technology, Shanghai 200235, China

<sup>4</sup>Department of Ultrasound, The First Affiliated Hospital of Xiamen University, Xiamen 361005, China

<sup>5</sup>Faculty of Computer Science, University of Sunderland, Sunderland, UK

<sup>6</sup>School of Software Engineering, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Wenfeng Wang; wangwenfeng@sit.edu.cn, Xiaodong Zhang; zxdon11@163.com, and Guokai Zhang; zhangguokai\_01@163.com

Received 2 February 2020; Accepted 6 March 2020; Published 7 April 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Wei 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.

The detection of recorded epileptic seizure activity in electroencephalogram (EEG) segments is crucial for the classification of seizures. Manual recognition is a time-consuming and laborious process that places a heavy burden on neurologists, and hence, the automatic identification of epilepsy has become an important issue. Traditional EEG recognition models largely depend on artificial experience and are of weak generalization ability. To break these limitations, we propose a novel one-dimensional deep neural network for robust detection of seizures, which composes of three convolutional blocks and three fully connected layers. Thereinto, each convolutional block consists of five types of layers: convolutional layer, batch normalization layer, nonlinear activation layer, dropout layer, and max-pooling layer. Model performance is evaluated on the University of Bonn dataset, which achieves the accuracy of 97.63%~99.52% in the two-class classification problem, 96.73%~98.06% in the three-class EEG classification problem, and 93.55% in classifying the complicated five-class problem.

## 1. Introduction

Electroencephalogram (EEG) is a noninvasive, effective technique used in clinical studies to decode the electrical activity of the brain. EEG is one of the critical technologies to identify an abnormality of the brain, such as detecting epileptic seizures. Seizures are transient neurological dysfunctions caused by abnormal brain neurons and excessive supersynchronized discharges. The visual inspection of EEG for seizure detection by expert neurologists is a time-consuming and laborious process, and the diagnosis may not be accurate because of the massive amounts of EEG data and the discrepant clinical judgment standards of different neurologists [1, 2]. Therefore, scientific research on EEG-based automatic detection of epilepsy has attracted much attention.

Numerous algorithms have been proposed in the literature for automatic detection of epileptic seizures. These methods can be roughly classified into two categories: conventional methods and deep learning- (DL-) based methods. Thereinto, most of the traditional methods use hand-engineered techniques for feature extraction from EEG signals and then conjunct with classifiers to recognize. The Bonn University EEG database is widely used, which is publicly available and labeled as A, B, C, D, and E. Details of the dataset are described in a later section. There is much published work using the Bonn dataset for epilepsy detection. They concern three main classification problems: the two-class seizure detection problem focuses on the classification between nonseizures and seizures; the three-class epileptic classification problem focuses on the grouping of



three different EEG categories (normal, interictal, and ictal); and the five-class recognition problem focuses on the classification of five distinct types (A, B, C, D, and E).

In 2009, Ocak [3] proposed a scheme for detecting epileptic seizures based on approximate entropy and discrete wavelet transform (DWT) of EEG signals. This framework obtained an accuracy of 96% for two-class EEG classification. Moreover, Tzallas et al. [4] demonstrated the suitability of the time-frequency analysis (TFA) to classify EEG segments for epileptic seizures. The authors employed the artificial neural network (ANN) as the classifier and achieved an accuracy of 100% for the two-class and three-class classification and 89% for the five-class case. In 2010, Subasi and Ismail Gursoy [5] employed principal component analysis, independent component analysis, and linear discriminant analysis to reduce the dimension of EEG signals, extracted statistical features from DWT, and then used support vector machine (SVM) for classification. This model yielded a seizure detection accuracy of 100% for two-class classification. In 2011, Orhan et al. [6] used the k-means algorithm to cluster from the wavelet coefficients and then classified a multilayer perceptron neural network (MLPNN). This model yielded maximum accuracy of two-class and three-class classifications that are 100% and 96.67%, respectively. In 2012, Acharya et al. [7] proposed a methodology for the automatic detection of normal, interictal, and ictal categories from EEG signals. They extracted four entropy features and then fed to a fuzzy classifier. This methodology achieved an accuracy of 98.1%. In 2014, Kaya et al. [8] used the one-dimensional local binary pattern (1-D-LBP) to extract features from raw EEG and, respectively, combined with five different classifiers, such as Bayes Net, SVM, ANN, logistic regression (LR), and functional tree (FT). The best-performing classifier was the Bayes Net classifier, which achieved 99.5% and 95.67% maximum accuracy for two-class and three-class classifications, respectively. The worst performing classifier was the LR classifier, which gained 96.50% and 66.67% maximum accuracy for two-class and three-class classifications, respectively. In 2015, Sharma and Pachori [9] proposed the features based on the phase space representation for the classification of epileptic seizure and seizure-free EEG signals. They employed the least squares support vector machine as a classifier, which gave 98.67% accuracy. In 2016, Sharmila and Geethanjali [10] studied the performance of the 14 different combinations of two-class epilepsy detection. They employed naive Bayes (NB) and k-nearest neighbor (KNN) classifiers for the derived statistical features from DWT, and the NB classifier obtained an accuracy of 100% in the classification of healthy eyes open and epileptic EEG data. In 2017, Zhang and Chen [1] employed local mean decomposition (LMD) to decompose raw EEG signals into several product functions (PFs) and then fed the features into five classifiers. The authors reported that the best-performing classifier was the SVM optimized by genetic algorithm (GA-SVM), and the average classification accuracy was equal to or higher than 98.1%. Bhattacharyya et al. [11] computed the Q-based entropy by decomposing the signal with the tunable-Q wavelet transform (TQWT) into the number of

subbands and estimating K-nearest neighbor entropies (KNNE) from various subband cumulatively and used the support vector machine classifier with the wrapper-based feature selection method to be the classifier. This method achieved an accuracy of 100% and 98.6% of maximum efficiency for two-class and three-class classifications, respectively. Zahra et al. [12] presented a data-driven approach to classify five-class EEG classification using the multivariate empirical mode decomposition (MEMD) algorithm. And ANN was employed to be a classifier, which achieved 87.2% accuracy.

These conventional methods for the detection of seizures use hand-engineered techniques to extract features from EEG signals. And many of these traditional methods show good accuracy for one problem but fail in performing accurately for others [2]. For example, they identify nonseizure and seizure cases (the two-class classification problem) with excellent accuracy but show poor performance for the detection of three-class epilepsy classification. Deep learning is a new research direction of machine learning that automatically learns the inherent laws and features of sample data. As both the available data and computational ability of hardware continue to increase, deep learning has addressed increasingly complex applications with ever-increasing accuracy [13–15]. Recently, automatic detection of epileptic seizures based on deep learning methods received much attention.

In 2018, Acharya et al. [16] implemented a 13-layer deep convolutional neural network (CNN) algorithm to detect normal, preictal, and seizure classes. This model includes five convolutional (Conv) layers, five max-pooling layers, and three fully connected (FC) layers. On this three-class detection problem, it achieved accuracy, specificity, and sensitivity of 88.67%, 90.00%, and 95.00%, respectively. Moreover, Ullah et al. [2] proposed an automatic system for epilepsy detection based on an ensemble of pyramidal one-dimensional convolutional neural network models. The core component of the system is a pyramidal one-dimensional convolutional neural network (P-1D-CNN) model, which consists of three main types of layers: Conv, batch normalization (BN), and FC layers. The classification performance of the P-1D-CNN model is not very satisfactory. Hence, the authors introduced the majority-vote (M-V) module in the final stage of the P-1D-CNN model, which significantly improved the performance of the algorithm. In almost all the cases of two-class and three-class concerning epilepsy detection problems, it has given the accuracy of  $99.1 \pm 0.9\%$ . In 2019, Turk and Ozerdem [17] obtained two-dimensional frequency-time scalograms by applying Continuous Wavelet Transform (CWT) to EEG records containing five different classes and used the CNN structure to learn the properties of the scalogram images. On all the two-class, three-class, and five-class classification problems involving seizures, its recognition accuracy is 98.5%~99.5%, 97.0%~99.0%, and 93.6%, respectively. Moreover, Hussein et al. [18] introduced a deep long short-term memory (LSTM) network to learn the high-level representations of different EEG patterns, using one FC layer to extract the most robust EEG features relevant to epileptic seizures. This

model achieved 100% accuracy of the two-class, three-class, and five-class classification problems.

Despite the encouraging seizure detection results gained using the CNN models mentioned above, several improvements can still be achieved. First, some of these CNN models have relatively single model structures. The second issue is the small number of available samples, which is not enough to train a deep neural network model. As such, we felt motivated to develop a CNN model for detecting seizures efficiently with raw EEG signals. To address these issues, first, we add the BN layer and dropout layer into the traditional convolutional blocks for learning features, which may help in detecting seizures efficiently. Second, the segments of raw EEG were divided into many nonoverlapping chunks to increase the number of samples for training and test, which may help in using a small amount of available data for fully training a deep model. Research findings have shown that the proposed approach is advantageous in detecting seizures using EEG signals.

## 2. Materials and Methods

**2.1. Description of EEG Dataset.** Our seizure recognition experiments are conducted using the widely used and publicly available EEG database produced by Bonn University [19]. This database consists of five diverse subsets (set A–E) denoted as Z, O, N, F, and S. Sets A and B are composed of surface EEG recordings of healthy volunteers in the wakeful state with eyes open and eyes closed, respectively. On the other hand, Sets C, D, and E are gathered from patients with epilepsy. Thereinto, Sets C and D were recorded during seizure-free intervals. Set C was recorded from the hippocampal formation of the opposite hemisphere of the brain. Set D was recorded from within the epileptogenic zone. Set E only included seizure activities. Each of these sets contains 100 single-channel recordings of EEG signals with a sampling rate of 173.61 Hz and a duration of 23.6 s. The corresponding time-series is sampled into 4097 data points. Besides, the Rochester Institute of Technology divided every 4097 data points into 23 chunks. Each chunk contains 178 data points for 1 second (<https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition>). To increase the number of samples for training a deep model, the Bonn dataset in this format is adopted, whose amount of sample increases 22 times. Therefore, the number of each category has 2300 EEG samples. Sample EEG signals of five EEG classes are shown in Figure 1.

**2.2. Architecture of the Proposed Network.** The deep CNN model [20] can automatically learn the features of EEG signals and performs classification in an end-to-end manner. The overall CNN architecture proposed in this paper is shown in Figure 2, which can perform feature extraction and classification. First, the input one-dimensional raw EEG data are normalized to zero mean and unit variance. Then, three convolutional blocks are adopted to learn features of the EEG signals, where each block consists of five layers. In detail, the first layer

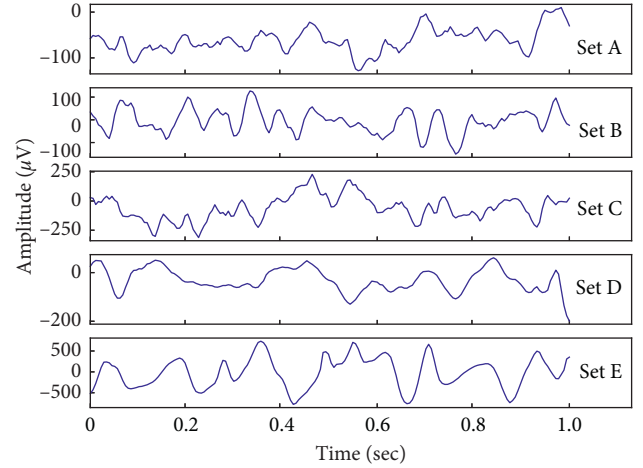


FIGURE 1: Sample EEG signals in this study.

computes multiple convolutions in parallel to generate a set of linear activation responses. The second layer is BN, which is used to solve the internal variable shift. Each linear activation response passes a nonlinear activation function in the layer. The activation function used in this work is the rectified linear unit (ReLU) [21]. In the fourth layer, the dropout technology [22] is employed to prevent overfitting. The last layer of the block is the max-pooling layer, which introduces translation invariance. In the structure, the second and third convolutional blocks are similar to the first.

At the end of the third convolutional block, the feature maps are flattened into a one-dimensional vector that is connected to the FC layer for integrating features. The first two FC layers employ ReLU as the activation function, followed by a dropout layer. The third FC layer applies softmax as the activation function which will output a vector of probabilities corresponding to each category. To choose better model parameters, we explored eight models with different specifications. Details are described in the Experimental Results and Discussion section. In this study, we select model M7. Table 1 shows the details of the proposed CNN structure.

**2.3. Convolution Operation.** A convolutional neural network (CNN) is a neural network designed to process data with similar network structures. The image can be regarded as a two-dimensional pixel grid. Similarly, time-series data can be considered as a one-dimensional grid formed by regularly sampling on time axis. The convolutional block of conventional CNN includes three layers: convolution, activation function, and pooling. For the one-dimensional EEG data used in this paper, the convolution operation is as follows:

$$s(t) = (x * w)(t) = \sum_a x(a)w(t - a). \quad (1)$$

Convolution network has the characteristics of sparse interaction. So, it means fewer parameters need to be stored, which not only reduces the storage requirements of the model but also simplifies the calculation. At the same time,

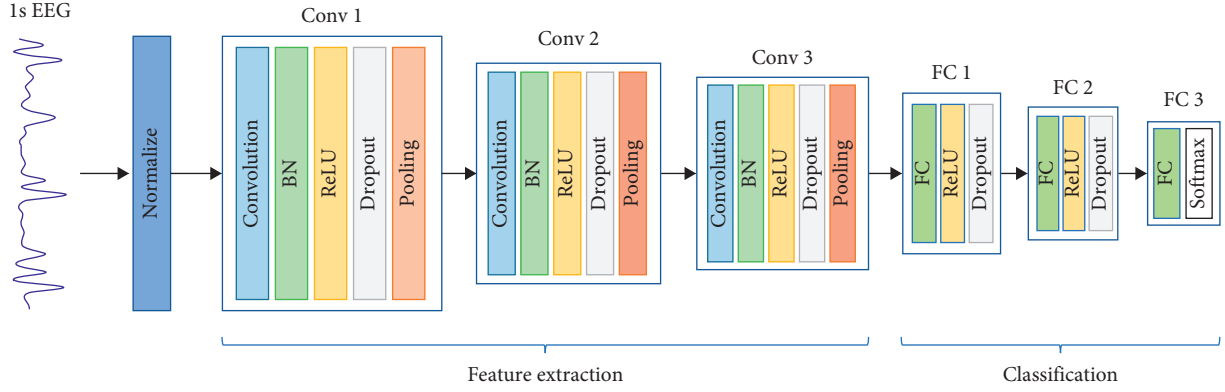


FIGURE 2: The proposed one-dimensional convolutional neural network architecture.

TABLE 1: Details of the CNN structure used in this research.

Block	Type	Number of neurons (Output layer)	Kernel size for each Output feature map	Stride
Conv 1	Convolution	$139 \times 20$	40	1
	BN	$139 \times 20$	—	—
	ReLU	$139 \times 20$	—	—
	Dropout	$139 \times 20$	—	—
	Max-pooling	$70 \times 20$	2	2
Conv 2	Convolution	$51 \times 40$	20	1
	BN	$51 \times 40$	—	—
	ReLU	$51 \times 40$	—	—
	Dropout	$51 \times 40$	—	—
	Max-pooling	$26 \times 40$	2	2
Conv 3	Convolution	$17 \times 80$	10	1
	BN	$17 \times 80$	—	—
	ReLU	$17 \times 80$	—	—
	Dropout	$17 \times 80$	—	—
	Max-pooling	$9 \times 80$	2	2
FC 1	FC	64	—	—
	ReLU	64	—	—
	Dropout	64	—	—
FC 2	FC	32	—	—
	ReLU	32	—	—
	Dropout	32	—	—
FC 3	FC	2 or 3 or 5	—	—

the parameters shared by the convolution kernel ensure that we only need to learn parameters that are many orders of magnitude smaller. Convolution is a kind of special linear operation, and activation function brings nonlinear characteristics into the network. The Rectified Linear Unit (ReLU) function is the most commonly used activation function in CNN, which overcomes the vanishing gradient problem, allowing models to learn faster and perform better. Equation (2) shows the ReLU function:

$$f(x) = \max\{0, x\}. \quad (2)$$

The pooling function can reduce the spatial size of the representation to reduce the number of parameters and

computation in the network. It replaces the output of the system at a specific position. For example, max-pooling gives the maximum value in several neighborhoods. The pooling can also help to make the representation approximately invariant to small translations of the input.

**2.4. Calculation of BN.** In this study, the BN layer and dropout layer are added to the traditional convolutional blocks. When training the deep neural network, the parameters of each layer are closely related to each other. The inconsistency in the distribution of layers' inputs causes a problem, called internal covariate shift. And the internal vary shift makes it difficult for us to choose an appropriate learning rate. To tackle this problem, Ioffe and Szegedy [23] developed BN technology which can almost reparametrize any deep networks, significantly reducing the problem of coordinated updates between multiple layers. The technology takes normalization as part of the model architecture and normalizes each mini-batch.

During training, BN calculates the sample mean and standard deviation for the mini-batch response  $H$  in backpropagation by

$$\begin{aligned} \mu &= \frac{1}{m} \sum_i H_i, \\ \sigma &= \sqrt{\delta + \frac{1}{m} \sum_i (H_i - \mu)^2}, \end{aligned} \quad (3)$$

where the delta component  $\delta$  is kept at a small positive value and is added only to avoid the gradient becoming undefined where the true standard deviation is zero. And they are used to normalize  $H$  by

$$H' = \frac{H - \mu}{\sigma}. \quad (4)$$

BN is also very useful in accelerating the convergence of the training phase and prevents overfitting. The technology has become a common practice, and the detail can be found in [23]. Therefore, we employ BN after every convolutional layer.

**2.5. Feature Fusion and Classification.** A deep neural network needs to learn a large number of parameters, which is likely to cause overfitting in the case of a small dataset. To address this issue, the authors [22] developed dropout technology to prevent the coadaptation of feature detectors. The critical idea of dropout is to randomly drop units with a predefined probability (along with their connections) from the neural network during training. It significantly reduces overfitting and gives significant improvements over other regularization methods. In the proposed model, we add the dropout lay after each ReLu activation function.

The output of the last convolutional block represents high-level features in the EEG signals. The fully connected layer is a usual manner of learning nonlinear combinations of these features. All the neurons in the last max-pooling layer are connected with all the neurons of the first FC layer. We used three FC layers. The number of neurons in the final FC layer (FC3) relies on the detection problem, e.g., for the two-class, three-class, and five-class epileptic classification problem, the number of neurons in FC3 is 2, 3, and 5, respectively.

The softmax activation function is a generalization of the binary form of logistic regression. It is commonly applied to the last layer of a deep neural network for constituting a categorical distribution over class labels and obtaining the probabilities of each input element belonging to a label. The softmax function, denoted by  $h_\theta(x^{(i)})$ , is defined as equation (5), which represent the respective probabilities of the  $i$ -th sample (denoted by  $x^{(i)}$ ) belonging to each category:

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix}, \quad (5)$$

where  $\theta_1, \theta_2, \dots, \theta_k$  are the softmax model parameters.

**2.6. Training of CNN Model.** Training the proposed model needs the weight parameters to be learned from the EEG data. For learning these parameters, we employed the conventional backpropagation algorithm with cross-entropy as the loss function. And, we used the stochastic gradient descent method with Adam optimizer that is based on the adaptive estimation of first-order and second-order moments. The hyperparameters of Adam algorithm are as follows: learning rate (0.0005), beta1(0.9), and beta2(0.999). The model was implemented in Keras, a powerful deep learning library, which runs on top of TensorFlow. The batch size of 100 is chosen in this work, which is used for each training update. To compare the performance measure, we trained all the models that are present in this work with 300 epochs.

**2.7. Performance Measures.** For evaluation, we adopted well-known performance metrics, such as accuracy (Acc),

precision (Pre), sensitivity (Sen), and specificity (Spe), F1. Thereinto, accuracy is one of the most commonly used metrics in the literature, and it is defined as a ratio between the correctly classified samples to the total number of samples. The definitions of these performance metrics are as follows:

$$\begin{aligned} \text{Acc} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Pre} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{Sen} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Spe} &= \frac{\text{TN}}{\text{FP} + \text{TN}}, \\ \text{F1} &= \frac{2 \times \text{Pre} \times \text{Sen}}{\text{Pre} + \text{Sen}}, \end{aligned} \quad (6)$$

where TP (true positive) is the number of abnormal EEG records, which are correctly identified as abnormal; TN (true negative) is the number of normal EEG cases that are correctly predicted as normal; FP (false positive) is the number of normal EEG cases that are predicted as abnormal; and FN (false negative) is the number of abnormal EEG records that are incorrectly classified as normal.

To reduce the statistical uncertainty of test error estimation caused by small-scale test datasets, we adopted 10-fold cross-validation for evaluation. The 2300 EEG signals of each category are randomly divided into ten nonoverlapping fold. During the  $i$ -th test, the  $i$ -th fold of the EEG signals is used for testing while the remaining 9 folds are used for training. The accuracy, sensitivity, and specificity values reported in the paper are the average values obtained from ten evaluations.

### 3. Experimental Results and Discussion

Datasets are grouped with different combinations for exploring a general classification model, which is classified into two classes (nonseizures and seizures), three categories (normal, interictal, and ictal), and five classes (A, B, C, D, and E). To choose better model parameters, we considered eight models with different configurations.

**3.1. Selection of Model.** We explored models with different parameters, including the size of the receptive field, the number of neurons, and the dropout probability of the FC layer, for comparison. Taking the five-class classification problem, for example, the experimental results using 10-fold cross-validation are shown in Table 2.

Experiments show that within experimental parameters, a larger size of the receptive field and more neurons in the FC layer make the recognition more effective. The dropout probability of 20% in the FC layers is more effective than a rate of 50%. Therefore, the parameters of the model M7 with



TABLE 2: Configurations of 8 models using 10-fold cross-validation for the A vs. B vs. C vs. D vs. E cases.

Block	Parameter	M1	M2	M3	M4	M5	M6	M7	M8
Conv 1	Number of kernels	20	20	20	20	20	20	20	20
	Size of receptive field	5	5	5	5	40	40	40	40
	Dropout rate	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Conv 2	Number of kernels	40	40	40	40	40	40	40	40
	Size of receptive field	3	3	3	3	20	20	20	20
	Dropout rate	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Conv 3	Number of kernels	80	80	80	80	80	80	80	80
	Size of receptive field	3	3	3	3	10	10	10	10
	Dropout rate	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
FC 1	Number of neurons	32	32	64	64	32	32	64	64
	Dropout rate	0.2	0.5	0.2	0.5	0.2	0.5	0.2	0.5
FC 2	Number of neurons	16	16	32	32	16	16	32	32
	Dropout rate	0.2	0.5	0.2	0.5	0.2	0.5	0.2	0.5
A vs. B vs. C vs. D vs. E	Acc	90.47	88.61	91.89	91.20	93.37	91.62	<b>93.55</b>	92.92
	Sen	75.98	70.86	79.66	77.79	83.33	78.85	83.73	82.30
	Spe	94.00	92.72	94.92	94.45	95.83	94.71	95.93	95.57

the best performance are used for experiments of two-class and three-class classifications with various combinations.

**3.2. Performance of the Proposed Model.** A multiple classification problem can be decomposed into multiple binary classification problems. The result of each classification can be listed as a confusion matrix, which reflects the original and predicted labels of each category. Table 3 shows the confusion matrix and evaluation metrics of classification normal (B) vs. preictal (D) vs. seizure (E), as well as the overall classification result. All the metrics are over 96%, especially the specificity, which is above 98% in each category, and the overall classification.

To check the robustness of the proposed model, we tested 20 combinations. The detail of 10-fold cross-validation results is shown in Table 4, in which the average accuracy is employed as overall accuracy. The accuracy of the two-class classification varies from 97.63% to 99.52%, which has the best performance for A vs. E and the worst performance for D vs. E. The accuracy of the three-class recognition problem is between 96.73%~98.06%. Notably, the accuracy is as high as 98.06% for B vs. D vs. E. The five-class classification problem is more complicated and harder to solve than the two-class and three-class problems but has an advantage in numerous clinical applications, and the proposed model still obtains an overall accuracy of 93.55%. The proposed model is suitable for various classification problems of the Bonn dataset and has a strong generalization ability.

**3.3. Comparisons with Previous Studies.** Numerous approaches have been presented in the literature for automated detection of epileptic seizures using the Bonn EEG database. Table 5 shows the results of the comparison of the recognition rate of this work with them on various classification problems. The binary classification problem is the problem of identifying nonseizures and seizures. Classification of healthy volunteers and seizures is A vs. E, B vs. E, and AB vs.

E. Due to the significant differences in this classification, the classification results of the various methods that appear in Table 5 are generally outstanding, all above 99%. The classification accuracy of interictal and ictal (C vs. E, D vs. E, and CD vs. E) is slightly lower than the first binary classification. In particular, both sets D and E are from the epileptogenic zone; therefore, it is difficult to distinguish. In the conventional methods of Table 5, Zhang et al. [1] obtained the best performance, which achieved 98.1% accuracy. In CNN-based technology, Ullah et al. [2] employed CNN and the majority-vote module to classify and gain 99.4% accuracy. Turk and Ozerdem [17] used CWT and CNN to recognize and achieved 98.50% accuracy. The proposed model of this work just employed CNN and obtained 97.63% accuracy.

The three-class classification problem further subdivides the EEG records to distinguish normal, interictal, and ictal EEG. We compared two types of three-class problem (B vs. D vs. E and AB vs. CD vs. E). The proposed model also achieved good performance. Especially in the case of B vs. D vs. E, its performance reaches the best accuracy of 98.06%, which is obviously better than another model [16] based on CNN only.

The five-class classification problem is more complicated and harder to classify than the two-class and three-class problems. It needs to identify the differentiation between EEG epochs belonging to the same class (e.g., sets A and B, which are both normal; sets C and D, which are both interictal). Therefore, in the literature, relatively some methods were proposed to address these three types of problems at the same time. The proposed CNN model achieved an accuracy of 93.55%, which is very close to the results of Turk and Ozerdem [17] and better than the conventional methods.

The experiment still needs to be implemented in reducing the learning rate and increasing the number of epochs, which will undoubtedly increase the accuracy of epilepsy recognition but, at the same time, will cost more time for training. For a limited number of training samples,



TABLE 3: Confusion matrix for the three-class problem (B vs. D vs. E.) across 10-folds.

		Predicted			Acc	Sen	Spe	Pre	F1
		Normal	Preictal	Seizure					
Original	Normal	2263	36	1	98.32	98.39	98.28	96.63	97.50
	Preictal	49	2220	31	97.54	96.52	98.04	96.10	96.31
	Seizure	30	54	2216	98.32	96.35	99.30	98.58	97.45
Overall		—	—	—	98.06	97.09	98.54	97.10	97.09

TABLE 4: Accuracies (%) of 10-fold cross-validation using model M7.

Data sets combination	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Mean
A vs. E	100	99.57	99.57	99.35	99.35	99.57	99.13	99.57	99.35	99.78	99.52
B vs. E	99.78	99.13	99.57	98.91	99.13	99.35	98.70	98.70	98.70	99.13	99.11
C vs. E	99.35	98.04	98.04	96.96	98.26	97.39	97.39	97.83	98.48	98.48	98.02
D vs. E	97.61	98.04	98.26	98.04	97.17	98.04	96.52	97.17	96.52	98.91	97.63
AB vs. E	99.57	99.13	99.57	99.57	99.57	99.13	99.57	99.13	99.57	98.99	99.38
AC vs. E	99.28	98.70	99.13	98.84	99.13	98.70	98.99	99.57	99.42	98.55	99.03
AD vs. E	98.12	97.83	98.41	98.70	98.41	98.41	98.55	98.55	99.13	98.84	98.50
BC vs. E	98.70	98.41	97.68	98.55	98.55	98.99	98.84	99.28	99.57	98.26	98.68
BD vs. E	97.39	97.10	97.54	98.84	98.26	97.54	98.41	97.97	97.83	97.39	97.83
CD vs. E	97.68	97.54	98.41	97.83	98.41	97.25	98.84	98.41	97.97	97.97	98.03
ABC vs. E	99.24	98.26	99.24	98.91	98.80	99.02	98.91	99.24	98.91	98.37	98.89
ABD vs. E	98.80	98.37	98.80	98.26	98.80	99.35	98.48	97.93	98.15	98.26	98.52
BCD vs. E	98.26	97.61	98.59	98.26	98.59	99.24	98.04	98.70	97.93	98.37	98.36
ABCD vs. E	98.96	99.22	98.70	98.52	98.35	99.22	98.78	98.61	99.13	98.09	98.76
A vs. C vs. E	96.04	97.05	97.00	97.39	94.98	97.58	97.00	96.09	96.81	97.39	96.73
A vs. D vs. E	97.63	97.10	97.54	95.94	97.00	96.67	97.39	97.87	96.81	96.43	97.04
B vs. C vs. E	97.63	97.97	98.12	97.68	98.36	97.20	97.87	99.03	97.68	97.58	97.91
B vs. D vs. E	98.35	98.30	98.07	97.49	98.26	97.97	97.20	98.45	98.45	98.06	98.06
AB vs. CD vs. E	96.70	97.10	97.74	96.43	96.72	97.97	94.96	97.91	96.96	97.25	96.97
A vs. B vs. C vs. D vs. E	92.99	94.37	94.00	93.41	93.36	92.73	93.74	93.25	93.74	93.91	93.55

TABLE 5: Comparison between the proposed method and other methods using the same dataset.

Data sets combination	Methodology	Study	Acc (%)	Our Acc (%)
A vs. E	TFA + ANN	Tzallas et al. [4]	100	99.52
	DWT + Kmeans + MLPNN	Orhan et al. [6]	100	
	1-D-LBP + FT/BN	Kaya et al. [8]	99.50	
	DWT + NB/KNN	Sharmila and Geethanjali [10]	100	
	TQWT + KNNE + SVM	Bhattacharyya et al. [11]	100	
	LMD + GA-SVM	Zhang and Chen [1]	100	
	CNN + M-V	Ullah et al. [2]	100	
	CWT + CNN	Turk and Ozerdem [17]	99.50	
B vs. E	DWT + NB/KNN	Sharmila and Geethanjali [10]	99.25	99.11
	TQWT + KNNE + SVM	Bhattacharyya et al. [11]	100	
	CNN + M-V	Ullah et al. [2]	99.6	
	CWT + CNN	Turk and Ozerdem [17]	99.50	
C vs. E	DWT + NB/KNN	Sharmila and Geethanjali [10]	99.62	98.02
	TQWT + KNNE + SVM	Bhattacharyya et al. [11]	99.50	
	CNN + M-V	Ullah et al. [2]	99.1	
	CWT + CNN	Turk and Ozerdem [17]	98.50	
D vs. E	1-D-LBP + FT/BN	Kaya et al. [8]	95.50	97.63
	DWT + NB/KNN	Sharmila and Geethanjali [10]	95.62	
	TQWT + KNNE + SVM	Bhattacharyya et al. [11]	98	
	LMD + GA-SVM	Zhang and Chen [1]	98.10	
	CNN + M-V	Ullah et al. [2]	99.4	
	CWT + CNN	Turk and Ozerdem [17]	98.50	

TABLE 5: Continued.

Data sets combination	Methodology	Study	Acc (%)	Our Acc (%)
AB vs. E	DWT + NB/KNN	Sharmila and Geethanjali [10]	99.16	99.38
	CNN + M-V	Ullah et al. [2]	99.8	
CD vs. E	1-D-LBP + FT/BN	Kaya et al. [8]	97.00	98.03
	DWT + NB/KNN	Sharmila and Geethanjali [10]	98.75	
	CNN + M-V	Ullah et al. [2]	99.7	
ABCD vs. E	DWT + Kmeans + MLPNN	Orhan et al. [6]	99.60	98.76
	DWT + NB/KNN	Sharmila and Geethanjali [10]	97.1	
	TQWT + KNNE + SVM	Bhattacharyya et al. [11]	99	
	LMD + GA-SVM	Zhang and Chen [1]	98.87	
	CNN + M-V	Ullah et al. [2]	99.7	
B vs. D vs. E	CNN	Acharya et al. [16]	88.7	98.06
	CWT + CNN	Turk and Ozerdem [17]	98.00	
AB vs. CD vs. E	DWT + Kmeans + MLPNN	Orhan et al. [6]	95.60	96.97
	TQWT + KNNE + SVM	Bhattacharyya et al. [11]	98.60	
	LMD + GA-SVM	Zhang and Chen [1]	98.40	
	CNN + M-V	Ullah et al. [2]	99.1	
A vs. B vs. C vs. D vs. E	TFA + ANN	Tzallas et al. [4]	89	93.55
	MEMD + ANN	Zahra et al. [12]	87.2	
	CWT + CNN	Turk and Ozerdem [17]	93.60	

we can also try to enhance the dataset, which may be useful for the generalization ability of the model. For example, we can divide the 23.6 seconds of EEG data into many overlapping chunks to further increase the number of samples.

#### 4. Conclusion

A novel model for robust detection of seizures has been proposed, which deals with two-class, three-class, and five-class classification problems. The proposed approach has been developed based on the one-dimensional convolutional neural network model, which takes the raw EEG signal as input. To improve the learning ability of the model, the BN and dropout layers have been introduced to the traditional convolutional block. To address the issue of the small datasets, the EEG has been divided into many nonoverlapping chunks for training and test. The experimental result shows that the proposed model performs well on various EEG classification problems on the Bonn dataset.

#### 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 Education and Scientific Research Project for Young and Middle-aged Teachers of Fujian Province (JAT191153 and JT180872), the National Natural Science Foundation of Shanghai, China

(19ZR1455200), and the National Natural Science Foundation of China (11601339).

#### References

- [1] T. Zhang and W. Chen, "LMD based features for the automatic seizure detection of EEG signals using SVM," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1100–1108, 2017.
- [2] I. Ullah, M. Hussain, E.-U.-H. Qazi, and H. Aboalsamh, "An automated system for epilepsy detection using EEG brain signals based on deep learning approach," *Expert Systems with Applications*, vol. 107, pp. 61–71, 2018.
- [3] H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2027–2036, 2009.
- [4] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in EEGs using time–frequency analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.
- [5] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [6] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13475–13481, 2011.
- [7] U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic EEG using entropies," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 401–408, 2012.
- [8] Y. Kaya, M. Uyar, R. Tekin, and S. Yildirim, "1D-local binary pattern based feature extraction for classification of epileptic EEG signals," *Applied Mathematics and Computation*, vol. 243, pp. 209–219, 2014.
- [9] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of

- intrinsic mode functions,” *Expert Systems with Applications*, vol. 42, no. 3, pp. 1106–1117, 2015.
- [10] A. Sharmila and P. Geethanjali, “DWT based detection of epileptic seizure from EEG signals using naive bayes and k-NN classifiers,” *IEEE Access*, vol. 4, pp. 7716–7727, 2016.
  - [11] A. Bhattacharyya, R. Pachori, A. Upadhyay et al., “Tunable-Q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals,” *Applied Sciences*, vol. 7, no. 4, Article ID 385, 2017.
  - [12] A. Zahra, N. Kanwal, N. ur Rehman, S. Ehsan, and K. D. McDonald-Maier, “Seizure detection from EEG signals using multivariate empirical mode decomposition,” *Computers in Biology and Medicine*, vol. 88, pp. 132–141, 2017.
  - [13] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
  - [14] C. Huang, Y. Lan, G. Xu et al., “A deep segmentation network of multi-scale feature fusion based on attention mechanism for IVOCT lumen contour,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.
  - [15] M. Li, S. Dong, Z. Gao et al., “Unified model for interpreting multi-view echocardiographic sequences without temporal information,” *Applied Soft Computing*, vol. 88, Article ID 106049, 2020.
  - [16] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, “Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals,” *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.
  - [17] O. Turk and M. S. Ozerdem, “Epilepsy detection by using scalogram based convolutional neural network from EEG signals,” *Brain Sciences*, vol. 9, no. 5, Article ID 115, 2019.
  - [18] R. Hussein, H. Palangi, R. K. Ward, and Z. J. Wang, “Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals,” *Clinical Neurophysiology*, vol. 130, no. 1, pp. 25–37, 2019.
  - [19] R. G. Andrzejak, K. Lehnertz, F. Mormann et al., “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state,” *Physical Review E*, vol. 64, no. 6, Article ID 061907, 2001.
  - [20] Y. LeCun, “Generalization and network design strategies,” in *Connectionism in Perspective*, pp. 143–155, Elsevier, Zurich, Switzerland, 1989.
  - [21] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 807–814, Haifa, Israel, June 2010.
  - [22] N. Srivastava, G. Hinton, A. Krizhevsky et al., “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
  - [23] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, pp. 448–456, Lille, France, July 2015.

## Research Article

# Magnetic Resonance Image Denoising Algorithm Based on Cartoon, Texture, and Residual Parts

Yanqiu Zeng<sup>1</sup>, Baocan Zhang<sup>1</sup>, Wei Zhao<sup>1</sup>, Shixiao Xiao<sup>1</sup>, Guokai Zhang<sup>2</sup>,  
Haiping Ren<sup>3</sup>, Wenbing Zhao<sup>4</sup>, Yonghong Peng<sup>5</sup>, Yutian Xiao<sup>6</sup>, Yiwen Lu,<sup>7</sup>  
Yongshuo Zong,<sup>7</sup> and Yimin Ding<sup>8</sup>

<sup>1</sup>Chengyi University College, Jimei University, Xiamen, China

<sup>2</sup>School of Software Engineering, Tongji University, Shanghai, China

<sup>3</sup>Jiangxi University of Science and Technology, Nanchang, China

<sup>4</sup>Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, OH 44115, USA

<sup>5</sup>Faculty of Computer Science, University of Sunderland, Sunderland, UK

<sup>6</sup>School of Informatics, Xiamen University, Xiamen, China

<sup>7</sup>Department of Computer Science, Tongji University, Shanghai, China

<sup>8</sup>College of Electronics and Information Engineering, Tongji University, Shanghai, China

Correspondence should be addressed to Guokai Zhang; zhangguokai\_01@163.com and Haiping Ren; chinahrp@163.com

Received 8 February 2020; Accepted 6 March 2020; Published 1 April 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Yanqiu Zeng 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.

Magnetic resonance (MR) images are often contaminated by Gaussian noise, an electronic noise caused by the random thermal motion of electronic components, which reduces the quality and reliability of the images. This paper puts forward a hybrid denoising algorithm for MR images based on two sparsely represented morphological components and one residual part. To begin with, decompose a noisy MR image into the cartoon, texture, and residual parts by MCA, and then each part is denoised by using Wiener filter, wavelet hard threshold, and wavelet soft threshold, respectively. Finally, stack up all the denoised subimages to obtain the denoised MR image. The experimental results show that the proposed method has significantly better performance in terms of mean square error and peak signal-to-noise ratio than each method alone.

## 1. Introduction

Magnetic resonance imaging (MRI) is one of the advanced imageological examination methods for modern medicine. MRI uses powerful magnets and computer-generated radio waves instead of injected contrast agents to create multidimensional images of human organs and tissues. It does not damage the body with ionizing radiation, so it is safer than emission computed tomography (ECT). For this reason, MRI is frequently used for imaging tests of the brain and spinal cord. However, compared with computed tomography (CT), MR image has a lower spatial resolution, longer scan time, and more artifacts. The longer the scanning time, the greater the thermal noise (a kind of Gaussian noise). Besides, medical images are always polluted by various noises during

collection, transmission, and storage. The magnitude of MRI data in the presence of noise generally follows a Rician distribution if acquired with single-coil systems [1]. Also, the Gaussian distribution can approximate Rician noise in high SNR (signal-to-noise ratio) regions [2]. Quite often, noise affecting the pixels in an image is Gaussian in nature and uniformly deters information pixels in the image [3]. MR image denoising, as an essential preprocessing step for MRI data processing, has been a hot topic in the related area.

Many scholars and researchers have performed much work on image denoising. Various image denoising methods can be broadly classified as five categories: spatial domain filtering, transform domain filtering, methods in other domains, sparse representation and dictionary learning methods, and hybrid methods [3].

The spatial domain filtering can be further divided into linear (such as Wiener filters) and nonlinear filters (such as median filters) [4]. Wiener filter, a denoising method used when the noise is a stationary random process, minimizes the mean square error between the output signal and the desired noise-free signal. It has a wide range of applications, regardless of whether the stationary random process is continuous or discrete, scalar or vector. Jingdong Chen et al. studied the quantitative performance behavior of the Wiener filter in the context of noise reduction in 2006 [5]. A new filtering method based on the neutrosophic set approach of the Wiener filter was proposed for MRI denoising in 2013 [6].

Transform domain filtering includes many classic methods, such as Fourier transform, wavelet transform, threshold function, curvelet, and contourlet. Wavelet transform uses a series of wavelets with different scales to decompose the original image to get the coefficients of different wavelets. In general, for a noisy image, the wavelet transform will decompose most of the noises and effective signals into coefficients with small and big moduli, respectively. So, it is easy to remove lots of the noises by removing the low-frequency parts. Wavelet thresholding is a powerful denoising method based on the wavelet transform. Donoho and Johnstone gave two kinds of thresholding functions: hard thresholding and soft thresholding [7, 8]. The deficiency of hard thresholding is its discontinuity, while the drawback of soft thresholding is that it causes constant deviation [9]. Fei Xiao and Yungang Zhang explored the properties of several representative thresholding techniques in wavelets denoising in 2011 [10]. Zhang et al. proposed an improved threshold function to overcome the drawbacks of hard thresholding and soft thresholding functions in 2019 [9]. Wavelet transform has been widely applied in image processing [11, 12].

Hybrid methods are popular since different denoising methods have different advantages. Here, we focus on the application of morphological component analysis (MCA) and wavelet thresholding. MCA is a signal separation method proposed by Starck et al. in 2005 [13], which combines the advantages of sparse representation and variational method. When the features contained in an image present different morphological aspects, we can separate multiple components with different shapes from the image (as shown in Figure 1). The morphological component analysis assumes that each morphological component can find a dictionary for sparse representation, and the dictionaries for different morphological components are independent. The authors in [14] proposed an image denoising method based on morphological component analysis (MCA) to remove the rain component successfully in 2013. Naimi et al. proposed a denoising approach basing on dual-tree complex wavelet and shrinkage with the Wiener filter technique in 2014 [15]. Deng and Liu proposed an improved image denoising method based on MCA and median filter to resist mixed noises in 2015 [16]. Cheng and Liu used APBT and BM3D to denoise texture and structure part, respectively [17]. MCA has been widely used in medical image processing [18].

In this paper, we propose a method to remove Gaussian noise in MR brain images based on image decomposition by MCA. Wiener filter (classical spatial domain filtering), wavelet hard threshold, and wavelet soft threshold (classical transform domain filtering) are employed to remove noise in the cartoon, texture, and residual parts, respectively. The experimental results show that the proposed method achieves better noise reduction, both objectively and subjectively.

## 2. Materials and Methods

In this study, we present a hybrid MR image denoising method based on MCA, Wiener filtering, and wavelet hard and soft thresholds. The whole denoising flow chart is shown in Figure 2. Firstly, add Gaussian white noise to the three original MR brain images. Secondly, use MCA to decompose the noisy images into three parts: cartoon, texture, and residual parts. In general, the residual part is regarded as a “noise component” and discarded directly by traditional MCA. Nevertheless, this study retains the residual part, which contains the outline of the brain. Thirdly, use Wiener filter, wavelet hard threshold, and wavelet soft threshold to remove noise in cartoon, texture, and residual parts, respectively. At last, superimpose the three denoised sub-images to get denoised MR images.

**2.1. Materials.** The original data are a very thorough set of sequences of MR brain imaging of Axial T2, which was obtained from the open-edit educational radiology resource radiopaedia.org [19]. This paper chooses three static images at different moments of the original data for research. They are noise-free and of high resolution.

**2.2. Wiener Filter.** The Wiener filter is an optimal linear filter proposed by Norbert Wiener in 1942. It seeks the linear time-invariant filter whose output comes as close as possible to the original signal. In other words, the goal is to minimize the mean square error (MSE) between the expected noise-free signal and the actual output signal. The Wiener filter assumes that the input is the sum of valuable signals and noise, both of which are generalized stationary processes, and their second-order statistical characteristics are known. Therefore, it is not adaptive and always implemented in the frequency domain.

Formally, let  $f(x, y)$  be the input image and  $g(x, y)$  be the degraded image with some point-spread function  $h(x, y)$  and additive noise  $\eta(x, y)$ . So, in the spatial domain, the blurred image is

$$g(x, y) = H(x, y) * f(x, y) + \eta(x, y), \quad (1)$$

where  $*$  means two-dimensional convolution,  $H(x, y)$  is the blurring function, and additive noise  $\eta(x, y)$  often refers to Gauss white noise, uniform noise, etc. The Wiener filter treats images and noises as random processes, and the objective is to find an estimate  $\hat{f}$  of the original image  $f(x, y)$  such that the MSE is minimum. The optimization problem is as follows:



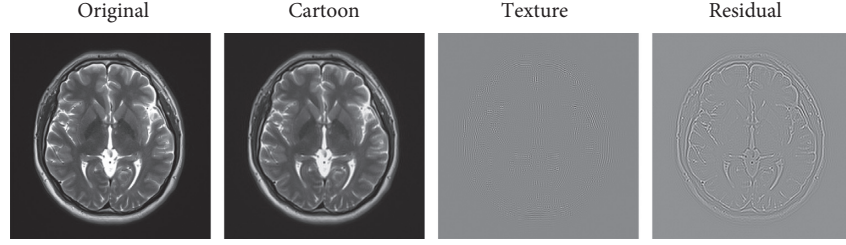


FIGURE 1: An original MRI and its decomposition based on MCA.

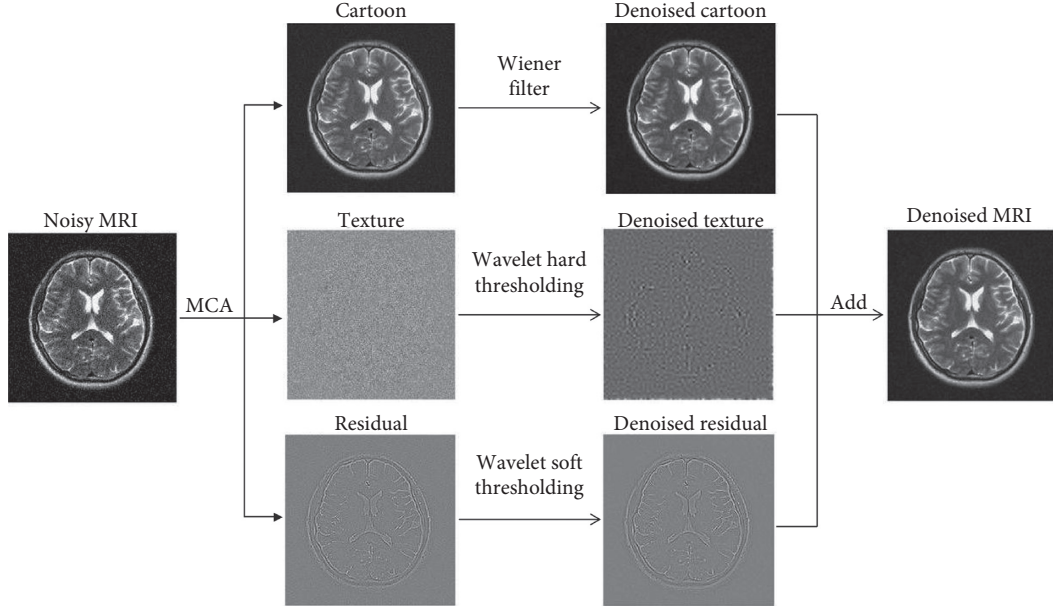


FIGURE 2: The flow chart of denoising an MR image.

$$\min e^2 = E\{(f - \hat{f})^2\}, \quad (2)$$

where  $E$  means mathematical expectation. In the frequency domain, the optimization solution is given by [20]

$$\hat{F}(u, v) = \frac{H^*(u, v)}{(|H(u, v)|^2 + S_\eta(u, v)/S_f(u, v))}, \quad (3)$$

where  $H^*(u, v)$  is the complex conjugate of  $H(u, v)$ ,  $S_\eta(u, v)$  is the power spectrum of the noise, and  $S_f(u, v)$  is the power spectrum of the original image. If  $(S_\eta(u, v)/S_f(u, v))$  is larger, the Wiener filter is smaller, so the frequency will be ignored. For more detailed information on the Wiener filter denoising, please refer to the textbook [21].

**2.3. Wavelet Transformation.** Wavelet denoising is widely used to remove noise from various signals, including one-dimensional signals (such as EEG) and two-dimensional signals (such as MR images). The algorithm is relatively simple to implement and has been proven effective in image denoising [22]. The energy obtained by using the wavelet transform usually concentrates on large coefficients, which correspond to the chief portion of the original signal because

common noises such as Gaussian white noise do not correlate with wavelets. Therefore, wavelet coefficients with large amplitudes are mostly the required signals, while wavelet coefficients with small amplitudes are usually noise. Due to this property, the most common technique for reducing image noise by wavelet is thresholding. In this study, we use hard and soft thresholds.

The hard threshold is defined as follows:

$$\hat{w}_{j,k} = \begin{cases} w_{j,k}, & |w_{j,k}| \geq \lambda, \\ 0, & |w_{j,k}| < \lambda. \end{cases} \quad (4)$$

Moreover, the soft threshold is defined as follows:

$$\hat{w}_{j,k} = \begin{cases} \text{sgn}(w_{j,k}) \left( |w_{j,k}| - \lambda \right), & |w_{j,k}| \geq \lambda, \\ 0, & |w_{j,k}| < \lambda. \end{cases} \quad (5)$$

In both definitions,  $w_{j,k}$  is the wavelet coefficient and  $\lambda$  is the threshold value. Then, we can use  $\hat{w}_{j,k}$  to perform the inverse wavelet transform to obtain the denoised image. Hard thresholding and soft thresholding are widely used and effective. It should be noticed that the hard threshold function is discontinuous when  $w_{j,k} = \pm\lambda$ , which would

cause oscillation in the denoised signal. On the contrary, the soft threshold method is continuous globally, but there is a constant error between  $w_{j,k}$  and  $\hat{w}_{j,k}$  when  $w_{j,k} \geq \lambda$ , which would reduce the accuracy of the approximation. Therefore, the choice of threshold is the primary concern for the wavelet transform denoising and should be validated by experiments. More detailed theories of the wavelet transform with its application could be found in many pieces of the literature, such as in [23].

**2.4. Our Proposed Method Based on MCA.** The main idea of MCA is to decompose an image into different additive layers, and each layer corresponds to a kind of morphological component of the image. Besides, the layer decomposition is required to optimize the sparsity of its representation. The core method of layer decomposition is to use adapted dictionaries, one for texture part representation and the other for cartoon part representation. The dictionaries are mutually unrelated. Each dictionary can only sparsely represent one morphological component and cannot sparsely represent other morphological components. The algorithm has been proven to perform well in many applications.

Formally, let  $s$  be a signal, which could be divided into  $K$  parts. Let

$$s = \sum_{k=1}^K s_k, \quad (6)$$

where each  $s_k$  represents a different type of signal decomposed from the signal  $s$ . For each possible representation  $s_k$ , there must be a dictionary  $\Phi_k \in M^{N \times L_k}$  (where  $L_k \gg N$  normally) such that the optimization problem

$$\bar{\alpha}_k = \arg \min_{\alpha} \|\alpha\|_0, \quad \text{subject to } s_k = \Phi_k \alpha, \quad (7)$$

Has a very sparse solution ( $\|\bar{\alpha}_k\|_0$  is very small). On the contrary, the optimization problem

$$\bar{\alpha}_l = \arg \min_{\alpha} \|\alpha\|_0, \quad \text{subject to } s_l = \Phi_k \alpha \ (k \neq l), \quad (8)$$

Does not have a sparse solution.

For the decomposition of a signal, the MCA requires to optimize the following equation:

$$\{\bar{\alpha}_1, \bar{\alpha}_2, \dots, \bar{\alpha}_K\} \arg \min_{\{\alpha_1, \alpha_2, \dots, \alpha_K\}} \sum_{k=1}^K \|\alpha_k\|_0, \quad \text{subject to } s = \sum_{k=1}^K \Phi_k \alpha_k. \quad (9)$$

For the decomposition of an image, the MCA usually decomposes it into three components: cartoon, texture, and additive noise. Then, we can throw away the noise component and add only cartoon and texture components as the denoised image.

In this study, MCA separated two morphological components with different features and different dictionaries from the MR image, namely, the cartoon and texture parts. Accordingly, the optimization equation is as follows: for an MR image  $s$ ,

$$\{\bar{\alpha}_c, \bar{\alpha}_t\} = \arg \min_{\{\alpha_c, \alpha_t\}} (\|\alpha_c\|_0 + \|\alpha_t\|_0), \quad \text{subject to } s = \Phi_c \alpha_c + \Phi_t \alpha_t, \quad (10)$$

where  $\Phi_c$  and  $\Phi_t$  represent the overcomplete dictionaries of cartoon and texture parts, respectively.  $\alpha_c$  and  $\alpha_t$  are the corresponding sparse coefficients. Because this is an NP-hard problem,  $l_1$ -norm would be used for approximation [24]. On the contrary, a noisy MR image cannot be accurately decomposed into cartoon and texture parts of sparse representations. So, a less strict constraint can be used to approximate the decomposition. The optimization equation used is as follows:

$$\{\bar{\alpha}_c, \bar{\alpha}_t\} \arg \min_{\{\alpha_c, \alpha_t\}} (\|\alpha_c\|_1 + \|\alpha_t\|_1), \quad \text{subject to } \|s - \Phi_c \alpha_c - \Phi_t \alpha_t\|_2 \leq \varepsilon, \quad (11)$$

where  $\varepsilon$  is the value of noise tolerance. The solution of the problem results in that the decomposition would leave out some components (i.e., the residual), which cannot be sparsely represented by both dictionaries. Let  $R$  be the residual part. Then, our decomposition model can be summarized as follows:

$$s = \Phi_c \alpha_c + \Phi_t \alpha_t + R. \quad (12)$$

As shown in Figure 1, the MCA decomposes the original MR image into the cartoon, texture, and residual parts, which represent a meaningful component, an insignificant component, and the residual part, respectively.

After the noise-added MR image is decomposed into three components, each component is denoised by different methods. Specifically, the Wiener filter, wavelet hard threshold, and wavelet soft threshold are used to denoise the cartoon, texture, and residual parts, respectively. Finally, all the denoised subimages are superimposed together as the final denoised MR image. The flow chart of denoising an MR image is shown in Figure 2.

**2.5. Evaluation Methods.** There are two ways to evaluate the performance of different image denoising methods: objective method and subjective method. The subjective evaluation method is to compare the original image and the denoised image visually with naked eyes. The objective evaluation method is an index to quantify denoising performance. Here, we employ two common objective evaluation indexes: mean square error (MSE) and peak signal-to-noise ratio (PSNR).

The mean square error (MSE) is calculated as follows:

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (f_{ij} - \hat{f}_{ij})^2, \quad (13)$$

where  $M$  and  $N$  represent the length and width of the image, respectively,  $f_{ij}$  denotes the pixel value of the original image, and  $\hat{f}_{ij}$  represents the pixel value of the denoised image.

The peak signal-to-noise ratio (PSNR) is computed as follows:

TABLE 1: Denoising results of MRI 1 in different ways.

Results	MSE of MRI 1					PSNR of MRI 1				
Noise variance	0.01	0.03	0.05	0.07	0.09	0.01	0.03	0.05	0.07	0.09
Wiener filter	87	257	422	577	723	28.75	24.03	21.88	20.52	19.54
Hard threshold	252	334	425	523	628	24.04	22.89	21.85	20.95	20.15
Soft threshold	74	231	382	702	1043	29.44	24.49	22.31	19.67	17.94
Proposed	73	174	284	399	522	29.48	25.73	23.6	22.13	20.95

TABLE 2: Denoising results of MRI 2 in different ways.

Results	MSE of MRI 2					PSNR of MRI 2				
Noise variance	0.01	0.03	0.05	0.07	0.09	0.01	0.03	0.05	0.07	0.09
Wiener filter	87	254	417	566	701	28.76	24.09	21.93	20.6	19.67
Hard threshold	233	308	397	493	586	24.46	23.25	22.14	21.2	20.45
Soft threshold	132	319	1682	436	2591	26.91	23.09	15.75	21.73	13.88
Proposed	68	167	272	388	501	29.79	25.92	23.78	22.24	21.13

TABLE 3: Denoising results of MRI 3 in different ways.

Results	MSE of MRI 3					PSNR of MRI 3				
Noise variance	0.01	0.03	0.05	0.07	0.09	0.01	0.03	0.05	0.07	0.09
Wiener filter	85	256	423	583	730	28.85	24.05	21.87	20.48	19.5
Hard threshold	232	311	404	510	618	24.45	23.2	22.07	21.06	20.22
Soft threshold	230	339	393	728	797	24.5	22.81	22.19	19.5	19.11
Proposed	68	167	278	400	530	29.85	25.9	23.7	22.11	20.89

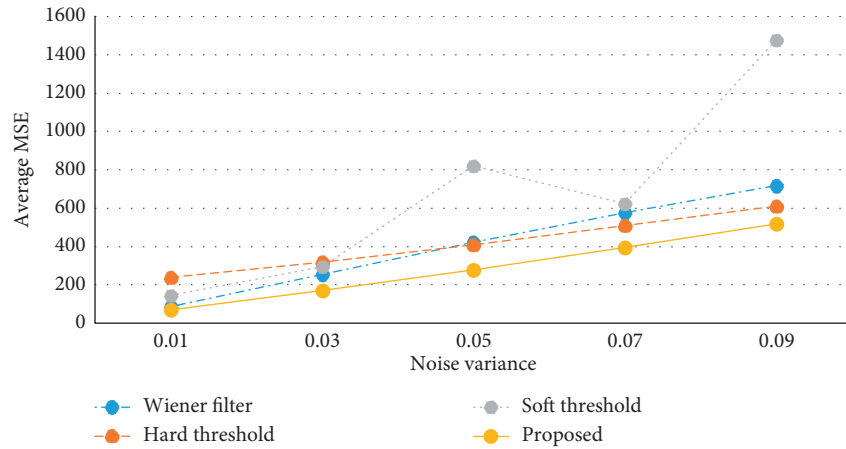


FIGURE 3: Average MSE values of the three MRI images in different ways.

$$\text{PSNR} = 10 * \log \frac{L^2}{\text{MSE}}, \quad (14)$$

where  $L$  represents the maximum grayscale value of the pixels in an image. Here,  $L = 255$ .

### 3. Analysis of Results

In this section, all of the simulation experiments were accomplished in Matlab 2012. The Wiener filter, wavelet hard threshold, wavelet soft threshold, and proposed method (a

combined method of the previous three methods based on MCA) are adopted for experimental comparison.

**3.1. Objective Evaluation.** In this section, the noise reduction effects are achieved when the mean value of Gaussian noise is zero, and the noise variance is 0.01, 0.03, 0.05, 0.07, and 0.09, respectively.

The objective comparison results of the three MRI images are shown in Tables 1–3, respectively. The tabulation results indicate that the MSE values of the proposed method

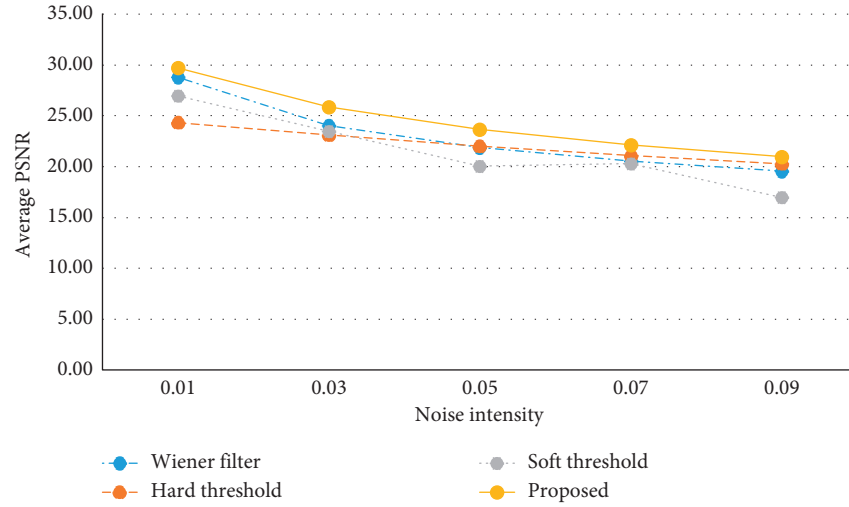


FIGURE 4: Average PSNR values of the three MRI images in different ways.

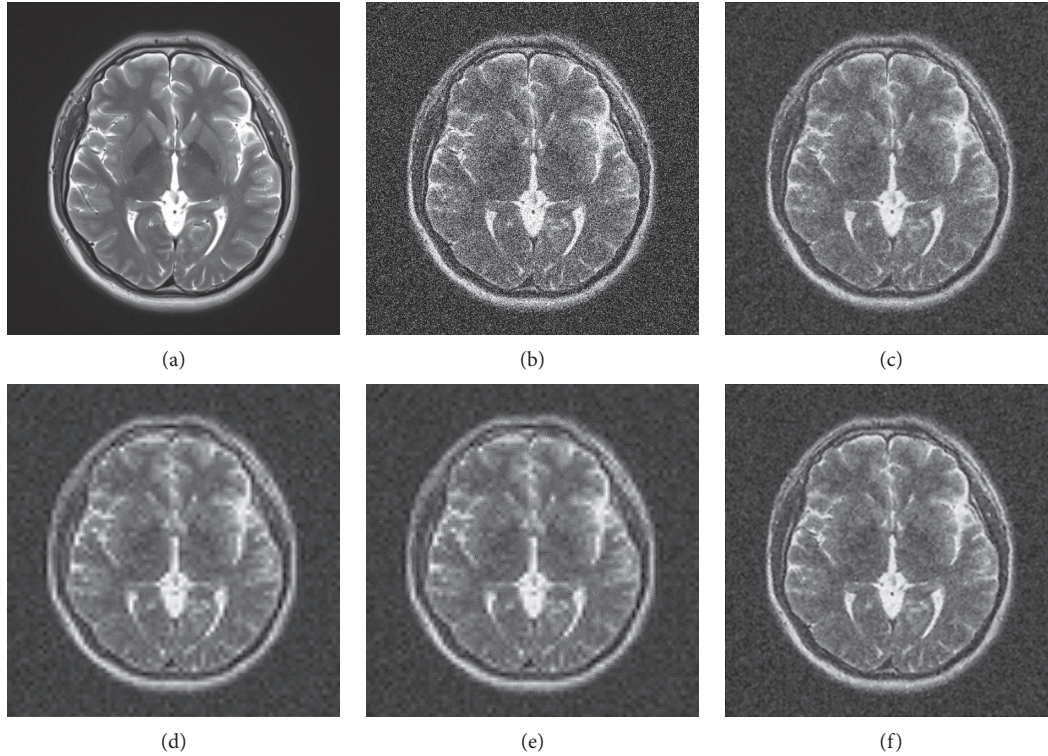


FIGURE 5: Subjective denoising results of MRI 1 in different ways. (a) Original image, (b) noisy image, (c) Wiener filter, (d) hard threshold, (e) soft threshold, and (f) proposed.

are always significantly lower than other methods, while the PSNR values of that are always higher than other methods. More intuitively, the line graphs in Figure 3 and Figure 4 visualize the average MSE values and the average PSNR values of the three MRI images, respectively. It is observed that the average MSE line of the proposed method is lower than other methods, and the average PSNR line of the proposed method is higher than other methods. To sum up, the proposed method achieves better denoising effects in terms of the MSE value and the PSNR value than each method alone.

**3.2. Subjective Evaluation.** In this section, the denoising effects are achieved when the mean value of Gaussian noise is zero and the variance of that is 0.05.

The denoising visual effects of the three MRI images are shown in Figures 5–7. The images denoised by hard and soft thresholds lose more important outline information than the other methods and therefore appear blurry. On the contrary, the images denoised by the Wiener filter retain most of the edge information as well as considerable noises. By contrast, the images denoised by



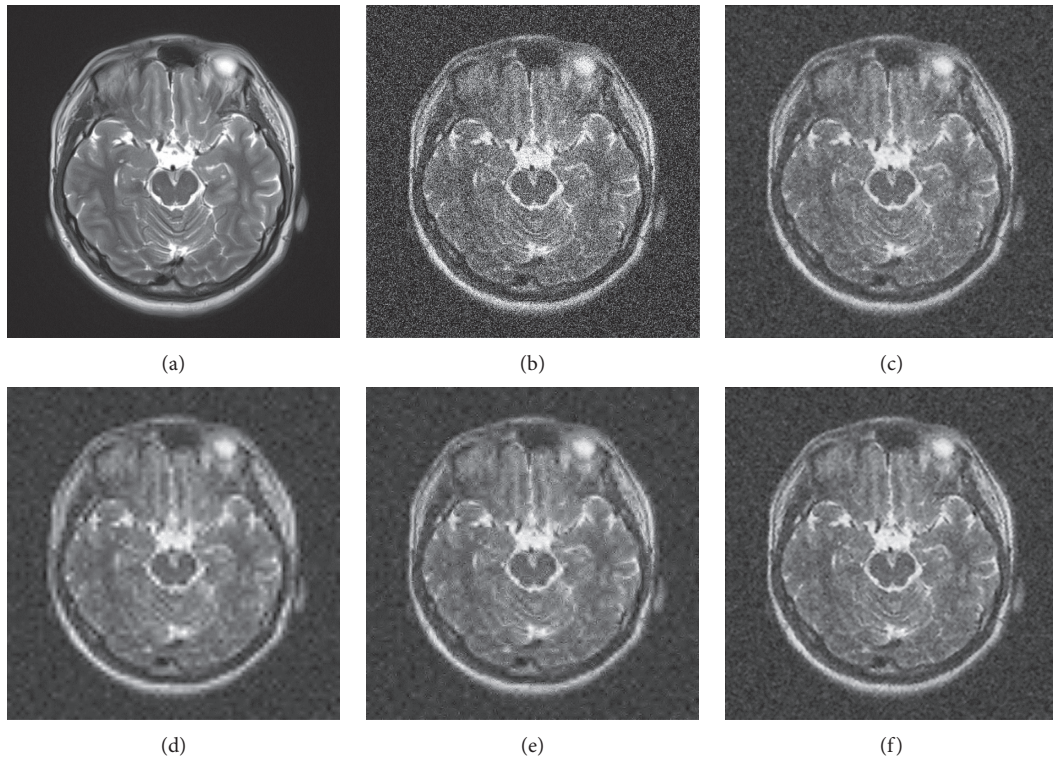


FIGURE 6: Subjective denoising results of MRI 2 in different ways. (a) Original image, (b) noisy image, (c) Wiener filter, (d) hard threshold, (e) soft threshold, and (f) proposed.

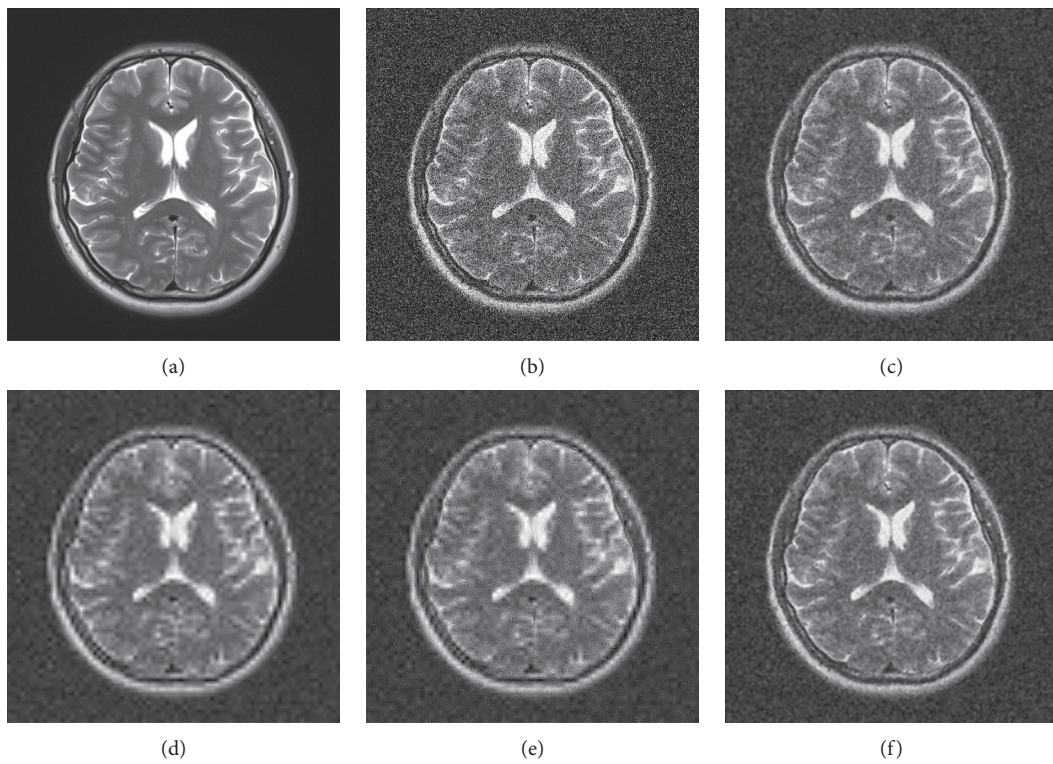


FIGURE 7: Subjective denoising results of MRI 3 in different ways. (a) Original image, (b) noisy image, (c) Wiener filter, (d) hard threshold, (e) soft threshold, and (f) proposed.



TABLE 4: PSNR values of denoising the cartoon, texture, and residual parts in order ( $W$ : Wiener filter,  $H$ : hard threshold, and  $S$ : soft threshold).

Noise variance	0.01	0.03	0.05	0.07	0.09
PSNR of MRI 1					
SHW	29.34	25.5	23.52	21.42	20.99
SWH	28.7	24.07	22	20.06	19.67
HSW	24.7	23.02	19.49	19.65	17.66
HWS	24.34	22.24	20.91	19.83	19.13
WSH	29.77	25.32	20.25	20.34	18.03
WHS (proposed)	29.48	25.73	23.6	22.13	20.95
Max	29.77	25.73	23.6	22.13	20.99
PSNR of MRI 2					
SHW	29.24	23.99	23.63	22.14	21.16
SWH	28.19	22.72	21.94	20.62	19.75
HSW	24.98	23.32	15.52	20.54	13.89
HWS	24.47	22.34	21.03	20.04	19.32
WSH	29.81	25.56	15.78	21.27	14.01
WHS (proposed)	29.79	25.92	23.78	22.24	21.13
Max	29.81	25.92	23.78	22.24	21.16
PSNR of MRI 3					
SHW	29.51	25.66	23.26	21.46	20.81
SWH	28.73	24	21.86	20.08	19.6
HSW	22.48	22.95	20.65	19.69	18.57
HWS	24.7	22.41	20.95	19.88	19.18
WSH	24.49	24.91	21.54	20.3	18.95
WHS (proposed)	29.85	25.9	23.7	22.11	20.89
Max	29.85	25.9	23.7	22.11	20.89

TABLE 5: Average PSNR values of denoising the cartoon, texture, and residual parts in order ( $W$ : Wiener filter,  $H$ : hard threshold, and  $S$ : soft threshold).

Noise variance	0.01	0.03	0.05	0.07	0.09
SHW	29.36	25.05	23.47	21.67	20.99
SWH	28.54	23.60	21.93	20.25	19.67
HSW	24.05	23.10	18.55	19.96	16.71
HWS	24.50	22.33	20.96	19.92	19.21
WSH	28.02	25.26	19.19	20.64	17.00
WHS (proposed)	29.71	25.85	23.69	22.16	20.99
Max	29.71	25.85	23.69	22.16	20.99

the proposed method preserve more edge information and less noise.

#### 4. Discussion

The proposed method achieves better denoising effects than other concerned methods, both objectively and subjectively. Let us explore the relative optimality of the proposed method.

Our goal is to use a combination of classical spatial and transform domain filters to achieve better denoising effects. The MCA can decompose an image into different morphological components, which enables us to combine different methods conveniently. The Wiener filter is a spatial domain filter that can preserve most of the edge information. The wavelet thresholding is a transform domain filter based on the property of sparsity. It can amplify the dissimilarity between noise and true signal by transformation and then use thresholding functions to reduce noise. There are two

common thresholding functions: hard thresholding and soft thresholding. Hard thresholding gives better results by preserving edge information in some cases [25]. Soft thresholding tends to over smoothen the restored image. Transform domain methods can represent textures and low contrast information [3].

Decompose the noise-added MR images by MCA into three parts: the cartoon, texture, and residual parts. The denoising methods used in the proposed method are the Wiener Filter, hard threshold, and soft threshold. As an experiment validation, we will use all possible combinations of the three methods to remove noises in the three parts. To keep things simple, denote by  $W$ ,  $H$ , and  $S$  the Wiener Filter, hard threshold, and soft threshold, respectively. Table 4 shows the PSNR values of different methods used in the three parts of the images. For example, WHS means using the Wiener filter, hard threshold, and soft threshold to remove noises in the cartoon, texture, and residual parts, respectively. In contrary, WHS is our proposed method. It is

observed that the proposed method always has the highest PSNR values except when the noise variance in MRI 1-2 is 0.01. However, the average PSNR values of the proposed method still reach the highest even when the noise variance is 0.01. Table 5 reveals the fact.

As an exploratory research method, the proposed MRI denoising method is relatively effective, but the denoising effect is not so satisfactory. In the future, we will study more excellent denoising methods for MRI and fMRI.

## 5. Conclusion

In summary, we can illustrate our work in three steps. Firstly, describe the merits and drawbacks of traditional image denoising methods: Wiener filter, wavelet hard threshold, and wavelet soft threshold. Secondly, propose a comprehensive denoising algorithm based on the morphological component analysis. It can be briefly described as follows. Separate a noise-added image into three parts: two components that can be sparsely represented (cartoon and texture parts) and one residual part that cannot be sparsely represented (the residual part). Then, use the Wiener filter, wavelet hard threshold, and wavelet soft threshold to denoise the cartoon, texture, and residual parts, respectively. Finally, reconstruct the denoised image by adding the three denoised parts. Thirdly, analyze the relative best performance of the proposed method objectively and subjectively, explain our original intention, and verify the experimental results.

## 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 Science Foundation for Youth Teacher of Fujian Educational Committee (2016JAT160695 and 2018JT180872).

## References

- [1] A. H. K. Upadhyay, B. Talawar, and J. Rajan, "GPU implementation of non-local maximum likelihood estimation method for denoising magnetic resonance images," *Journal of Real-Time Image Processing*, vol. 13, no. 1, pp. 181–192, 2017.
- [2] L. Liu, H. Yang, J. Fan, R. W. Liu, and Y. Duan, "Rician noise and intensity nonuniformity correction (NNC) model for MRI data," *Biomedical Signal Processing and Control*, vol. 49, pp. 506–519, 2019.
- [3] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, "Image denoising review: from classical to state-of-the-art approaches," *Information Fusion*, vol. 55, pp. 220–244, 2020.
- [4] M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, and M. Adjouadi, "A comprehensive survey on impulse and Gaussian denoising filters for digital images," *Signal Processing*, vol. 157, pp. 236–260, 2019.
- [5] J. Jingdong Chen, J. Benesty, Y. Yiteng Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1218–1234, 2006.
- [6] J. Mohan, V. Krishnaveni, and Y. Guo, "A new neutrosophic approach of wiener filtering for MRI denoising," *Measurement Science Review*, vol. 13, no. 4, pp. 177–186, 2013.
- [7] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation BY wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [8] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [9] Y. Zhang, W. Ding, F. Pan et al., "Improved wavelet threshold for image de-noising," *Frontiers in Neuroscience*, vol. 13, 2019.
- [10] F. Xiao and Y. Zhang, "A comparative study on thresholding methods in wavelet-based image denoising," *Procedia Engineering*, vol. 15, pp. 3998–4003, 2011.
- [11] V. K. Sudarshan, M. R. K. Mookiah, U. R. Acharya et al., "Application of wavelet techniques for cancer diagnosis using ultrasound images: a Review," *Computers in Biology and Medicine*, vol. 69, pp. 97–111, 2016.
- [12] C. Huang, Y. Xie, Y. Lan et al., "A new framework for the integrative analytics of intravascular ultrasound and optical coherence tomography images," *IEEE Access*, vol. 6, pp. 36408–36419, 2018.
- [13] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1570–1582, 2005.
- [14] S. Janardhana, J. Jaya, K. J. Sabareesaan et al., "Image noise removal framework based on morphological component analysis," in *Proceedings of the International Conference on Current Trends in Engineering and Technology (ICCTET)*, vol. 2013, pp. 63–66, Coimbatore, India, July 2013.
- [15] L. Mitiche, A. B. Houada Adamou-Mitiche, and H. Naimi, "Medical image denoising using dual tree complex thresholding wavelet transform," in *Proceedings of the IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–5, Amman, Jordan, December 2013.
- [16] X. Deng and Z. Liu, "An improved image denoising method applied in resisting mixed noise based on MCA and Median filter," in *Proceedings of the 11th International Conference on Computational Intelligence and Security (CIS)*, pp. 162–166, Shenzhen, China, December 2015.
- [17] Y. Cheng and Z. Liu, "Image denoising algorithm based on structure and texture part," in *Proceedings of the 2016 12th International Conference on Computational Intelligence And Security (CIS)*, pp. 147–151, Xi'an, China, December 2016.
- [18] C. Huang and Y. Lan, "Patient-specific coronary artery 3D printing based on intravascular optical coherence tomography and coronary angiography," *Complexity*, vol. 2019, Article ID 5712594, 10 pages, 2019.
- [19] Dr. Ian Bickle, Normal MRI Brain: Adult, radiopaedia.org [DB/OL], <https://radiopaedia.org/cases/normal-mri-brain-adult>, 15 Apr 2017.
- [20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Pearson Upper Saddle River, NJ, USA, 2017.
- [21] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, Hoboken, NJ, USA, 2000.
- [22] M. Kazubek, "Wavelet domain image denoising by thresholding and Wiener filtering," *IEEE Signal Processing Letters*, vol. 10, no. 11, pp. 324–326, 2003.

- [23] P. S. Adsion, *The Illustrated Wavelet Transform Handbook*, Institute of Physics, London, UK, 2002.
- [24] S. Chen, D. Donoho, and M. Saunder, "Atomic decomposition by basic pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1999.
- [25] Y. Chen and C. Han, "Adaptive wavelet threshold for image denoising," *Electronics Letters*, vol. 41, no. 10, pp. 586-587, 12 May 2005.

## Research Article

# Bayesian Estimation of Gumbel Type-II Distribution under Type-II Censoring with Medical Applications

Kamran Abbas <sup>1</sup>, Zamir Hussain,<sup>2</sup> Noreen Rashid,<sup>3</sup> Amjad Ali <sup>4</sup>, Muhammad Taj,<sup>5</sup> Sajjad Ahmad Khan,<sup>4</sup> Sadaf Manzoor,<sup>4</sup> Umair Khalil,<sup>6</sup> and Dost Muhammad Khan <sup>6</sup>

<sup>1</sup>Department of Statistics, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan

<sup>2</sup>Research Centre for Modeling and Simulation, National University of Sciences and Technology, H-12 Campus, Islamabad, Pakistan

<sup>3</sup>Department of Statistics, Allama Iqbal Open University, Islamabad, Pakistan

<sup>4</sup>Department of Statistics, Islamia College University, Peshawar, Khyber Pakhtunkhwa, Pakistan

<sup>5</sup>Department of Mathematics, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan

<sup>6</sup>Department of Statistics, Abdul Wali Khan University, Mardan, Khyber Pakhtunkhwa, Pakistan

Correspondence should be addressed to Kamran Abbas; [kamiuajk@gmail.com](mailto:kamiuajk@gmail.com)

Received 11 December 2019; Revised 9 February 2020; Accepted 10 February 2020; Published 26 March 2020

Guest Editor: Pritee Khanna

Copyright © 2020 Kamran Abbas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The time to event or survival time usually follows certain skewed probability distributions. These distributions encounter vital role using the Bayesian framework to analyze and project the maximum life expectancy in order to inform decision-making. The Bayesian method provides a flexible framework for monitoring the randomized clinical trials to update what is already known using prior information about specific phenomena under uncertainty. Additionally, medical practitioners can use the Bayesian estimators to measure the probability of time until tumor recurrence, time until cardiovascular death, and time until AIDS for HIV patients by considering the prior information. However, in clinical trials and medical studies, censoring is present when an exact event occurrence time is not known. The present study aims to estimate the parameters of Gumbel type-II distribution based on the type-II censored data using the Bayesian framework. The Bayesian estimators cannot be obtained in explicit forms, and therefore we use Lindley's approximation based on noninformative prior and various loss functions such as squared error loss function, general entropy loss function, and LINEX (linear exponential) loss function. The maximum likelihood and Bayesian estimators are compared in terms of mean squared error by using the simulation study. Furthermore, two data sets about remission times (in months) of bladder cancer patients and survival times in weeks of 61 patients with inoperable adenocarcinoma of the lung are analyzed for illustration purposes.

## 1. Introduction

In medical research, data supporting the time until the occurrence of a particular event, such as the death of a patient, are frequently encountered. Such data are referred to as survival time data which has generally right-skewed distribution, and Gumbel type-II distribution can be used for this purpose. It was introduced by the German mathematician Gumbel in [1] and is useful to model “extreme values” such as floods, earthquakes, and natural disasters and also used in life expectancy tables, hydrology, and

rainfall. The probability density function (PDF) of Gumbel type-II distribution is

$$f(x|\alpha, \beta) = \alpha\beta x^{-(\alpha+1)} \exp(-\beta x^{-\alpha}), \quad x > 0, \alpha, \beta > 0, \quad (1)$$

where “ $\alpha$ ” is the shape and “ $\beta$ ” is the scale parameter of the distribution. The corresponding cumulative distribution function (CDF) is

$$F(x|\alpha, \beta) = \exp(-\beta x^{-\alpha}). \quad (2)$$

A common feature of lifetime data is that the data points are possibly censored. In manifold reliability and life-testing

studies, experiments are generally windup before failure times of all items are observed. Therefore, adequate information and results on failure times of all objects cannot be obtained. During experimentation, these situations occur due to loss or removal of objects before they fail. Therefore, generally, such experiments are preplanned and purposeful to save time and cost of these testing. Data obtained from such experiments are called censored. The type-I and type-II censoring are two well-known censoring schemes. In type-II censoring scheme, the number of failure units are fixed in advanced. For example, the investigator may decide to terminate the study after four of the six rats have developed tumors. There is an enormous literature accessible on estimation of parameters of distributions using type-II censoring, for example, Abbas and Tang [2] considered ML and least square estimators of Frechet distribution using type-II censored samples. Okasha [3] estimated the unknown parameters, reliability, and hazard functions of Lomax distribution under type-II censoring using Bayesian and E-Bayesian estimation. Abu-Zinadah [4] studied on exponentiated Gompertz distribution based on type-II and complete censored data. El-Sagheer [5] studied the generalized pareto distribution under the different censoring schemes.

Recently, many authors have worked on Gumbel type-II distribution and Bayesian estimation using different loss functions. Abbas et al. [6] worked on Gumbel type-II distribution and obtained the Bayes estimators under different loss functions. Feroze and Aslam [7] obtained Bayes estimators of two components of Gumbel type-II distribution. Malinowska and Szyal [8] also derived Bayes estimators for Gumbel type-II distribution on  $k$ th lower record values. Sultana et al. [9] worked on a three-component mixture of Gumbel type-II distribution using Bayesian estimation under different priors such as informative and noninformative. Moreover, Metiri et al. [10] worked on the properties of the Lindley distribution. The Bayes estimates were derived under LINEX loss function using informative and noninformative priors (Reyad and Ahmed [11]). Preda et al. [12] developed Bayes estimators of modified Weibull distribution under squared error loss function (SELF) and LINEX loss function.

However, Bayesian estimation of Gumbel type-II distribution based on type-II censoring is not frequently discussed; therefore, we are interested in estimating the unknown parameters of Gumbel type-II distribution under type-II censored data. Including this introduction section, the rest of the paper is arranged as follows: in Section 2, maximum likelihood estimators (MLEs) for the parameters are obtained. In Section 3, Bayesian estimators based on different loss functions by taking noninformative and gamma priors are derived. The proposed estimators are compared in terms of their mean squared error (MSE) in Section 4. Section 5 illustrates the applications of proposed estimators with two examples, namely, data set of remission times for bladder cancer and survival times of inoperable adenocarcinoma of the lung. Finally, conclusions and recommendations are presented in Section 6.

## 2. Maximum Likelihood Estimation

Suppose that  $X_1 < X_2 < \dots < X_r$  is a type-II censored sample of size " $r$ " obtained from a life test on " $n$ " items whose life times have the Gumbel type-II distribution with parameters " $\alpha$ " and " $\beta$ ." The likelihood function of " $r$ " failures and  $(n - r)$  censored values may be written as

$$L = \left( \prod_{i=1}^r f(x_i, \theta) \right) (1 - F(x_r, \theta))^{n-r}, \quad (3)$$

$$L = \left( \prod_{i=1}^r \alpha \beta x_i^{-(\alpha+1)} \exp(-\beta x_i^{-\alpha}) \right) (1 - \exp(-\beta x_r^{-\alpha}))^{n-r}. \quad (4)$$

It is more convenient to work with log-likelihood. The log-likelihood function is

$$\ln L = r \ln \alpha + r \ln \beta - (\alpha + 1) \sum_{i=1}^r \ln x_i - \beta \sum_{i=1}^r x_i^{-\alpha} + (n - r) \ln(1 - \exp(-\beta x_r^{-\alpha})). \quad (5)$$

To get the ML estimator of  $\alpha$  and  $\beta$ , differentiate equation (5) with respect to  $\alpha$  and  $\beta$  and the resulting equations are

$$\frac{\partial \ln L}{\partial \alpha} = \frac{r}{\alpha} - \sum_{i=1}^r \ln x_i + \beta \sum_{i=1}^r x_i^{-\alpha} \ln x_i + (n - r) \cdot \left( \frac{(-\exp(-\beta x_r^{-\alpha})) (\beta x_r^{-\alpha} \ln x_r)}{1 - \exp(-\beta x_r^{-\alpha})} \right) = 0, \quad (6)$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{r}{\beta} - \sum_{i=1}^r x_i^{-\alpha} + (n - r) \cdot \left( \frac{-\exp(-\beta x_r^{-\alpha}) (-x_r^{-\alpha})}{1 - \exp(-\beta x_r^{-\alpha})} \right) = 0. \quad (7)$$

Equations (6) and (7) cannot be written in closed form. Therefore, here, we use the Laplace approximation to get the point estimates of  $\alpha$  and  $\beta$ .

## 3. Bayesian Estimation

In Bayesian estimation, we consider different loss functions such as squared error loss function (SELF) proposed by Legendre [13] and Gauss [14], LINEX (Varian [15]), and general entropy loss function (GELF) introduced by Calabria and Pulcini [16]. As both parameters are unknown, independent noninformative form of priors can be used. Supposed that  $\alpha$  and  $\beta$  have independent Gamma ( $a, b$ ) and Gamma ( $c, d$ ) priors, respectively, for  $a, b, c, d > 0$ , i.e.,

$$\pi_1(\alpha) \propto \alpha^{a-1} (\exp(-b\alpha)), \quad (8)$$

$$\pi_2(\beta) \propto \beta^{c-1} (\exp(-d\beta)). \quad (9)$$

The joint prior distribution of parameters is



$$\begin{aligned}\Phi'(\alpha, \beta | x) &= K(\alpha^{(a+r-1)})(\beta^{(c+r-1)}) \\ &\cdot \left( \exp\left(-b\alpha - d\beta - \beta \sum_{i=1}^r x_i^{-\alpha}\right) \right) \left( \prod_{i=1}^r x_i^{-\alpha-1} \right) \\ &\cdot \left( 1 - \exp\left(\sum_{i=1}^r x_i^{-\alpha}\right) \right),\end{aligned}\quad (10)$$

where  $K$  is the normalizing constant that makes  $\Phi'(\alpha, \beta | x)$  a proper PDF. Thus,

$$\begin{aligned}K^{-1} &= \int_{\alpha, \beta} (\alpha^{(a+r-1)})(\beta^{(c+r-1)}) \\ &\cdot \left( \exp\left(-b\alpha - d\beta - \beta \sum_{i=1}^r x_i^{-\alpha}\right) \right) \left( \prod_{i=1}^r x_i^{-\alpha-1} \right) \\ &\cdot \left( 1 - \exp\left(\sum_{i=1}^r x_i^{-\alpha}\right) \right) d\alpha d\beta.\end{aligned}\quad (11)$$

Therefore, the joint posterior density under any loss function is

$$\Phi'(\alpha, \beta | x) = \frac{\int_{\alpha} \int_{\beta} g(\alpha, \beta) L(\alpha, \beta) \pi_1(\alpha) \pi_2(\beta) d\alpha d\beta}{\int_{\alpha} \int_{\beta} L(\alpha, \beta) \pi_1(\alpha) \pi_2(\beta) d\alpha d\beta}. \quad (12)$$

Posterior distribution (12) takes a ratio form that cannot be reduced to a closed form. Therefore, we use Lindley approximation [17] to get the Bayesian estimates, which can be written as

$$\begin{aligned}\hat{g} &= g(\hat{\alpha}, \hat{\beta}) + \frac{1}{2} \left[ \sum_{i=1}^2 \sum_{j=1}^2 l_{ij} S_{ij} + l_{30} A_{12} + l_{03} A_{21} + l_{21} B_{12} + l_{12} B_{21} \right] \\ &+ q_1 C_{12} + q_2 C_{21},\end{aligned}\quad (13)$$

where  $q_1 = (\partial \ln \pi(\alpha, \beta) / \partial \alpha)$ ;  $q_2 = (\partial \ln \pi(\alpha, \beta) / \partial \beta)$ ;  $l_{11} = (\partial^2 g(\alpha, \beta) / \partial \alpha^2)$ ;  $l_{12} = (\partial^2 g(\alpha, \beta) / \partial \alpha \partial \beta)$ ;  $l_{21} = (\partial^2 g(\alpha, \beta) / \partial \beta \partial \alpha)$ ;  $l_{22} = (\partial^2 g(\alpha, \beta) / \partial \beta^2)$ ;  $l_1 = (\partial g(\alpha, \beta) / \partial \alpha)$ ;  $l_2 = (\partial g(\alpha, \beta) / \partial \beta)$ ;  $A_{ij} = (l_i S_{ii} + l_j S_{ij}) S_{ii}$ ;  $B_{ij} = 3l_i S_{ii} S_{ij} + l_j (S_{ii} S_{jj} + 2S_{ij}^2)$ ;  $C_{ij} = l_i S_{ii} + l_j S_{ji}$ ,  $j = 1, 2$ . The detail of equation (13) is provided in Appendix. The approximate Bayesian estimators of “ $\alpha$ ” and “ $\beta$ ” based on SELF are

$$\begin{aligned}\hat{\alpha}_{\text{BSELF}} &= \hat{\alpha} + \frac{1}{2} \left\{ \left[ \frac{2r}{\hat{\alpha}^3} + \hat{\beta} \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^3 - \left( \frac{(n-r)(1-A)\hat{\beta}(\ln x_r)^3 x_r^{-\hat{\alpha}}}{A^3} \right) \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (2-A) - A(1+A) \right) + A^2 \right) \right] \right. \\ &S_{11}^2 + \left\{ \frac{2r}{\hat{\beta}^3} - \frac{(n-r)x_r^{-3\hat{\alpha}}(1-A)(A-2)}{A^3} \right\} S_{21} S_{22} \\ &+ 3 \left\{ - \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^2 - \left( \frac{(1-A)(n-r)x_r^{-\hat{\alpha}} (\ln x_r)^2}{A^3} \right) \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (A-2) + A(A+1) \right) - A^2 \right) \right\} \\ &S_{11} S_{12} + \left\{ - \frac{(n-r)(1-A)x_r^{-2\hat{\alpha}} \ln x_r}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} (2-A) - 2A \right) \right\} \left( S_{22} S_{21} + 2S_{21}^2 \right) \\ &+ \left( \frac{a-1}{\hat{\alpha}} - b \right) S_{11} + \left( \frac{c-1}{\hat{\beta}} - d \right) S_{21},\end{aligned}\quad (14)$$

$$\begin{aligned}
\hat{\beta}_{\text{BSELF}} = & \hat{\beta} + \frac{1}{2} \left[ \left\{ \frac{2r}{\hat{\alpha}^3} + \hat{\beta} \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^3 - \left( \frac{(n-r)(1-A)\hat{\beta}(\ln x_r)^3 x_r^{-\hat{\alpha}}}{A^3} \right) \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (2-A) - A(1+A) \right) + A^2 \right) \right\} \right. \\
& S_{12} S_{11} + \left\{ \frac{2r}{\hat{\beta}^3} - \frac{(n-r)x_r^{-3\hat{\alpha}}(1-A)(A-2)}{A^3} \right\} S_{22}^2 + \left\{ - \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^2 - \left( \frac{(1-A)(n-r)x_r^{-\hat{\alpha}}(\ln x_r)^2}{A^3} \right) \right. \\
& \cdot \left. \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (A-2) + A(A+1) \right) - A^2 \right) \right\} \\
& (S_{11} S_{22} + 2S_{12}^2) + 3 \left\{ - \left( \frac{(n-r)(1-A)x_r^{-2\hat{\alpha}} \ln x_r}{A^3} \right) \left( \hat{\beta} x_r^{-\hat{\alpha}} (2-A) - 2A \right) \right\} S_{22} S_{21} \Bigg] \\
& + \left( \frac{a-1}{\hat{\alpha}} - b \right) S_{21} + \left( \frac{c-1}{\hat{\beta}} - d \right) S_{22}.
\end{aligned} \tag{15}$$

Similarly, the Bayesian estimators of “ $\alpha$ ” and “ $\beta$ ” under LINEX loss function are

$$\begin{aligned}
\hat{\alpha}_{\text{BLINEX}} = & -\frac{1}{k} \ln \left[ e^{-k\hat{\alpha}} + \frac{1}{2} \left\{ k^2 e^{-k\hat{\alpha}} S_{11} - k e^{-k\hat{\alpha}} \left( \frac{2r}{\hat{\alpha}^3} + \hat{\beta} \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^3 - \left( \frac{(n-r)(1-A)\hat{\beta}(\ln x_r)^3 x_r^{-\hat{\alpha}}}{A^3} \right) \right. \right. \right. \\
& \left. \left. \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (2-A) - A(1+A) \right) + A^2 \right) \right) S_{11}^2 \right. \right. \\
& - k e^{-k\hat{\alpha}} \left( \frac{2r}{\hat{\beta}^3} - \frac{(n-r)x_r^{-3\hat{\alpha}}(1-A)(A-2)}{A^3} \right) S_{21} S_{22} \\
& - 3k e^{-k\hat{\alpha}} \left( - \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^2 - \left( \frac{(1-A)(n-r)x_r^{-\hat{\alpha}}(\ln x_r)^2}{A^3} \right) \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (A-2) + A(A+1) \right) - A^2 \right) \right) S_{11} S_{12} \\
& - k e^{-k\hat{\alpha}} \left( - \left( \frac{(n-r)(1-A)x_r^{-2\hat{\alpha}} \ln x_r}{A^3} \right) \left( \hat{\beta} x_r^{-\hat{\alpha}} (2-A) - 2A \right) \right) (S_{22} S_{21} + 2S_{21}^2) \Bigg\} \\
& - k e^{-k\hat{\alpha}} \left( \frac{a-1}{\hat{\alpha}} - b \right) S_{11} - k e^{-k\hat{\alpha}} \left( \frac{c-1}{\hat{\beta}} - d \right) S_{21} \Bigg],
\end{aligned} \tag{16}$$

$$\begin{aligned}
\hat{\beta}_{\text{BLINEX}} = & -\frac{1}{k} \ln \left[ e^{-k\hat{\beta}} + \frac{1}{2} \left\{ k^2 e^{-k\hat{\beta}} S_{22} - k e^{-k\hat{\beta}} \right. \right. \\
& \cdot \left( \frac{2r}{\hat{\alpha}^3} + \hat{\beta} \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^3 - \frac{(n-r)(1-A)\hat{\beta}(\ln x_r)^3 x_r^{-\hat{\alpha}}}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (2-A) - A(1+A) \right) + A^2 \right) \right) \\
& S_{12} S_{11} - k e^{-k\hat{\beta}} \left( \frac{2r}{\hat{\beta}^3} - \frac{(n-r)x_r^{-3\hat{\alpha}}(1-A)(A-2)}{A^3} \right) S_{22}^2 \\
& - k e^{-k\hat{\beta}} \left( - \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^2 - \frac{(1-A)(n-r)x_r^{-\hat{\alpha}}(\ln x_r)^2}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (A-2) + A(A+1) \right) - A^2 \right) \right) \\
& \cdot (S_{11} S_{22} + 2S_{12}^2) \\
& - 3k e^{-k\hat{\beta}} \left( - \frac{(n-r)(1-A)x_r^{-2\hat{\alpha}} \ln x_r}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} (2-A) - 2A \right) \right) S_{22} S_{21} \Big\} \\
& \left. - k e^{-k\hat{\beta}} \left( \frac{a-1}{\hat{\alpha}} - b \right) S_{21} - k e^{-k\hat{\beta}} \left( \frac{c-1}{\hat{\beta}} - d \right) S_{22} \right]. \tag{17}
\end{aligned}$$

The Bayesian estimators of “ $\alpha$ ” and “ $\beta$ ” under GELF are

$$\begin{aligned}
\hat{\alpha}_{\text{BGELF}} = & \left[ \hat{\alpha}^{-k} + \frac{1}{2} \left\{ k(k+1) \hat{\alpha}^{-(k+2)} S_{11} - k \hat{\alpha}^{-(k+1)} \right. \right. \\
& \left( \frac{2r}{\hat{\alpha}^3} + \hat{\beta} \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^3 - \frac{(n-r)(1-A)\hat{\beta}(\ln x_r)^3 x_r^{-\hat{\alpha}}}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (2-A) - A(1+A) \right) + A^2 \right) \right) S_{11}^2 \\
& - k \hat{\alpha}^{-(k+1)} \left( \frac{2r}{\hat{\beta}^3} - \frac{(n-r)x_r^{-3\hat{\alpha}}(1-A)(A-2)}{A^3} \right) S_{21} S_{22} \\
& - 3k \hat{\alpha}^{-(k+1)} - \left( \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^2 - \frac{(1-A)(n-r)x_r^{-\hat{\alpha}}(\ln x_r)^2}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( \left( \hat{\beta} x_r^{-\hat{\alpha}} - A \right) (A-2) + A(A+1) \right) - A^2 \right) \right) S_{11} S_{12} \\
& - k \hat{\alpha}^{-(k+1)} \left( - \frac{(n-r)(1-A)x_r^{-2\hat{\alpha}} \ln x_r}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} (2-A) - 2A \right) \right) (S_{22} S_{21} + 2S_{21}^2) \Big\} \\
& \left. - k \hat{\alpha}^{-(k+1)} \left( \frac{a-1}{\hat{\alpha}} - b \right) S_{11} - k \hat{\alpha}^{-(k+1)} \left( \frac{c-1}{\hat{\beta}} - d \right) S_{21} \right]^{-1/k}, \tag{18}
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_{\text{BGELF}} = & \left[ \hat{\beta}^{-k} + \frac{1}{2} \left\{ k(k+1) \hat{\beta}^{-(k+2)} S_{22} \right. \right. \\
& - k \hat{\beta}^{-(k+1)} \left( \frac{2r}{\hat{\alpha}^3} + \hat{\beta} \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^3 - \frac{(n-r)(1-A) \hat{\beta} (\ln x_r)^3 x_r^{-\hat{\alpha}}}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( (\hat{\beta} x_r^{-\hat{\alpha}} - A) (2-A) - A(1+A) \right) + A^2 \right) \right) S_{12} S_{11} \\
& - k \hat{\beta}^{-(k+1)} \left( \frac{2r}{\hat{\beta}^3} - \frac{(n-r) x_r^{-3\hat{\alpha}} (1-A)(A-2)}{A^3} \right) S_{22}^2 - k \hat{\beta}^{-(k+1)} \\
& \cdot \left( - \sum_{i=1}^r x_i^{-\hat{\alpha}} (\ln x_i)^2 - \frac{(1-A)(n-r) x_r^{-\hat{\alpha}} (\ln x_r)^2}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} \left( (\hat{\beta} x_r^{-\hat{\alpha}} - A) (A-2) + A(A+1) \right) - A^2 \right) \right) (S_{11} S_{22} + 2S_{12}^2) \\
& - 3k \hat{\beta}^{-(k+1)} \left( - \frac{(n-r)(1-A) x_r^{-2\hat{\alpha}} \ln x_r}{A^3} \left( \hat{\beta} x_r^{-\hat{\alpha}} (2-A) - 2A \right) \right) S_{22} S_{21} \left. \right\} \\
& - k \hat{\beta}^{-(k+1)} \left( \frac{a-1}{\hat{\alpha}} - b \right) S_{21} - k \hat{\beta}^{-(k+1)} \left( \frac{c-1}{\hat{\beta}} - d \right) S_{22} \left. \right]^{-1/k},
\end{aligned} \tag{19}$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are the ML estimators of  $\alpha$  and  $\beta$  which can be obtained from equations (6) and (7), respectively.

#### 4. Simulation Study

The performance of the proposed Bayesian estimators with their ML counterpart in terms of MSE, different sample sizes, and different values of parameters are considered using Monte Carlo simulation based on prespecified different percentages of failures, i.e., 40%, 60%, and 80%. Monte Carlo simulation is conducted as follows:

- (i) Take the initial values of  $\alpha$  and  $\beta$ , respectively, and the samples are generated from the Gumbel type-II distribution using inverse transformation technique, i.e.,  $X(F) = (-(\log U/\beta))^{-(1/\alpha)}$ , where  $U \sim \text{uniform}(0, 1)$ .
- (ii) First, calculate the MLE using Laplace approximation, and then Bayesian estimates under non-informative priors are obtained via Lindley approximation.
- (iii) The process is replicated 5000 times for each sample size and averages of these estimates and the corresponding MSEs (within parenthesis) were calculated for each method using the R software version (i386 3.6.1), which approximately takes around half an hour.

The results are reported in Tables 1–4 for comparison purposes. Tables 1 and 2 contain simulation results for the case where loss function parameter  $k=1$  and values of hyperparameters are considered as  $a=b=c=d=2$ , whereas Tables 3 and 4 comprise the results for the case where  $k=1.5$  and values of hyperparameters are  $a=1$ ,  $b=2$ ,  $c=2.25$ , and  $d=1.5$  for the simulation study. From the results of the simulation study, conclusions are drawn regarding the behavior of the estimators, which are summarized as follows:

- (i) In terms of MSEs, the ML and Bayesian estimators become closer by increasing the sample sizes.

- (ii) For fixed percentage of failures, as sample size increases, it is observed that the MSEs of all the estimators decrease because as for large sample sizes, prior has minimal effect on the posterior.
- (iii) For fixed values of  $\alpha$  and  $\beta$ , the MSEs of ML and Bayesian estimators decrease when both increase the sample size and percentage of failures.
- (iv) When  $k=1$  and  $a=b=c=d=2$ , the Bayesian estimators based on GELF and LINEX loss function are smaller as compared to ML estimators in terms of MSEs. Therefore, Bayes estimators are much stable than ML estimators.
- (v) Generally, the ML and Bayesian estimators are closed for the large sample in terms of MSE.

#### 5. Data Analysis

In this section, we consider two examples for illustration purposes.

**5.1. Example 1.** The real data about remission times (in months) of a random sample of 128 bladder cancer patients presented in Table 5 were reported by Lee and Wang [20]. A total of 128 patients with different prespecified percentages of events, i.e., 40%, 50%, 60%, and 80%, represented patients whose treatment was terminated and rest of the percentages are censored. Clearly, Figure 1 confirms that the histogram is slightly skewed to the right and is leptokurtic. Moreover, ML and Bayesian estimates can also be envisioned in Figure 1, in which the  $x$ -axis represents the remission times (in months) of bladder cancer patients, while the Gumbel type-II density function is taken on the  $y$ -axis. Therefore, it would be appropriate to select positively skewed distributions for describing the behavior of remission times of bladder cancer patients. Amongst the skewed distributions, Gumbel type-II distribution is fitted and the parameter estimates using ML and Bayesian methods are presented in Table 6 for comparison purposes.

TABLE 1: Average ML and Bayesain estimates with corresponding MSEs (within parenthesis) using different percentages of failures for  $\alpha = 1.5$  (when  $a = b = c = d = 2$ ,  $k = 1$ ).

$n$	Percent	ML	BSELF	BGELF	BLINEX
25	40	1.7162 (0.2808)	1.7948 (0.3524)	1.4560 (0.0989)	1.4666 (0.0981)
	60	1.6408 (0.1565)	1.6893 (0.1822)	1.4633 (0.0794)	1.4708 (0.0758)
	80	1.6160 (0.1090)	1.6527 (0.1231)	1.4883 (0.0645)	1.4947 (0.0620)
50	40	1.6263 (0.0985)	1.6580 (0.1112)	1.5114 (0.0587)	1.5179 (0.0565)
	60	1.5610 (0.0579)	1.5826 (0.0624)	1.4836 (0.0429)	1.4891 (0.0416)
	80	1.5667 (0.0412)	1.5840 (0.0446)	1.5084 (0.0308)	1.5126 (0.0303)
80	40	1.5708 (0.0521)	1.5889 (0.0561)	1.5042 (0.0384)	1.5095 (0.0375)
	60	1.5516 (0.0348)	1.5648 (0.0369)	1.5041 (0.0278)	1.5080 (0.0274)
	80	1.5269 (0.0223)	1.5373 (0.0233)	1.4926 (0.0196)	1.4958 (0.0193)
100	40	1.5540 (0.0420)	1.5681 (0.0445)	1.5023 (0.0334)	1.5069 (0.0328)
	60	1.5345 (0.0278)	1.5448 (0.0289)	1.4975 (0.0237)	1.5008 (0.0234)
	80	1.5292 (0.0176)	1.5374 (0.0184)	1.5020 (0.0155)	1.5045 (0.0154)
150	40	1.5327 (0.0243)	1.5448 (0.0253)	1.4992 (0.0210)	1.5026 (0.0207)
	60	1.5234 (0.0162)	1.5381 (0.0167)	1.4991 (0.0146)	1.5016 (0.0145)
	80	1.5143 (0.0128)	1.5208 (0.0131)	1.4977 (0.0119)	1.4995 (0.0118)
200	40	1.5317 (0.0188)	1.5384 (0.0195)	1.5067 (0.0165)	1.5092 (0.0164)
	60	1.5220 (0.0114)	1.5271 (0.0117)	1.5040 (0.0104)	1.5058 (0.0103)
	80	1.5119 (0.0089)	1.5160 (0.0091)	1.4987 (0.0084)	1.5001 (0.0083)

ML: maximum likelihood estimators, BSELF: Bayesian estimators under squared error loss function, BGELF: Bayesian estimators under general entropy loss function, and BLINEX: Bayesian estimators under linear exponential loss function.

TABLE 2: Average ML and Bayesian estimates with corresponding MSEs (within parenthesis) using different percentages of failures for  $\beta = 1.5$  (when  $a = b = c = d = 2$ ,  $k = 1$ ).

$n$	Percent	ML	BSELF	BGELF	BLINEX
25	40	1.5418 (0.1538)	1.6030 (0.1852)	1.3879 (0.1131)	1.3988 (0.1042)
	60	1.5907 (0.1522)	1.6488 (0.1737)	1.4357 (0.0886)	1.4448 (0.0842)
	80	1.5725 (0.1338)	1.6280 (0.1651)	1.4255 (0.0827)	1.4346 (0.0782)
50	40	1.5211 (0.0635)	1.5473 (0.0684)	1.4455 (0.0548)	1.4521 (0.0526)
	60	1.5396 (0.0575)	1.5644 (0.0632)	1.4677 (0.0467)	1.4735 (0.0451)
	80	1.5358 (0.0565)	1.5609 (0.0624)	1.4658 (0.0458)	1.4715 (0.0442)
80	40	1.5147 (0.0351)	1.5303 (0.0371)	1.4676 (0.0320)	1.4721 (0.0312)
	60	1.5215 (0.0349)	1.5364 (0.0364)	1.4774 (0.0311)	1.4814 (0.0304)
	80	1.5262 (0.0341)	1.5415 (0.0363)	1.4830 (0.0297)	1.4869 (0.0291)
100	40	1.5026 (0.0298)	1.5149 (0.0306)	1.4654 (0.0285)	1.4691 (0.0279)
	60	1.5220 (0.0260)	1.5338 (0.0272)	1.4868 (0.0237)	1.4901 (0.0233)
	80	1.5158 (0.0249)	1.5278 (0.0262)	1.4817 (0.0224)	1.4850 (0.0221)
150	40	1.5033 (0.0191)	1.5114 (0.0194)	1.4784 (0.0184)	1.4809 (0.0182)
	60	1.5118 (0.0165)	1.5194 (0.0170)	1.4886 (0.0157)	1.4909 (0.0155)
	80	1.5067 (0.0163)	1.5145 (0.0168)	1.4842 (0.0153)	1.4865 (0.0151)
200	40	1.5060 (0.0130)	1.5121 (0.0133)	1.4873 (0.0123)	1.4892 (0.0122)
	60	1.5105 (0.0119)	1.5162 (0.0122)	1.4931 (0.0113)	1.4949 (0.0112)
	80	1.5110 (0.0116)	1.5169 (0.0118)	1.4940 (0.0112)	1.4958 (0.0111)

TABLE 3: Average ML and Bayesian estimates with corresponding MSEs (within parenthesis) using different percentages of failures for  $\alpha = 1.5$  (when  $a = 1$ ,  $b = 2$ ,  $c = 2.25$ ,  $d = 1.5$ , and  $k = 1.5$ ).

$n$	Percent	ML	BSELF	BGELF	BLINEX
50	40	1.5879 (0.0923)	1.6185 (0.1025)	1.4239 (0.0629)	1.4251 (0.0605)
	60	1.5757 (0.0594)	1.5975 (0.0649)	1.4604 (0.0422)	1.4615 (0.0407)
	80	1.5404 (0.0451)	1.5575 (0.0477)	1.4576 (0.0374)	1.4590 (0.0363)
80	40	1.5434 (0.0408)	1.5607 (0.0435)	1.4439 (0.0338)	1.4454 (0.0326)
	60	1.5704 (0.0382)	1.5838 (0.0408)	1.4973 (0.0296)	1.4983 (0.0270)
	80	1.5493 (0.0200)	1.5597 (0.0214)	1.4479 (0.0278)	1.4982 (0.0151)
100	40	1.5280 (0.0337)	1.5419 (0.0354)	1.4599 (0.0171)	1.4495 (0.0288)
	60	1.5145 (0.0179)	1.5247 (0.0202)	1.4973 (0.0161)	1.4612 (0.0167)
	80	1.5389 (0.0173)	1.5472 (0.0186)	1.4971 (0.0154)	1.4982 (0.0150)
200	40	1.5370 (0.0158)	1.5439 (0.0165)	1.4959 (0.0131)	1.4968 (0.0129)
	60	1.5184 (0.0114)	1.5248 (0.0117)	1.4903 (0.0104)	1.4910 (0.0103)
	80	1.5216 (0.0113)	1.5257 (0.0116)	1.5007 (0.0103)	1.5013 (0.0103)



TABLE 4: Average ML and Bayesian estimates with corresponding MSEs (within parenthesis) using different percentages of failures for  $\beta = 1.5$  (when  $a = 1$ ,  $b = 2$ ,  $c = 2.25$ ,  $d = 1.5$ , and  $k = 1.5$ ).

$n$	Percent	ML	BSELF	BGELF	BLINEX
50	40	1.5341 (0.0661)	1.5818 (0.0734)	1.5067 (0.0546)	1.5085 (0.0521)
	60	1.5564 (0.0627)	1.5719 (0.0693)	1.4987 (0.0524)	1.5006 (0.0501)
	80	1.5464 (0.0494)	1.5599 (0.0536)	1.4829 (0.0430)	1.4856 (0.0408)
80	40	1.5052 (0.0382)	1.5459 (0.0407)	1.5008 (0.0342)	1.5024 (0.0332)
	60	1.5308 (0.0319)	1.5208 (0.0330)	1.4741 (0.0302)	1.4763 (0.0292)
	80	1.5187 (0.0285)	1.5337 (0.0301)	1.4901 (0.0260)	1.4919 (0.0254)
100	40	1.5209 (0.0261)	1.5333 (0.0272)	1.4954 (0.0241)	1.4970 (0.0235)
	60	1.5079 (0.0254)	1.5195 (0.0263)	1.4846 (0.0240)	1.4862 (0.0234)
	80	1.5075 (0.0212)	1.5193 (0.0221)	1.4851 (0.0200)	1.4867 (0.0195)
200	40	1.5210 (0.0151)	1.5269 (0.0156)	1.5194 (0.0148)	1.4794 (0.0146)
	60	1.5310 (0.0149)	1.5371 (0.0155)	1.5092 (0.0144)	1.5099 (0.0142)
	80	1.4907 (0.0147)	1.4967 (0.0149)	1.4783 (0.0138)	1.5201 (0.0137)

TABLE 5: Remission times (in months) of a random sample of 128 bladder cancer patients.

0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98  
 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50  
 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28  
 9.74, 14.76, 6.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64  
 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59,  
 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01,  
 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66,  
 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02,  
 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13,  
 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03,  
 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63,  
 22.69

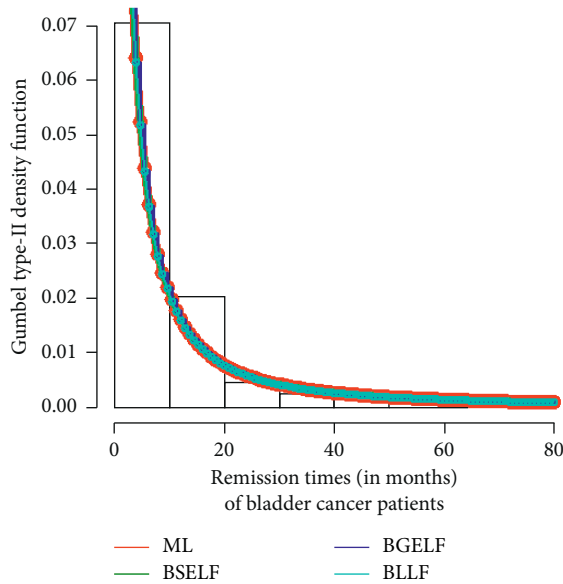


FIGURE 1: Comparison of estimation methods.

It is concluded that the proposed estimators of Gumbel type-II distribution fit the data well. Therefore, it is recommended that the Bayesian estimators can be more

TABLE 6: Point estimates of  $\alpha$  and  $\beta$  using different percentages of failures when  $a = 1$ ,  $b = 2$ ,  $c = 2.25$ ,  $d = 1.5$ , and  $k = 1$ .

	Percent	ML	BSELF	BGELF	BLINEX
$\alpha$	40	0.5433	0.5452	0.5287	0.5329
	50	0.5909	0.5921	0.5779	0.5812
	60	0.6339	0.6348	0.6225	0.6252
	80	0.6988	0.6994	0.6898	0.6917
$\beta$	40	2.4491	2.4702	2.3960	2.3915
	50	2.4252	2.4440	2.3729	2.3689
	60	2.4139	2.4312	2.3623	2.3594
	80	2.4095	2.4253	2.3588	2.3551

TABLE 7: Survival times in weeks of 61 patients with inoperable adenocarcinoma of the lung.

<i>28 Censored observations</i>	
0.14, 0.14, 0.29, 0.43, 0.57, 0.57, 1.86, 3.00, 3.00, 3.29, 3.29, 6.00 6.00, 6.14, 8.71, 10.57, 11.86, 15.57, 16.57, 17.29, 18.71, 21.29, 23.86, 26.00, 27.57, 32.14, 33.14, 47.29	
<i>33 Uncensored observations</i>	
0.43, 2.86, 3.14, 3.14, 3.43, 3.43, 3.71, 3.86, 6.14, 6.86, 9.00, 9.43 10.71, 10.86, 11.14, 13.00, 14.43, 15.71, 18.43, 18.57, 20.71, 29.14, 29.71, 40.57, 48.57, 49.43, 53.86, 61.86, 66.57, 68.71, 68.96, 72.86, 72.86	

beneficial to address the uncertainty in medical-related censored data.

5.2. Example 2. The survival times, in weeks, of 61 patients with unoperable lung cancer treated with cyclophosphamide considered in Lagakos and Williams ([18]) and in Lee and Wolfe ([19]) are presented in Table 7. There are 33 uncensored observations and 28 censored observations, representing the patients whose treatment was terminated because of a devolving condition. The point estimates of  $\alpha$  and  $\beta$  obtained by all the methods are summarized in Table 8. Figure 2 shows the results of different estimation methods and depicts that Gumbel type-II distribution fits the data better, in which  $x$ -axis comprises the survival times in weeks of 61 patients with inoperable adenocarcinoma of

TABLE 8: Point estimates of  $\alpha$  and  $\beta$  when  $a = 1$ ,  $b = 2$ ,  $c = 2.25$ ,  $d = 1.5$ , and  $k = 1$ .

Parameter	ML	BSELF	BGELF	BLINEX
$\alpha$	0.4354817	0.4398853	0.4218823	0.4277962
$\beta$	2.121879	2.156254	2.049693	2.047704

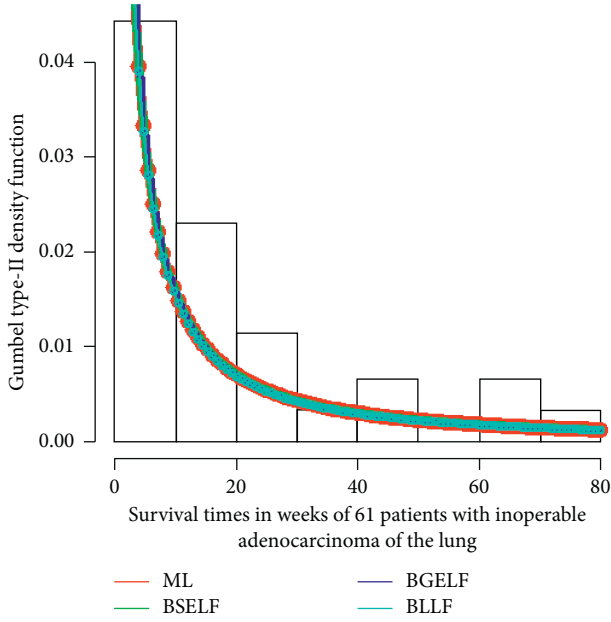


FIGURE 2: Comparison of estimation methods.

the lung as the Gumbel type-II density function is taken on the  $y$ -axis.

## 6. Conclusion and Recommendations

In medical decision-making, Bayesian tools incorporate the state of uncertainty and provide a rational framework for studying such problems. Usually, medical data are generally skewed to the right, and positively skewed distributions can be most suitable for describing unimodal medical data. In this study, an attempt has been made to develop the Bayesian estimators for Gumbel type-II distribution based on type-II censored data using squared error loss, GELF, and LINEX loss functions via Lindley's approximation. It is concluded that ML and Bayesian estimators become closer by increasing the sample sizes and prespecified percentages of failures. Based on the outcomes of this research study, we may suggest that this study can be further extended by using other skewed distributions considering the Bayesian framework with other loss functions using medical data.

## Appendix

### Observed Fisher Information Matrix

The observed Fisher information matrix (FIM) is computed by taking the 2<sup>nd</sup> partial derivatives with respect to " $\alpha$ " and " $\beta$ ," respectively. Therefore, the matrix may be defined as

$$I_{(\hat{\alpha}, \hat{\beta})} = \begin{bmatrix} -\frac{\partial^2 \ln L}{\partial \alpha^2} & -\frac{\partial^2 \ln L}{\partial \alpha \partial \beta} \\ -\frac{\partial^2 \ln L}{\partial \beta \partial \alpha} & -\frac{\partial^2 \ln L}{\partial \beta^2} \end{bmatrix}. \quad (\text{A.1})$$

The components of observed FIM are

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \alpha^2} &= -\frac{r}{\alpha^2} - \beta \sum_{i=1}^r x_i^{-\alpha} (\ln x_i)^2 - (n-r)\beta \ln x_r \\ &\quad \cdot (1-A)x_r^{-\alpha} \ln x_r \left[ \frac{x_r^{-\alpha} \beta - A}{A^2} \right], \\ &\quad \cdot \frac{r}{\alpha^2} + \beta \sum_{i=1}^r x_i^{-\alpha} (\ln x_i)^2 + (n-r)\beta \ln x_r \\ &\quad \cdot (1-A)x_r^{-\alpha} \ln x_r \left[ \frac{x_r^{-\alpha} \beta - A}{A^2} \right] = U, \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \beta^2} &= -\frac{r}{\beta^2} - (n-r)x_r^{-2\alpha} \left[ \frac{(1-A)}{A^2} \right], \\ &\quad \cdot \frac{r}{\beta^2} + (n-r)x_r^{-2\alpha} \left[ \frac{(1-A)}{A^2} \right] = V, \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} &= \sum_{i=1}^r x_i^{-\alpha} \ln x_i + (n-r) \left[ \frac{(1-A)x_r^{-\alpha} \ln x_r (\beta x_r^{-\alpha} - A)}{A^2} \right], \\ &\quad - \sum_{i=1}^r x_i^{-\alpha} \ln x_i - (n-r) \\ &\quad \cdot \left[ \frac{(1-A)x_r^{-\alpha} \ln x_r (\beta x_r^{-\alpha} - A)}{A^2} \right] = W. \end{aligned} \quad (\text{A.4})$$

The observed FIM matrix is rewritten as

$$I_{(\alpha, \beta)} = \begin{bmatrix} U & W \\ W & V \end{bmatrix}. \quad (\text{A.5})$$

The inverse of  $I_{(\alpha, \beta)}$  is

$$\begin{aligned} I_{(\alpha, \beta)}^{-1} &= \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \\ S_{11} &= \frac{V}{UV - W^2}, \\ S_{12} &= -\frac{W}{UV - W^2}, \\ S_{21} &= -\frac{W}{UV - W^2}, \\ S_{22} &= \frac{U}{UV - W^2}, \end{aligned} \quad (\text{A.6})$$

where  $l_{ij} = (\partial^{i+j}(\alpha_1, \alpha_2)/\partial\alpha_1^i\alpha_2^j)$ ;  $i, j=0, 1, 2, 3$ .  $q_1 = (\partial\ln\pi(\alpha, \beta)/\partial\alpha)$ ,  $q_2 = (\partial\ln\pi(\alpha, \beta)/\partial\beta)$ ,  $l_{11} = (\partial^2 g(\alpha, \beta)/\partial\alpha^2)$ ,  $l_{12} = (\partial^2 g(\alpha, \beta)/\partial\alpha\partial\beta)$ ,  $l_{21} = (\partial^2 g(\alpha, \beta)/\partial\beta\partial\alpha)$ ,  $l_{22} = (\partial^2 g(\alpha, \beta)/\partial\beta^2)$ ,  $l_1 = (\partial g(\alpha, \beta)/\partial\alpha)$ ,  $l_2 = (\partial g(\alpha, \beta)/\partial\beta)$ ,  $A_{ij} = (l_i S_{ii} + l_j S_{ij})$ ,  $B_{ij} = 3l_i S_{ii} S_{ij} + l_j (S_{ii} S_{jj} + 2S_{ij}^2)$ , and  $C_{ij} = l_i S_{ii} + l_j S_{jj}$ ,  $i, j=1, 2$ .

$$\sum_{i=1}^2 \sum_{j=1}^2 l_{ij} S_{ij} = l_{11} S_{11} + l_{12} S_{12} + l_{21} S_{21} + l_{22} S_{22}, \quad (\text{A.7})$$

$$l_1 = 1, l_2 = l_{11} = l_{12} = l_{21} = l_{22} = 0,$$

$$\sum_{i=1}^2 \sum_{j=1}^2 l_{ij} S_{ij} = 0, \quad (\text{A.8})$$

$$\begin{aligned} A_{ij} &= (l_i S_{ii} + l_j S_{ij}) S_{ii}, \quad A_{12} = (l_1 S_{11} + l_2 S_{12}) S_{11}, \quad A_{12} = S_{11}^2, \quad A_{21} = (l_2 S_{22} + l_1 S_{21}) S_{22}, \\ A_{21} &= S_{21} S_{22}, \quad B_{ij} = 3l_i S_{ii} S_{ij} + l_j (S_{ii} S_{jj} + 2S_{ij}^2), \quad B_{12} = 3l_1 S_{11} S_{12} + l_2 (S_{11} S_{22} + 2S_{12}^2), \\ B_{12} &= 3S_{11} S_{12}, \quad B_{21} = 3l_2 S_{22} S_{21} + l_1 (S_{22} S_{11} + 2S_{21}^2), \quad B_{21} = (S_{22} S_{11} + 2S_{21}^2), \\ C_{ij} &= l_i S_{ii} + l_j S_{jj}, \quad C_{12} = l_1 S_{11} + l_2 S_{21}, \quad C_{12} = S_{11}, \quad C_{21} = l_2 S_{22} + l_1 S_{12}, \\ C_{21} &= S_{12}, \quad q_1 = (((a-1)/\hat{\alpha}) - b), \quad q_2 = (((c-1)/\hat{\beta}) - d), \end{aligned}$$

$$\begin{aligned} \frac{\partial^3 \ln L}{\partial \alpha^3} &= \frac{2r}{\alpha^3} + \beta \sum_{i=1}^r x_i^{-\alpha} (\ln x_i)^3 - (n-r)\beta (\ln x_r)^2 \left[ \frac{(1-A)}{A^2} x_r^{-\alpha} (-\beta x_r^{-\alpha} \ln x_r + \exp(-\beta x_r^{-\alpha}) (\beta x_r^{-\alpha} \ln x_r)) \right. \\ &\quad + \frac{(1-A)}{A^2} (\beta x_r^{-\alpha} - A) (-x_r^{-\alpha} \ln x_r) \\ &\quad \left. + \frac{x_r^{-\alpha} (\beta x_r^{-\alpha} - A)}{(1 - \exp(-\beta x_r^{-\alpha}))^4} ((1 - \exp(-\beta x_r^{-\alpha}))^2 \exp(-\beta x_r^{-\alpha}) \beta x_r^{-\alpha} \ln x_r + 2(\exp(-\beta x_r^{-\alpha}))^2 (1 - \exp(-\beta x_r^{-\alpha})) \beta x_r^{-\alpha} \ln x_r) \right], \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} \frac{\partial^3 \ln L}{\partial \beta^3} &= \frac{2r}{\beta^3} - \frac{(n-r)x_r^{-2\alpha}}{(1 - \exp(-\beta x_r^{-\alpha}))^4} \\ &\quad \left[ (1 - \exp(-\beta x_r^{-\alpha}))^2 \exp(-\beta x_r^{-\alpha}) (-x_r^{-\alpha}) - (\exp(-\beta x_r^{-\alpha}))^2 2(1 - \exp(-\beta x_r^{-\alpha})) (x_r^{-\alpha}) \right], \end{aligned} \quad (\text{A.10})$$

$$\begin{aligned} \frac{\partial^3 \ln L}{\partial \alpha^2 \partial \beta} &= - \sum_{i=1}^r x_i^{-\alpha} (\ln x_i)^2 - (n-r)x_r^{-\alpha} (\ln x_r)^2 \\ &\quad \left[ \frac{(1-A)(\beta x_r^{-\alpha} - A)}{A^2} + \frac{(1-A)A\beta x_r^{-\alpha}}{A^2} + \frac{\beta(\beta x_r^{-\alpha} - A)}{A^4} A(1-A)(x_r^{-\alpha})(-A - 2(1-A)) \right], \end{aligned} \quad (\text{A.11})$$

$$\begin{aligned} \frac{\partial^3 \ln L}{\partial \alpha \partial \beta^2} &= -(n-r) \left[ \frac{-2(1-A)x_r^{-2\alpha} \ln x_r}{A^2} + \frac{x_r^{-2\alpha}}{(1 - \exp(-\beta x_r^{-\alpha}))^4} ((1 - \exp(-\beta x_r^{-\alpha}))^2 \exp(-\beta x_r^{-\alpha}) (\beta x_r^{-\alpha} \ln x_r) \right. \\ &\quad \left. + 2(\exp(-\beta x_r^{-\alpha}))^2 (1 - \exp(-\beta x_r^{-\alpha})) \beta x_r^{-\alpha} \ln x_r) \right]. \end{aligned} \quad (\text{A.12})$$

## Data Availability

This work is mainly a methodological development and has been applied on secondary data, but if required, data will be provided.

## Conflicts of Interest

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

## References

- [1] E. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, USA, 1958.
- [2] K. Abbas and Y. Tang, "Estimation of parameters for Frechet distribution based on type-II censored samples," *Caspian Journal of Applied Sciences Research*, vol. 2, no. 7, pp. 36–43, 2013.
- [3] H. M. Okasha, "E-Bayesian estimation for the Lomax distribution based on type-II censored data," *Journal of the Egyptian Mathematical Society*, vol. 22, no. 3, pp. 489–495, 2014.

- [4] H. H. Abu-Zinadah, "Bayesian estimation on the exponentiated Gompertz distribution under type II censoring," *International Journal of Contemporary Mathematical Sciences*, vol. 9, no. 11, pp. 497–505, 2014.
- [5] R. M. El-Sagheer, "Bayesian prediction based on general progressive censored data from generalized Pareto distribution," *Journal of Statistics Applications & Probability*, vol. 5, no. 1, pp. 43–51, 2016.
- [6] K. Abbas, J. Fu, and Y. Tang, "Bayesian estimation of Gumbel type-II distribution," *Data Science Journal*, vol. 12, pp. 33–46, 2013.
- [7] N. Feroze and M. Aslam, "Bayesian estimation of two-component mixture of gumbel type II distribution under informative priors," *International Journal of Advanced Science and Technology*, vol. 53, pp. 11–30, 2013.
- [8] I. Malinowska and D. Szynal, "On characterization of certain distributions of  $k^{\text{th}}$  lower (upper) record values," *Applied Mathematics and Computation*, vol. 202, no. 1, pp. 338–347, 2008.
- [9] T. Sultana, M. Aslam, and M. Raftab, "Bayesian estimation of 3-component mixture of Gumbel type-II distributions under non-informative and informative priors," *Journal of the National Science Foundation of Sri Lanka*, vol. 45, no. 3, pp. 287–306, 2017.
- [10] F. Metiri, H. Zeghdoudi, and M. R. Remita, "On Bayes estimates of Lindley distribution under Linex loss function: informative and non informative priors," *Global Journal of Pure and Applied Mathematics*, vol. 12, no. 1, pp. 391–400, 2016.
- [11] H. Reyad and S. O. Ahmed, "E-Bayesian analysis of the Gumbel type-II distribution under type-II censored scheme," *International Journal of Advanced Mathematical Sciences*, vol. 3, no. 2, pp. 108–120, 2015.
- [12] V. Preda, E. Panaitescu, and A. Constantinescu, "Bayes estimators of modified- weibull distribution parameters using Lindleys approximation," *WSEAS Transactions on Mathematics*, vol. 9, no. 7, pp. 539–549, 2010.
- [13] A. Legendre, *New Method for the Dermination of Orbits of Comets*, Courcier, Paris, France, 1805.
- [14] C. F. Gauss, *Least Squares Method for the Combinations of Observation*, (Translated by J. Bertrand 1955), Mallet-Bachelier, Paris, France, 1810.
- [15] H. R. Varian, *A Bayesian Approach to Real Estate Assessment*, pp. 195–208, North Holland, Amsterdam, Netherlands, 1975.
- [16] R. Calabria and G. Pulcini, "Point estimation under asymmetric loss functions for left-truncated exponential samples," *Communications in Statistics—Theory and Methods*, vol. 25, no. 3, pp. 585–600, 1996.
- [17] D. V. Lindley, "Approximate bayesian methods," *Trabajos de Estadística Y de Investigacion Operativa*, vol. 31, no. 1, pp. 223–245, 1980.
- [18] E. T. Lee and J. W. Wang, Eds., *Statistical Methods for Survival Data Analysis*, John Wiley and Sons, New York, NY, USA, 3rd edition, 2003.
- [19] S. W. Lagakos and J. S. Williams, "Models for censored survival analysis: a cone class of variable-sum models," *Biometrika*, vol. 65, pp. 181–189, 1978.
- [20] S. Lee and R. A. Wolfe, "A simple test for independent censoring under the proportional hazards model," *Biometrics*, vol. 54, pp. 1176–1182, 1998.

## Research Article

# Influential Factors of an Asynchronous BCI for Movement Intention Detection

Sura Rodpongpun , Thapanan Janyalikit, and Chotirat Ann Ratanamahatana 

*Department of Computer Engineering, Chulalongkorn University, Pathumwan, Bangkok 10330, Thailand*

Correspondence should be addressed to Chotirat Ann Ratanamahatana; [chotirat.r@chula.ac.th](mailto:chotirat.r@chula.ac.th)

Received 5 December 2019; Revised 2 February 2020; Accepted 10 February 2020; Published 23 March 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Sura Rodpongpun 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, asynchronous brain computer interface (BCI) systems have been utilized in many domains such as robot controlling, assistive technology, and rehabilitation. In such BCI systems, movement intention detection algorithms are used to detect movement desires. In recent years, movement-related cortical potential (MRCP), an electroencephalogram (EEG) pattern representing voluntary movement intention, attracts wide attention in movement intention detection. Unfortunately, low MRCP detection accuracy makes the asynchronous BCI system impractical for real usage. In order to develop an effective MRCP detection algorithm, EEG data have to be properly preprocessed. In this work, we investigate the relationship and effects of three factors including frequency bands, spatial filters, and classifiers on MRCP classification performance to determine best settings. In particular, we performed a systematic performance investigation on combinations of five frequency bands, five spatial filters, and six classifiers. The EEG data were acquired from subjects performing series of self-paced ankle dorsiflexions. Analysis of variance (ANOVA) statistical test was performed on F1 scores to investigate effects of these three factors. The results show that frequency bands and spatial filters depend on each other. The combinations directly affect the F1 scores, so they have to be chosen carefully. The results can be used as guidelines for BCI researchers to effectively design a preprocessing method for an advanced asynchronous BCI system, which can assist the stroke rehabilitation.

## 1. Introduction

A brain computer interface (BCI) system is a system that translates human minds to control signals for external devices. These external devices include simple feedback systems and more complex devices such as prosthetic organs [1]. Moreover, a BCI system can be used for communication between persons. It transmits one's messages as brain activity signals to other persons through a system that can decode the signals to human recognizable messages. In this sense, patients suffering from a brain disease including locked-in syndrome (LIS) or completely locked-in state (CLIS), as they cannot move, talk, or even blink, can get benefits from the BCI system [2]. Therefore, large amount of research effort has been devoted to solving communication problem for these patients [1, 3–5].

In the past decade, there has been evidence based on the Hebbian theory [6] showing that motor function loss can be

restored by brain plasticity induction through the BCI system for rehabilitation [7]. This phenomenon can be a great advantage for stroke patients, whose brain regions were partially destroyed by blood clots in the brain vessel. This situation is a cause of mobility and speech function impairment that affects the patients' daily life. After some treatments and a long-term rehabilitation program, some patients can partially or even fully recover their motor functions; it depends on how long the symptom lasts, how much the brain region is destroyed, how well the brain plasticity develops, and even how good the rehabilitation program is.

BCI systems can be categorized by their functions into two types: synchronous and asynchronous. For the synchronous BCI system, aka a cue-based BCI system, users have to perform specific task in response to given cues while using the system. This scenario often causes discomfort and fatigue. For this reason, asynchronous BCI systems have



been proposed since the last decade [1]. The asynchronous BCI system, aka a self-paced BCI system, allows its users to perform tasks upon their desire. In this sense, the asynchronous BCI system overcomes the ordinary synchronous BCI system by providing more comfort and causing less fatigue. These properties are very important to the BCI system for rehabilitation because the users have to interact with the system for a long period. However, designing an asynchronous BCI system is much more challenging due to the fact that the acquired signals are noisier because users can pay less concentration to use the system.

Recently, the asynchronous BCI system has been used in stroke rehabilitation for the first time to induce brain plasticity [8]. The asynchronous BCI system is applied in stroke rehabilitation by using brain signals measured from the scalp, known as electroencephalogram (EEG), to drive a rehabilitation tool. Particularly, movement intention is detected from EEG data and translated to a control signal to instantaneously switch on an electrical stimulator. Upon completion of some rehabilitation sessions, stroke impact scale (SIS) appears to improve in all patients, demonstrating that impacts on stroke patients could be alleviated. The key of these improvements is the preciseness of the electrical stimulator activation when patients start to move or start to imagine about moving the affected limb [9]. Although there have been evidences showing that the recent BCI systems for stroke rehabilitation can induce brain plasticity in stroke patients [8–10], its limitation in terms of movement intention detection accuracy makes it not widely used.

To detect the movement intention, there are two types of signals that are usually utilized: movement-related cortical potential (MRCP) [11] and event-related desynchronization/synchronization (ERD/ERS). These two come from different domains; MRCP is extracted from a time domain as a pattern of time series signal while ERD/ERS is extracted from a frequency domain. In terms of movement intention detection accuracy, MRCP and ERD/ERS provide comparable detection accuracy. However, it has been recently shown that MRCP is superior in terms of detection latency and therefore is preferred to ERD/ERS for movement intention detection [12, 13].

Over the past decade, many algorithms have been proposed to detect the movement intention before real movement execution [14–18]; however, it is still unclear about how to filter or clean the acquired signals before feeding them to a classifier. For example, Niazi et al. [16] and Lin et al. [17] proposed methods utilizing matched filter (MF) and locality sensitivity discriminant analysis (LSDA) to detect movement intentions from MRCP, but they suffered from difficult choices in selecting both an appropriate frequency band and a spatial filtering technique. Instead of finding an appropriate solution, the choices were made arbitrarily.

Although some researchers were aware of these problems and attempted to explore an appropriate configuration or combination of computation methods for the movement intention or the movement imagination detection in BCI system [14, 19–24], none of them explored these problems thoroughly. The most systematic study was in [14] where

combinations among three factors of spatial filter, temporal filter, and classifier were analyzed. However, the work only focuses on movement intention classification between left and right hand, not on movement intention detection. Moreover, not only MRCP or EEG data in time domain but also ERD/ERS data were used as it can be an obstacle for a real-time usage. In [24], the authors studied the effects of frequency and spatial filters on the contingent negative variation (CNV), a cue-based version of MRCP. Although their experiments were very systematic and concerned many aspects, they did not focus on self-paced experiments. In addition, effects of classifiers were also not mentioned in their work except for linear discrimination analysis (LDA). Furthermore, their results are limited to only offline usage due to the delay of the frequency filter. For other works, experiments were done under some constraints (e.g., a preset classifier, a preset frequency band, or a preset spatial filter). These situations could cause problems; for example, changing one factor may affect other factors as will be shown in our study. Also, the data were not acquired in a self-paced manner, nor the detection cannot be employed in real time [19], which makes the result unusable for movement intention detection problem.

In this paper, we make the first attempt to study effects and relationship among the 3 factors—frequency bands, spatial filtering techniques, and classifiers, specifically on EEG data acquired in a self-paced manner for movement intention detection. In detail, we recorded 19-channel EEG data from 9 subjects while performing a series of self-paced ankle dorsiflexions. After labeling and extracting the data, we employed frequency filters in the range of 0.01–5 Hz, where MRCP can be observed [25, 26]. Five well-known spatial filtering techniques and six classifiers were used. The five spatial filtering techniques are (1) no spatial filtering (NoF), (2) surface Laplacian (SL), (3) independent component analysis (ICA), (4) common spatial pattern analysis (CSP), and (5) principle component analysis (PCA). The six classifiers include (1) linear discrimination analysis (LDA), (2) support vector machine (SVM), (3) one-nearest neighbor utilizing Euclidean distance (1-NN-ED), (4) one-nearest neighbor utilizing dynamic time warping distance (1-NN-DTW), (5) shape-based template matching (TM), and (6) matched filter (MF). To emphasize, this study mainly focuses on finding relationship and effects of various factors and on determining good combinations among them. Our results provide valuable insights towards a development in an effective movement intention detection algorithm for asynchronous BCI rehabilitation systems.

Additionally, a novel shape-based template matching algorithm is added to our study to investigate a possibility of utilizing a shape of MRCP via time series mining techniques in BCI applications. Our study provides insight that can be used to help improve performance and efficiency of movement intention detection in asynchronous BCI systems.

Contributions and impact of this work can be summarized as follows:

- (i) This work investigates relationship and effects of the three critical factors in asynchronous BCI systems

for stroke rehabilitation. The results demonstrate that these factors are quite sensitive. Different combinations of the factors can substantially influence the system's performance.

- (ii) This work can be used as primitive guidelines to process self-paced EEG data in an asynchronous BCI system, which can assist the stroke rehabilitation.
- (iii) This work explores a possibility of utilizing shapes of time series signals to detect MRCP from self-paced EEG data using unconventional time series mining techniques.

## 2. Methodology

**2.1. Subjects.** Nine subjects (seven males and two females; age ranges from 22 to 26 years) participated in the experiment. None of the subjects had any known neurological disorders, nor had experiences with any BCI systems prior to the experiment. All subjects gave their signed informed consents for the experiments, and the experiment protocol was approved by the research ethics review committee for research involving human research participants, Health Science Group, Chulalongkorn University (COA No. 049/2018).

**2.2. Experimental Protocol and Data Acquisition.** At the beginning of the recording session, each participant was asked to sit on a chair in a comfortable position with both legs rested on the ground. After that, 19 channels of monopolar EEG were collected from each subject in the 10–20 system using Electro-Caps (Electro-Caps, Electro-Cap International Inc.) and a Nicolet w10-20HB amplifier (Natus Medical Inc.). The electrodes were located at Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2. The ground electrode was the ground node of the cap that is in the middle among Fp1, Fp2, and Fz, and the reference was placed on the left earlobe. With the use of electrolyte gel, the impedance of all electrodes was calibrated to be less than the 5k-Ohm threshold before starting the data acquisition. One channel of surface electromyography (EMG) was also recorded by a Nicolet w10-20HB amplifier (with disposable electrodes) for the purpose of EEG signal labeling. EMG was recorded from bipolar derivation from the tibialis anterior (TA) muscle and on the bony surface of the knee of the dominant leg (right knees in all subjects). All of the EEG and EMG signals were sampled at 1024 Hz. EEG signals were filtered by a Butterworth band-pass filter with a frequency band of 0.01 to 30 Hz and a notch filter to filter out 49–51 Hz power-line distortion [27]. In the recording sessions, subjects were instructed to perform self-paced ballistic ankle dorsiflexions. The duration between consecutive trials was roughly 3 to 7 seconds. Roughly, there are about 280,000 data points per channel per session that need labeling. During the recording session to reduce artifacts, each subject was asked to stay relaxed, closing the eye lids, and not to move other body parts. The protocol is shown in Figure 1. Each subject took about six recording sessions with resting

periods of two to ten minutes in between. During the recording sessions, videos were recorded for EMG signal synchronization. The recording was made in an environment similar to a hospital setting.

**2.3. Data Labeling.** After the EEG data acquisition, each period of the movement execution had to be labeled as a baseline for classification. To detect the onset and offset of each movement initiation and termination, respectively, EMG data were utilized in this process; EMG was visually inspected to discover movement onsets and offsets. Only trials whose EMG movement onset were at least 4 seconds apart from the previous offset were considered. Recorded raw data and their label are made publicly available as stated in the Data Availability section so that the experiment in this work can be reproduced.

**2.4. Data Segmentation.** The data were manually segmented (into 2-second sequences) for classification. For movement intention segments, we extracted a sequence starting before each actual movement and ending after the movement. In particular, each potential MRCP segment was extracted by considering a 2-second duration before each movement onset, and each non-MRCP segment was extracted by considering a 2-second duration right after each movement offset. After segmentation, there are 1,674 MRCP segments and 1,544 non-MRCP segments in total. All of the acquired segmented data were  $z$ -normalized. Further explanations and examples of segmented data can be downloaded via the link described in the Data Availability section.

**2.5. Spatial Filtering Techniques.** Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activities, i.e., voltage fluctuations within the neurons of the brain. However, these voltage fluctuations are so small ( $\mu V$ ) that the EEG signal is easily contaminated by noise or even electric fields from other brain regions, resulting in a low signal-to-noise ratio of the acquired EEG. Many spatial filtering techniques have been introduced to accentuate localized EEG activities to maximize the signal-to-noise ratio. In this section, we describe some details of the widely used spatial filtering techniques that have been applied in our experiments to reduce noises or to enhance the quality of EEG signals.

**2.5.1. Surface Laplacian.** SL is a technique used to reduce contamination in an electrode caused by other surrounding electrodes. SL is calculated by subtracting the weighted voltage of the surrounding electrodes from the working electrode. The weight is usually the distance from the electrode of interest to each of the surrounding electrodes. According to [28], the formula of SL is shown in the following equation:

$$V_{\text{surrogate}} = V_{\text{main}} - \sum_{j \in \text{NN}} \frac{1/d_j}{\sum_{j \in \text{NN}} (1/d_j)} V_j, \quad (1)$$

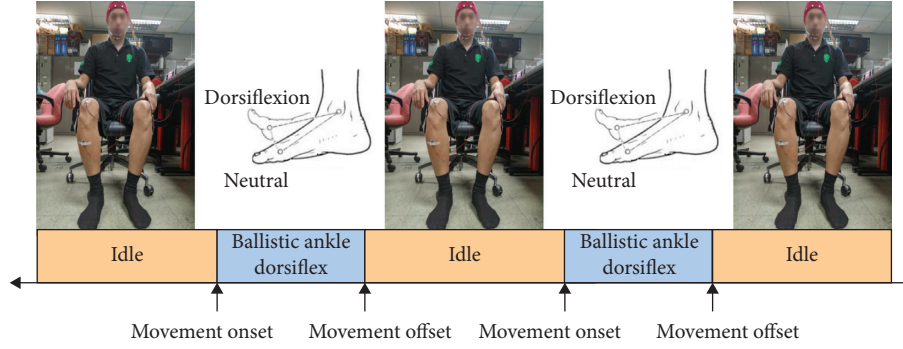


FIGURE 1: The experimental protocol shows the task of ballistic ankle dorsiflexion sequence followed by idle states. Idle states are as long as the user desires. These idle states normally take about 3 to 7 seconds.

where  $V_{\text{surrogate}}$  is the output voltage of SL,  $V_{\text{main}}$  is the voltage between the electrode of interest and a reference electrode, NN is a set of surrounding electrodes,  $d_j$  is the distance between  $j$  surrounding electrodes and the electrode of interest, and  $V_j$  is the voltage between the electrode  $j$  and the reference electrode. In our experiments,  $d_j$  was set to 1 for all  $j \in \text{NN}$  and only nine channels, i.e., F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4, were used as working electrodes as these electrodes have surrounding electrodes.

**2.5.2. Independent Component Analysis.** ICA was invented to deal with problems similar to the cocktail-party problem. It can be used to separate a mixed signal into independent signals. In EEG domain, EEG signals are simultaneously acquired from many channels, thus a signal from one channel can be contaminated by the others. In this case, ICA is used to separate acquired signals into independent signals. According to [29], assume that we acquired  $n$  linear mixed signals from  $n$  channels  $x_1, \dots, x_n$ . Each signal  $x_i$  is originated from  $n$  independent components as illustrated in the following equation:

$$x_1 = a_1s_1 + a_2s_2 + \dots + a_ns_n. \quad (2)$$

We can rewrite equation (2) in terms of vectors and matrix as shown in the following equation:

$$x = As, \quad (3)$$

where  $x$  is a column vector form of  $n$  EEG channels,  $s$  is a column vector form of  $n$  independent components, and  $A$  is a matrix of size  $n \times n$ . Instead of finding  $x$ , we want to find the inverse of  $A$ , denoted by  $W$ , to obtain the independent component  $s$ . A well-known algorithm for ICA is fastICA [29]. Instead of rejecting noisy components and reconstructing the original data as performed in [27], we used  $W$  to form independent components as new representations and feed them to the classifiers.

**2.5.3. Common Spatial Pattern.** CSP is a method used for constructing a new representation from a high dimensional data to a lower dimensional data whose variance is maximized between two classes of data. In EEG domain, inputs of CSP are two classes of multichannel EEG data, i.e., MRCP

and non-MRCP data. CSP then calculates a projection matrix that is used to project multichannel data into a low dimensional spatial subspace by a linear transformation [30]. The first and the last row of the projection matrix are most suitable for discriminating two different classes. They provide a transformation that maximizes a variance in one class and minimizes a variance in another. In this work, the first row of the projection matrix was utilized as a weight for multichannel EEG to make a new representation as performed in [16].

**2.5.4. Principle Component Analysis.** PCA has been used as a dimensionality reduction technique for most data miners. However, PCA can also be used in data compression as it can bring out strong patterns from complex data while preserving variability. PCA projects the original data onto a new set of axes via its eigenvector, which is called principle component. Apart from the dimensionality reduction, PCA can also be used to decompose original data into a new set of decorrelated data by projecting the data onto orthogonal axes. Details of PCA, its principle, calculation, and application are well described in [31, 32]. In this work, we decomposed multichannel EEG data into independent components and then used them similarly to ICA.

## 2.6. Classifiers

**2.6.1. Linear Discriminant Analysis.** LDA is one of the most widely used classification algorithms in BCI systems due to its simplicity and efficiency. It is based on the use of hyperplanes to separate different classes of data. A hyperplane can be determined by minimizing the intraclass variance and maximizing the distance between the means of two classes of data.

**2.6.2. Support Vector Machine.** SVM [33] is based on an idea of hyperplane construction, which provides a maximum margin that can be used to separate the data. If the data are linearly separable in the input space, a hyperplane is constructed from the input data. Otherwise, a mapping of features to a higher dimensional space is done via a kernel function before a hyperplane construction. In this work, we used the *svm* function in MATLAB and performed



optimization on all parameters using grid search with default settings.

**2.6.3. One-Nearest Neighbor.** 1-NN is a widely used classifier in a time series domain. A distance measure is used to measure similarity between two time series, in this case, two epochs of EEG data. A simple distance measure is Euclidean distance (ED), and a more complex one is dynamic time warping (DTW). The main difference between these two measures is that DTW has a nonlinear alignment ability, which makes DTW achieve higher accuracy than ED. The nonlinear alignment of DTW can be restricted to prevent unreasonable warping through the use of a global constraint, which in turn further increases the accuracy. In this work, 1-NN was utilized, while both ED and DTW were used as distance measures. For DTW, we used the Sakoe–Chiba band, which is one of the widely used global constraint, with  $r = 10\%$ . More information about DTW and the global constraints can be found in [34].

**2.6.4. Matched Filter.** MF is a technique to extract a template, which is a known signal, in an unknown signal. MF can be calculated by convolving the unknown signal with a time-reversed template. In this work, an average of MRCP data was used to create a template.

**2.6.5. Template Matching.** TM is similar to MF but different in template construction and similarity measurement. In time series mining, template construction can also be done through shape-based averaging. One of the most well-known shape-based averaging techniques for time series data is dynamic time warping barycenter averaging (DBA) [35], which had shown to outperform other existing averaging methods [35–37]. For similarity measurement, TM utilizes DTW as a similarity measure, while MF uses a convolution technique. In this work, we used DBA with a medoid of MRCP data as an initial template. The global constraint was set to 10% [38] for preventing unreasonable warp.

**2.7. Experiment Setup.** In the experiment, we started by filtering the acquired EEG data with a causal second-order Butterworth band-pass filter with frequency bands of [0.01–1], [1–2], [2–3], [3–4], and [4–5] Hz, followed by data segmentation. However, we set the order of Butterworth filter to 2<sup>nd</sup> order and did not take other filter orders into account in this experiment due to an unstable problem of IIR filters as discussed in [20]. It is worth noting that there are infinitely many possible ways to extract frequency bands in the range of [0.01–5] Hz, where MRCP can be observed; thus, to make it feasible to extract only a prominent band and reject irrelevant ones, we decided to decompose the frequency into a bin of 1 Hz. Afterwards, MRCP and non-MRCP data were shuffled, and then stratified sampling was performed to create training and test sets with the ratio of 2/3 and 1/3, respectively. Overall numbers of MRCP and non-MRCP data for training set and test set of each subject are shown in Table 1. Then, five different spatial filtering techniques

TABLE 1: Number of samples in MRCP class vs. non-MRCP class of all participants.

Subject	Training set		Test set	
	MRCP	Non-MRCP	MRCP	Non-MRCP
1	122	108	62	55
2	120	113	61	57
3	124	106	63	54
4	118	111	60	57
5	116	103	59	53
6	99	90	51	46
7	142	135	72	69
8	140	135	71	69
9	129	121	65	62

including no spatial filtering (NoF) were performed. For ICA and PCA, which require continuous data, MRCP and non-MRCP data of each training set were concatenated to form continuous data. To classify the MRCP data, six different classifying methods were used. To summarize, there were 5 frequency bands, 19 channels without spatial filter, 9 channels for SL, 19 components for ICA and PCA, 1 channel for CSP, and 6 classifying methods; thus, there were  $5 \times (19 + 9 + 19 + 19 + 1) \times 6 = 2,010$  combinations in total for each individual. By aiming to achieve high detection rate while providing low false alarm rate, we evaluated the results by measuring F1 scores. F1 scores are calculated by taking both precision and recall into account to maintain balance between them. Precision is a number of correctly predicted MRCP samples divided by a number of total samples predicted as MRCP. Recall is a number of correctly predicted MRCP samples divided by a number of total MRCP samples. F1 scores are calculated as shown in the following equation:

$$F1 \text{ score} = \left( \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \right). \quad (4)$$

The best channel and component of each spatial technique and the threshold of the classifiers were selected from the best one that achieves the highest F1 score.

We employed three-way repeated measures analysis of variance (ANOVA) with Greenhouse–Geisser adjustment, which is used to adjust the degree of freedom, to investigate the effects of 3 independent variables, i.e., 5 spatial filters, 5 frequency bands, and 6 classifiers, on F1 scores. The total number of repeated measurements was  $5 \times 5 \times 6 = 150$ . By employing ANOVA, we can investigate whether the variation of factors lead to significant difference of experimental results. In other words, the ANOVA can be used as an evidence to reject or accept the null hypothesis, which means the mean values of tested groups are the same. If the null hypothesis is rejected, the mean values of tested groups are considered different, and we can imply that the factors affect the results. On the other hand, if the null hypothesis is accepted, the mean values of tested groups can be considered similar, and the factors do not affect the results. The post hoc test used the Bonferroni adjustment to compensate for multiple comparisons. For the ANOVA test and the post hoc test, a significance level of  $P < 0.01$  was adopted in this work. Due to the fact that our sample size is

not considered very large, we decided to set the significance level of  $P < 0.01$  instead of  $P < 0.1$  or  $P < 0.05$  as used in many previous works [8, 14, 16] for compensation and giving more confidence. The statistical analysis was performed by IBM SPSS Statistics 22.0.

### 3. Results

**3.1. Three-Way Repeated Measures ANOVA.** The output of three-way repeated measures ANOVA is shown in Table 2. There is a significant interaction effect between spatial filter and frequency,  $SF * Freq \cdot (F(3.659, 29.275) = 10.811, P < 0.01)$ . None of other interaction effects are significant. This means the effect of frequency and the effect of spatial filter are dependent in some circumstances, which will also be explored in this work. There is also a significant effect of the frequency,  $Freq(F(2.167, 17.339) = 21.384, P < 0.01)$ , and the classifier,  $Classifier(F(1.24, 9.921) = 13.102, P = 0.003)$ . No other significant effects are found.

To visualize the three-way interaction, we illustrate multiple comparison graphs of the three-way interaction in Figure 2. The highest estimated marginal mean of F1 scores can be found by utilizing EEG data in frequency [0.01-1] Hz with SL, regardless of the choice of classifier. For the classifier, also shown in Figure 2, LDA and SVM provide comparable F1 scores; the rest provide relatively lower F1 scores than LDA and SVM, while providing comparable F1 scores to each other.

The results in Figure 2 also show that these factors are quite sensitive. Choosing different combinations of the factors can substantially affect the F1 scores. The scores could range from as low as 44.4% with the choice of [3-4] Hz frequency + CSP filter + MF classifier to as high as 82.3% with the choice of [0.01-1] Hz frequency + SL spatial filter + SVM classifier.

After a post hoc test of the classifier, we found that SVM is the most prominent classifier, being an independent factor without any intervention from other factors. The estimated marginal mean of SVM is 66.4% with a significantly higher F1 score than the others including LDA, as shown in Figure 3.

The estimated marginal mean difference between SVM and LDA is about 4%. Moreover, the multiple comparison results of the classifiers revealed that LDA is the second most prominent classifier providing high F1 scores with a significant difference from the other classifiers. 1-NN-ED, 1-NN-DTW, and TM provide comparable F1 scores among each other, i.e., 59.6%, 58.6%, and 58.1%, respectively. MF is the worst classifier providing the lowest F1 score of 53.7%.

From the three-way repeated measures ANOVA test, the result tells us that frequency has a significant impact while spatial filter has no significant impact. However, since frequency and spatial filter are highly correlated, we need to conduct one-way repeated measures ANOVA to investigate simple effects of spatial filter at every single level of frequency and also simple effects of frequency at every single level of spatial filter.

**3.2. One-Way Repeated Measures ANOVA.** The results of simple effects of spatial filter at every single level of frequency are shown in Table 3. Interestingly, a significant simple effect of spatial filter can be found only in the frequency band of [0.01-1] Hz, SF at [0.01-1] Hz ( $F(3.463, 183.548) = 76.338, P < 0.01$ ). No other significant effect of spatial filter can be found when using other levels of frequency. This result means that a spatial filter can be arbitrarily chosen in any frequency band except in [0.01-1] Hz, which has to be precisely chosen to provide valuable results.

To visualize the effects, we plot a multiple comparison graph for the simple effects of spatial filter at each level of frequency in Figure 4.

In particular, NoF, SL, ICA, CSP, and PCA for EEG data in [0.01-1] Hz frequency provide estimated marginal means of F1 scores of 61%, 78%, 66.5%, 57.4%, and 62.6%, respectively.

After post hoc tests, there is a significant difference when comparing NoF with SL and ICA, but there is no significant difference between CSP and PCA. SL provides the highest F1 score, and it is significantly different from the rest when the frequency is fixed to [0.01-1] Hz. ICA is the second best choice of spatial filter in this bandwidth. It provides the second highest F1 score with a significant difference from other spatial filters. CSP and PCA are also significantly different from each other but they are insignificantly different from NoF.

Next, we provide a further exploration to investigate simple effects of frequency at each type of spatial filters. We therefore conducted one-way repeated measures ANOVA by fixing spatial filter types and adjusting the frequency bandwidth. The output is shown in Table 4. Frequency gives significant effects on NoF, SL, ICA, and PCA:  $Freq$  on NoF ( $F(3.303, 175.053) = 5.536, P = 0.001$ ),  $Freq$  on SL ( $F(3.167, 167.856) = 70.402, P < 0.01$ ),  $Freq$  on ICA ( $F(3.085, 163.490) = 10.265, P < 0.01$ ), and  $Freq$  on PCA ( $F(3.433, 181.937) = 6.094, P < 0.01$ ), respectively. No significant effect of frequency is found on CSP.

After post hoc tests, pairwise comparison of estimated marginal mean of F1 scores reveals that for NoF, only one pair of [0.01-1] and [2-3] Hz frequency bands does have significantly different scores; the rest of the frequency bands are found not to be significantly different from each other. For SL and ICA, the [0.01-1] Hz frequency band provides a prominent estimated marginal mean of F1 score, which made this bandwidth significantly different to all others, while the rest are insignificantly different. For PCA, there are only two pairs of frequency bands that provide significant difference of estimated marginal mean of F1 score; the first pair is [0.01-1] and [3-4] Hz; the second is [0.01-1] and [4-5] Hz. These results are illustrated in Figure 4.

**3.3. Localization of MRCP Detection Performance.** Furthermore, to investigate MRCP localization diagnosis accuracy, we provide a topoplot of F1 score in Figure 5 corresponding to a surrogate channel of SL at frequency [0.01-1] Hz, which is the best combination among these factors for each subject. The channel that provides



TABLE 2: Results of the three-way repeated measures ANOVA test of F1 scores.

Source	Type III sum of squares	Corrected degree of freedom	Mean square	F value	P value	Partial eta squared
SF	0.478	1.273	0.375	7.163	0.018	0.472
Error (SF)	0.534	10.185	0.052			
Freq	0.958	2.167	0.442	21.384	<b>0.000</b>	0.728
Error (freq)	0.359	17.339	0.021			
Classifier	2.135	1.24	1.721	13.102	<b>0.003</b>	0.621
Error (classifier)	1.303	9.921	0.131			
SF * Freq	0.999	3.659	0.273	10.811	<b>0.000</b>	0.575
Error (SF * Freq)	0.739	29.275	0.025			
SF * Classifier	0.172	4.088	0.042	1.661	0.182	0.172
Error (SF * Classifier)	0.827	32.705	0.025			
Freq * Classifier	0.135	4.859	0.028	1.676	0.165	0.173
Error (Freq * Classifier)	0.644	38.870	0.017			
SF * Freq * Classifier	0.397	6.268	0.063	1.299	0.274	0.14
Error (SF*Freq * Classifier)	2.445	50.143	0.049			

P value less than 0.01 is shown in bold, showing significant effects of the parameter to F1 scores. Note that there is a significant interaction effect between spatial filter and frequency; further analysis is needed for these two parameters.

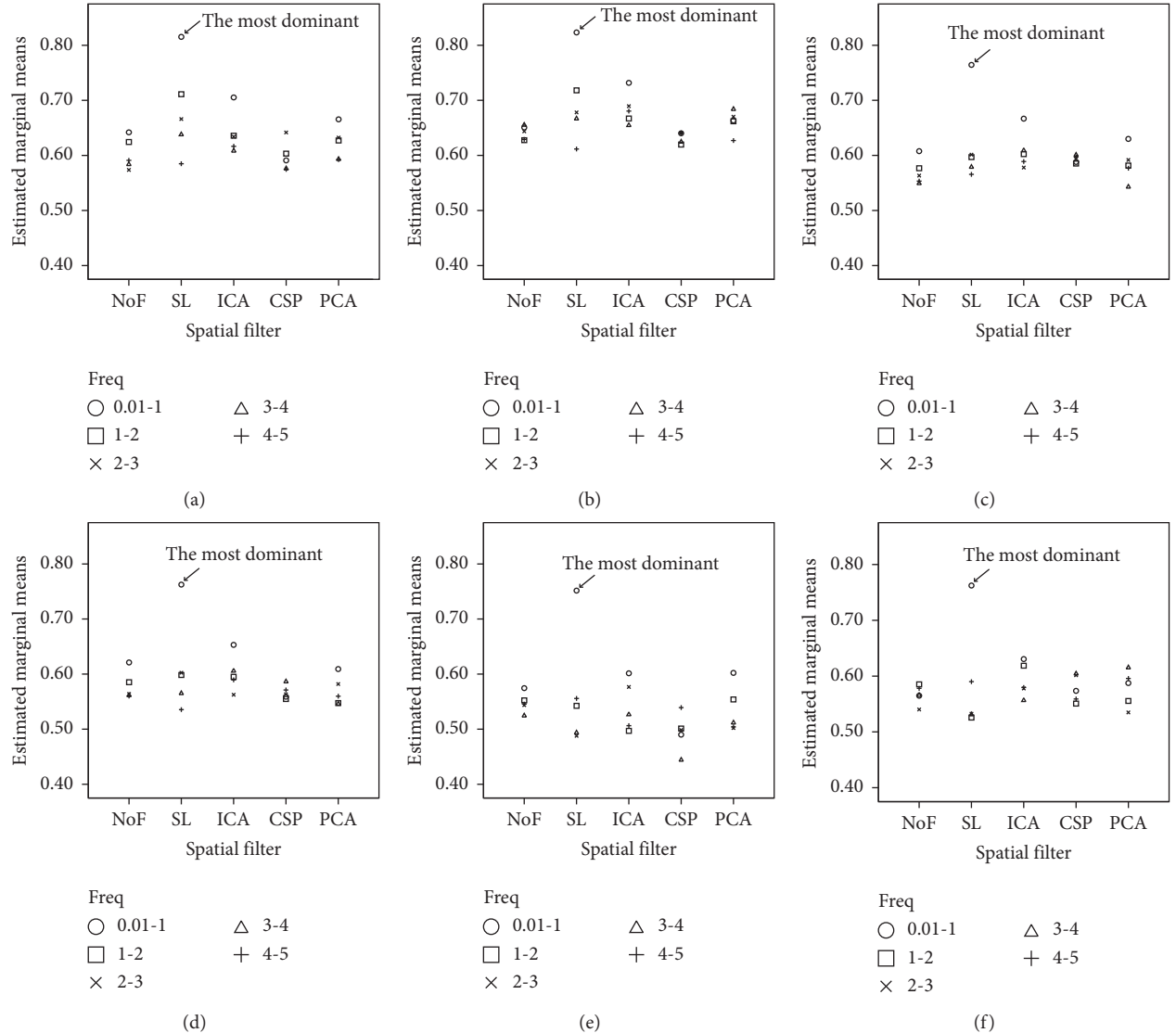


FIGURE 2: Multiple comparison results of the three-way interactions among different spatial filter, frequency, and classifier. The best combination is when classifier = SVM, spatial filter = SL, and frequency = [0.01–1] Hz. (a) Estimated marginal means of F1 score with classifier = LDA. (b) Estimated marginal means of F1 score with classifier = SVM. (c) Estimated marginal means of F1 score with classifier = 1-NN-ED. (d) Estimated marginal means of F1 score with classifier = 1-NN-DTW. (e) Estimated marginal means of F1 score with classifier = MF. (f) Estimated marginal means of F1 score with classifier = TM.

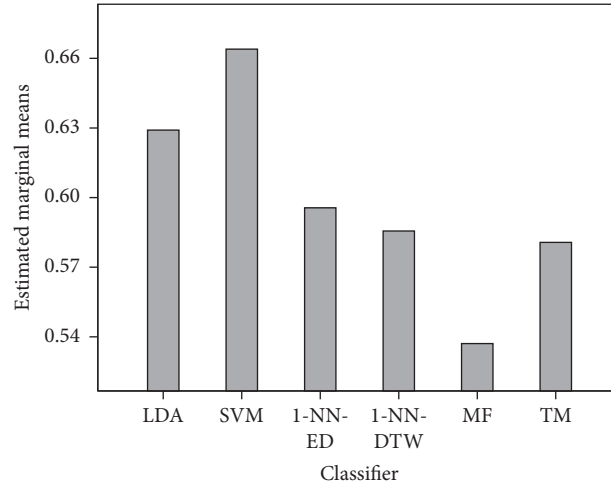


FIGURE 3: Comparison of the estimated marginal means of F1 scores in various classifiers, showing SVM as the dominant classifier.

TABLE 3: Simple effect results of spatial filter with each frequency band from the one-way repeated measures ANOVA test of F1 scores.

Source	Type III sum of squares	Corrected degree of freedom	Mean square	F value	P value	Partial eta squared
SF at [0.01-1] Hz	1.354	3.463	0.391	76.338	<b>0.000</b>	0.590
Error (SF at [0.01-1] Hz)	0.940	183.548	0.005			
SF at [1-2] Hz	0.064	3.551	0.018	2.768	0.034	0.050
Error (SF at [1-2] Hz)	1.232	188.184	0.007			
SF at [2-3] Hz	0.030	3.212	0.009	1.397	0.244	0.026
Error (SF at [2-3] Hz)	1.132	170.222	0.007			
SF at [3-4] Hz	0.015	2.809	0.005	0.755	0.513	0.014
Error (SF at [3-4] Hz)	1.082	148.867	0.007			
SF at [4-5] Hz	0.014	3.318	0.004	0.990	0.404	0.018
Error (SF at [4-5] Hz)	0.728	175.831	0.004			

The only significant simple effect of the spatial filter can be found only in the frequency band of [0.01-1] Hz, where the  $P$  value is less than 0.01.

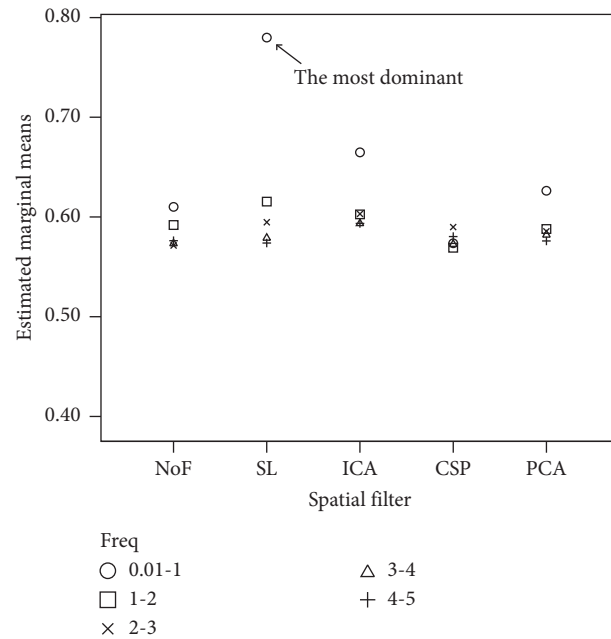


FIGURE 4: Comparison of the estimated marginal means of F1 scores in various frequency bands on different spatial filters. SL spatial filter at [0.01-1] Hz frequency clearly provides the best result.

TABLE 4: Simple effects of frequency with each type of spatial filter from the one-way repeated measures ANOVA test of F1 scores.

Source	Type III sum of squares	Corrected degree of freedom	Mean square	F value	P value	Partial eta squared
Freq on NoF	0.058	3.303	0.018			
Error (freq on NoF)	0.557	175.053	0.003	5.536	<b>0.001</b>	0.095
Freq on SL	1.602	3.167	0.506			
Error (freq on SL)	1.206	167.856	0.007	70.402	<b>0.000</b>	0.571
Freq on ICA	0.196	3.085	0.064			
Error (freq on ICA)	1.014	163.490	0.006	10.265	<b>0.000</b>	0.162
Freq on CSP	0.015	2.743	0.005			
Error (freq on CSP)	1.196	145.357	0.008	0.643	0.575	0.012
Freq on PCA	0.086	3.433	0.025			
Error (freq on PCA)	0.746	181.937	0.004	6.094	<b>0.000</b>	0.103

There are significant simple effects of frequency band on all types of spatial filter but not CSP, where  $P$  value is more than 0.01.

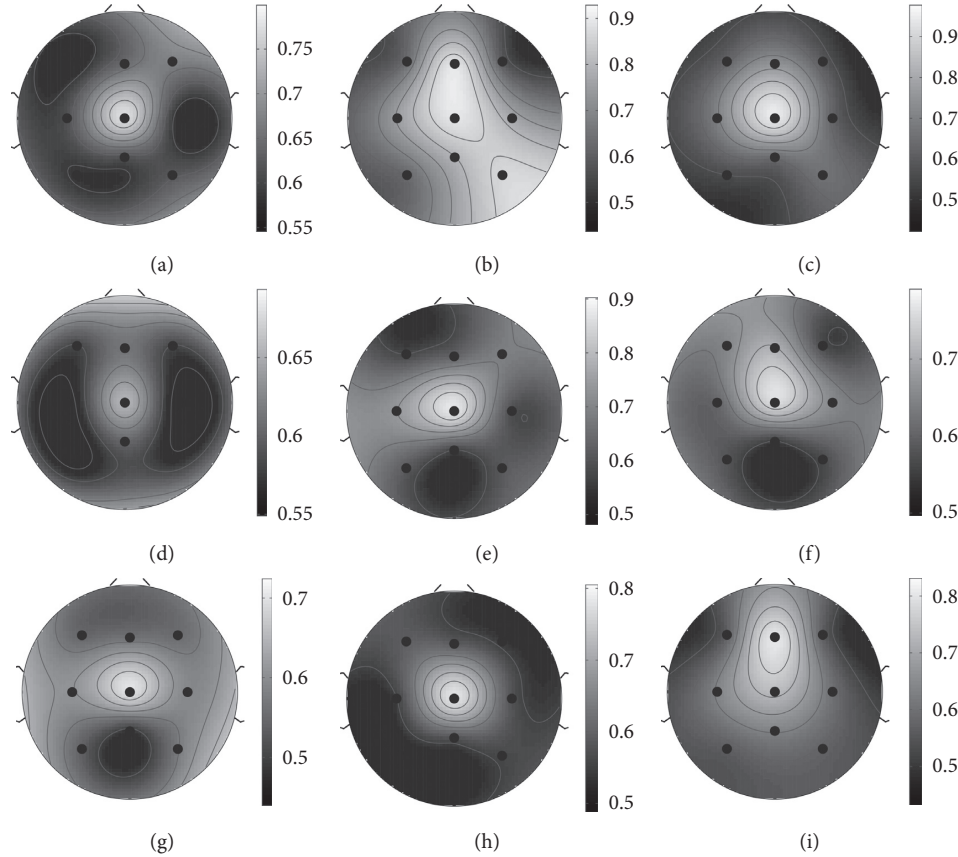


FIGURE 5: The topoplot demonstrates the F1 score of each channel for every subject when frequency was set with [0.01-1] Hz using SL spatial filter. The channels with higher discriminant power appear brighter. (a) Subject 1. (b) Subject 2. (c) Subject 3. (d) Subject 4. (e) Subject 5. (f) Subject 6. (g) Subject 7. (h) Subject 8. (i) Subject 9.

prominent F1 scores for each subject is Cz in most subjects except in subject 2 where both Cz and Fz are prominent and in subject 9 where Fz provides higher F1 score than Cz.

The results in Figure 2 also show that these factors are quite sensitive. Choosing different combinations of the factors can substantially affect the F1 scores. The scores could range from as low as 44.4% with the choice of [3-4] Hz frequency + CSP filter + MF classifier to as high as 82.3% with the choice of [0.01-1] Hz frequency + SL spatial filter + SVM classifier.

After a post hoc test of the classifier, we found that SVM is the most prominent classifier, being an independent factor

without any intervention from other factors. The estimated marginal mean of SVM is 66.4% with a significantly higher F1 score than the others including LDA, as shown in Figure 3.

#### 4. Discussion

This study aimed to investigate factors that affect performance of MRCP detection in asynchronous BCI, which can facilitate stroke rehabilitation systems. To tackle this problem, we carried out a systematic analysis on effects of three prevalent

factors including 5 different frequency bands of the EEG data, 5 spatial filters, and 6 classification algorithms. To focus on both high detection rate and quality of the detection, we analyzed these factors using F1 scores. Series of ANOVA tests were conducted to investigate relationship and effects among these factors based on F1 scores.

Interestingly, the results show that there is no significant relationship in three-way interaction. However, in two-way interaction, there is a significant interactive effect between frequency and spatial filter. This means that the effect of frequency and the effect of spatial filter are dependent in some circumstances. When EEG data were filtered to the frequency band of [0.01-1] Hz, applying different spatial filters will provide significantly different F1 scores. Nonetheless, results were not significantly different regarding different spatial filters in other frequency bands. In other words, the spatial filter only has to be carefully and precisely chosen when used with EEG data in [0.01-1] Hz range. More specifically, the experiment results show that utilizing frequency in [0.01-1] Hz with SL spatial filter provided best F1 scores. This can be interpreted as the discriminant feature of MRCP may be discovered only in frequency band of [0.01-1] Hz and not all types of spatial filter can reveal this discriminant feature. However, we also respect the fact that different experiment protocols and different methods in acquiring EEG data may affect the setting in the preprocessing step; our experiment intends to reveal an influence, effect, and interaction among these factors more than to discover the best setting for the preprocessing step under some preset factors as done in previous works [20, 21, 24]. By pointing out that not all the settings can enhance the accuracy, we suggest researchers who conduct research in this area select these factors in the preprocessing method with extra care.

The classifying method does not have any interaction effects with other factors, meaning that an appropriate classifier will provide good F1 scores regardless of the frequency bands and spatial filters used. The most prominent classifier for MRCP detection is SVM, while LDA is the second best. 1-NN-ED, 1-NN-DTW, and TM provide similar results, while MF provides the worst F1 scores. These results are according to the fact that EEG data are non-stationary signals; thus, a classification algorithm which is a stable learning algorithm like SVM and LDA can provide superior results than 1-NN-ED and 1-NN-DTW [26]. Although MF and TM are members of stable learning algorithm, neither MF nor TM provides comparable F1 score to SVM and LDA. We notice that compressing nonstationary signals into one template as done in MF and TM is not a good solution for dealing with EEG data. However, TM provides comparable F1 scores to 1-NN-ED and 1-NN-DTW, but not MF. This is resulted from the shape-based averaging that tries to average the data based on its shape.

Moreover, the distance measures do not significantly affect the results when EEG data are used in time domain as 1-NN-ED and 1-NN-DTW provide similar results. If memory space is taken into account, TM will be superior to 1-NN-ED and 1-NN-DTW. These three classifiers provide comparable F1 scores; however, 1-NN-ED and 1-NN-DTW

use a lazy learning method that keeps all training instances in memory to make a decision. Instead of keeping all available training instances, MF and TM create templates of groups of training data. Thus, it reduces memory usage and decision time. For MF and TM that needs template constructions, TM is a better choice than MF because TM employs a shape-based averaging technique, whereas MF uses a simple mean averaging method. Furthermore, utilizing TM could provide comparable F1 scores with 1-NN-DTW, which means that the shape-based averaging method such as DBA can reserve important information, while the number of kept instances in memory is minimized.

## 5. Conclusion

We performed a comprehensive statistical analysis on performance of different frequency bands, spatial filters, and classifying methods to reveal effects and relationship among these factors. The performance was analyzed based on F1 scores. The results showed that the classifier is an independent factor, whereas the frequency and the spatial filter are dependent factors. Thus, the frequency and the spatial filter have to be considered simultaneously. These results can be used as guidelines for research in MRCP detection, especially on lower limb movement. For time series miners who are interested in developing an MRCP detection algorithm, either a more sophisticated data representation or projection for MRCP is required to provide superior results.

## Data Availability

The BCI data used to support the findings of this study are available at <http://www.cu-timeseries.com> or from the corresponding author upon request.

## Conflicts of Interest

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

## Acknowledgments

This work was supported by Rachadapisek Sompote Fund for Postdoctoral Fellowship, Chulalongkorn University (to Sura Rodpongpun); Thailand Research Fund and Chulalongkorn University given through Royal Golden Jubilee Ph.D. Program (grant number PHD/0057/2557, Thapanan Janyalikit); the 90 Anniversary of Chulalongkorn University, Rachadapisek Sompote Fund (to Thapanan Janyalikit).

## References

- [1] J. R. Millan, F. Renkens, J. Mourino, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human eeg," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1026–1033, 2004.
- [2] U. Chaudhary, B. Xia, S. Silvoni, L. G. Cohen, and N. Birbaumer, "Brain-computer interface-based communication in the completely locked-in state," *PLoS Biology*, vol. 15, no. 1, Article ID e1002593, 2017.

- [3] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.
- [4] N. Birbaumer, N. Ghanayim, T. Hinterberger et al., "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, 1999.
- [5] S. Muthong, P. Vateekul, and M. Sriyudthsak, "An adjustment strategy on multi-session eeg data for online left/right hand imagery classification," in *Proceedings of the 2016 8th International Conference on Knowledge and Smart Technology (KST)*, pp. 179–183, IEEE, Chiangmai, Thailand, February 2016.
- [6] D. O. Hebb, *The Organization of Behavior: A Neuro-psychological Theory*, Psychology Press, London, UK, 2005.
- [7] J. J. Daly and J. R. Wolpaw, "Brain-computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, 2008.
- [8] J. Ibáñez, E. Monge-Pereira, F. Molina-Rueda et al., "Low latency estimation of motor intentions to assist reaching movements along multiple sessions in chronic stroke patients: a feasibility study," *Frontiers in Neuroscience*, vol. 11, 2017.
- [9] N. Mrachacz-Kersting, S. R. Kristensen, I. K. Niazi, and D. Farina, "Precise temporal association between cortical potentials evoked by motor imagination and afference induces cortical plasticity," *The Journal of Physiology*, vol. 590, no. 7, pp. 1669–1682, 2012.
- [10] K. K. Ang and C. Guan, "Brain-computer interface in stroke rehabilitation," *Journal of Computing Science and Engineering*, vol. 7, no. 2, pp. 139–146, 2013.
- [11] H. H. Kornhuber and L. d. Deecke, "Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: bereitchaftspotential und reafferente potentiale," *Pflügers Archiv European Journal of Physiology*, vol. 284, no. 1, pp. 1–17, 1965.
- [12] D. Liu, W. Chen, Z. Pei, and J. Wang, "Detection of lower-limb movement intention from eeg signals," in *Proceedings of the 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 84–89, IEEE, Siem Reap, Cambodia, June 2017.
- [13] R. Xu, N. Jiang, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "Factors of influence on the performance of a short-latency non-invasive brain switch: evidence in healthy individuals and implication for motor function rehabilitation," *Frontiers in Neuroscience*, vol. 9, p. 527, 2016.
- [14] O. Bai, P. Lin, S. Vorbach, J. Li, S. Furlani, and M. Hallett, "Exploration of computational methods for classification of movement intention during human voluntary movement from single trial eeg," *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2637–2655, 2007.
- [15] E. Yom-Tov and G. F. Inbar, "Detection of movement-related potentials from the electro-encephalogram for possible use in a brain-computer interface," *Medical & Biological Engineering & Computing*, vol. 41, no. 1, pp. 85–93, 2003.
- [16] I. K. Niazi, N. Jiang, O. Tiberghien, J. F. Nielsen, K. Dremstrup, and D. Farina, "Detection of movement intention from single-trial movement-related cortical potentials," *Journal of Neural Engineering*, vol. 8, no. 6, Article ID 066009, 2011.
- [17] C. Lin, B.-H. Wang, N. Jiang, R. Xu, N. Mrachacz-Kersting, and D. Farina, "Discriminative manifold learning based detection of movement-related cortical potentials," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 9, pp. 921–927, 2016.
- [18] R. Xu, N. Jiang, C. Lin, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "Enhanced low-latency detection of motor intention from eeg for closed-loop brain-computer interface applications," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 288–296, 2014.
- [19] G. Garipelli, R. Chavarriaga, and J. d. R. Millán, "Single trial recognition of anticipatory slow cortical potentials: the role of spatio-spectral filtering," in *Proceedings of the 2011 5th International IEEE/EMBS Conference on Neural Engineering (NER)*, IEEE, Cancun, Mexico, pp. 408–411, May 2011.
- [20] C. Bibián, E. López-Larraz, N. Irastorza-Landa, N. Birbaumer, and A. Ramos-Murguialday, "Evaluation of filtering techniques to extract movement intention information from low-frequency eeg activity," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, Jeju Island, Korea, pp. 2960–2963, July 2017.
- [21] E. Lew, R. Chavarriaga, S. Silvoni, and J. d. R. Millán, "Detection of self-paced reaching movement intention from eeg signals," *Frontiers in Neuroengineering*, vol. 5, p. 13, 2012.
- [22] S. Aliakbarhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, "Classification of eeg signals to identify variations in attention during motor task execution," *Journal of Neuroscience Methods*, vol. 284, pp. 27–34, 2017.
- [23] A. Ghaemi, E. Rashedi, A. M. Pourrahimi, M. Kamandar, and F. Rahdari, "Automatic channel selection in eeg signals for classification of left or right hand movement in brain computer interfaces using improved binary gravitation search algorithm," *Biomedical Signal Processing and Control*, vol. 33, pp. 109–118, 2017.
- [24] G. Garipelli, R. Chavarriaga, and J. del R. Millán, "Single trial analysis of slow cortical potentials: a study on anticipation related potentials," *Journal of Neural Engineering*, vol. 10, no. 3, Article ID 036014, 2013.
- [25] D. J. Wright, P. S. Holmes, and D. Smith, "Using the movement-related cortical potential to study motor skill learning," *Journal of Motor Behavior*, vol. 43, no. 3, pp. 193–201, 2011.
- [26] A. Shakeel, M. S. Navid, M. N. Anwar, S. Mazhar, M. Jochumsen, and I. K. Niazi, "A review of techniques for detection of movement intention using movement-related cortical potentials," *Computational and Mathematical Methods in Medicine*, vol. 201513 pages, 2015.
- [27] N. Jiang, L. Gizzi, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "A brain-computer interface for single-trial detection of gait initiation from movement related cortical potentials," *Clinical Neurophysiology*, vol. 126, no. 1, pp. 154–159, 2015.
- [28] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for eeg-based communication," *Electroencephalography and Clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.
- [29] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, 2000.
- [30] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selection in motor imagery based brain-computer interface," in *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society*, IEEE, Shanghai, China, pp. 5392–5395, September 2005.
- [31] L. I. Smith, "A tutorial on principal components analysis," Technical Report, Cornell University, Ithaca, NY, USA, 2002.



- [32] G. Clifford, "Singular value decomposition & independent component analysis for blind source separation," in *Course Materials for HST.582J/6.555J/16.456J, Biomedical Signal and Image Processing, Spring 2007, MIT OpenCourseWare*, Massachusetts Institute of Technology, Cambridge, MA, USA, 2007, <http://ocw.mit.edu>.
- [33] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [34] T. Janyalikit, P. Sathianwiriyaikhun, H. Sivaraks, and C. A. Ratanamahatana, "An enhanced support vector machine for faster time series classification," in *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, Springer, Pa Tong, Thailand, pp. 616–625, 2016.
- [35] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.
- [36] F. Petitjean, G. Forestier, G. I. Webb, A. E. Nicholson, Y. Chen, and E. Keogh, "Dynamic time warping averaging of time series allows faster and more accurate classification," in *Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM)*, IEEE, Dallas, TX, USA, pp. 470–479, 2014.
- [37] P. Sathianwiriyaikhun, T. Janyalikit, and C. A. Ratanamahatana, "Fast and accurate template averaging for time series classification," in *Proceedings of the 2016 8th International Conference on Knowledge and Smart Technology (KST)*, IEEE, Chiangmai, Thailand, pp. 49–54, February 2016.
- [38] C. A. Ratanamahatana and E. Keogh, "Three myths about dynamic time warping data mining," in *Proceedings of the 2005 SIAM International Conference on Data Mining*, SIAM, Newport Beach, CA, USA, pp. 506–510, April 2005.

## Research Article

# Research of Low-Rank Representation and Discriminant Correlation Analysis for Alzheimer's Disease Diagnosis

Zhigang Li,<sup>1</sup> Aimei Dong ,<sup>1,2</sup> and Jing Zhou<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Science), Jinan 250353, China

<sup>2</sup>Key Lab of Pulp and Paper Science & Technology Ministry of Education, Qilu University of Technology (Shandong Academy of Science), Jinan 250353, China

Correspondence should be addressed to Aimei Dong; [amdong@qlu.edu.cn](mailto:amdong@qlu.edu.cn)

Received 12 January 2020; Accepted 21 February 2020; Published 19 March 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Zhigang 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.

As population aging is becoming more common worldwide, applying artificial intelligence into the diagnosis of Alzheimer's disease (AD) is critical to improve the diagnostic level in recent years. In early diagnosis of AD, the fusion of complementary information contained in multimodality data (e.g., magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF)) has obtained enormous achievement. Detecting Alzheimer's disease using multimodality data has two difficulties: (1) there exists noise information in multimodal data; (2) how to establish an effective mathematical model of the relationship between multimodal data? To this end, we proposed a method named LDF which is based on the combination of low-rank representation and discriminant correlation analysis (DCA) to fuse multimodal datasets. Specifically, the low-rank representation method is used to extract the latent features of the submodal data, so the noise information in the submodal data is removed. Then, discriminant correlation analysis is used to fuse the submodal data, so the complementary information can be fully utilized. The experimental results indicate the effectiveness of this method.

## 1. Introduction

Alzheimer's disease (AD) is a type of neurodegenerative disease, which is caused by many factors. With the increase in the aging population, the incidence and mortality of Alzheimer's disease increase year by year [1, 2]. Alzheimer's care and treatment cost up to \$290 billion a year. Timely detection is the key to the treatment of Alzheimer's disease, but it is very difficult to diagnose Alzheimer's disease due to the diversity of its causes. Using medical imaging technology [3–6] to assist clinical staff in diagnosis is the primary method to detect Alzheimer's disease. Each imaging [3, 5, 6] device can display pathological information of different tissues and organs of the human body, as well as different forms of pathological information of the same organ. Because of the diversity of the causes of Alzheimer's disease and the phenomenon of brain atrophy, it is difficult to capture all the immunological information contained in the medical image only by the naked eye. So in the past decade, people

began to identify Alzheimer's disease by machine learning. The computer can capture various fine-grained pathological information which cannot be obtained by human eyes. Computer-aided diagnosis [7–11] has grown up to be an important basis for clinicians to diagnose diseases. Machine learning offers a theoretical basis for it.

In recent years, the research shows that the performance of diagnosis can be substantially improved by using multimodal data with complementary information. Using multimodal data to detect Alzheimer's disease has grown up to be a research hotspot. A double-layer polynomial network [12] method is proposed. Firstly, the first-layer polynomial network extracts the high-level semantic features of MRI and PET data, and the second-layer polynomial network is employed to multimodal data fusion. This method reduces the noise of data, but will cause the loss of latent features. Zhu et al. [13] proposed a method for combining feature selection and subspace learning to identify and select features in a unified framework. Nevertheless, it ignores the

internal feature structure of data. Liu et al. [14] proposed a new multitype diagnosis framework, which is composed of an automatic encoder and soft ax layers. The multimodal data are shared through the automatic encoder network spatial representation. This method can effectively learn the potential features of multimodal data, ignoring the relationship between modes. In literature [15], a new approach for HGR is proposed which is based on Quartile Deviation of Normal Distribution (QDOND) for mortal extraction and Bayesian model along with binomial distribution for features fusion and best features selection. This method does not consider the influence of noise value. In literature [16], a fusion method based on phase consistency and local Laplace energy weighting is proposed. The high-frequency and low-frequency features of diverse modal data are obtained by NSCT, in which the high-frequency features are fused by phase consistency rules and the low-frequency features are fused by local Laplace energy weighting. But the computational efficiency of this method needs to be improved.

There are for two main challenges in multimodal data fusion: (1) dealing with noise data and redundant information; (2) modeling the relationship between multimodal data effectively. The above method uses joint representation to learn the shared potential features of multimodal data. Noise information and redundant information in the data are not effectively processed, and the relationship between the multimodal data is also suppressed. In view of the above problems, we propose a feature fusion method founded on the combination of low-rank representation and discriminant correlation analysis. The proposed method has three advantages: (1) noise reduction and subspace feature learning of the original data reduce the noise value and redundancy of the data; (2) maximize the pairwise correlation between the submodal and the submodal features and effectively simulate the relationship between the modes; (3) replace the original features with fusion features to avoid noise information to the greatest extent.

The rest of the work is as follows: In Section 2, we introduce the method named LDF built on low-order representation and discriminant correlation. We further describe and review the experimental results in Section 3. Finally, we summarize this paper in Section 4.

## 2. Method

In this part, we will introduce the proposed method LDF in detail. The LDF method is mainly composed of four parts: data completion, feature processing, feature fusion, and SVM classification. Due to the lack of data, the KNN algorithm is used to complete the data first; secondly, low-rank representation is used to extract the potential features of the data and denoise the multimodal data; thirdly, discriminant correlation analysis is used to model the submodal data and get the fusion matrix; finally, the fusion results are input into the support vector machine classifier for classification, see Figure 1 for detail.

**2.1. Data Completion.** There are missing data in the original dataset. Existing studies [17] have shown that deleting missing data can affect the accuracy of experiments. The existing research proves that the KNN algorithm is superior to other algorithms in the supplement of missing value. At the same time, we have carried out a large number of experiments and the results demonstrated the superiority of the KNN method to other data completion methods. The core idea of the KNN [18, 19] algorithm is tantamount to the distance between the missing data items and the complete dataset, and the K value is selected closest to the missing data for weighted average, as the supplement of the true value. We choose the value of K as 5. The KNN algorithm replaces the missing value by finding the nearest K number and finding the weighted average sum of these K numbers:

$$\text{dist}(X_i, Y_j) = \sqrt{\sum_{l=1}^n (X_{il} - Y_{jl})^2}, \quad (1)$$

$$\bar{x} = \frac{\sum_{i=1}^K w x_{ij}}{K},$$

where  $\text{dist}(X_i, Y_j)$  is the European distance,  $\bar{x}$  is the imputed data, and  $w = 1/d_i / \sum_{i=1}^K 1/d_i$  is the weight.

**2.2. Low-Rank Representation.** The original data exist in high-dimensional space and contain noise data. High-dimensional data usually contain hidden features. Low-rank representation [20–22] is a powerful and applicable tool to extract hidden features and remove noise information from high-dimensional data. Our goal is to extract latent features from high-dimensional space and remove the noise information contained in the original data. We define raw data as  $\mathbf{A}$ :

$$\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n), \quad (2)$$

where  $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n$  represents single submodal feature data.

The purpose of low-rank representation [20] is to determine the internal relationship between the sample points and extract a global feature. The low-rank representation of the matrix is primarily obtained through the convex optimization algorithm of gradual approximation.

In order to extract the hidden features contained in the original data and remove the noise information contained in the original data, we divide matrix  $\mathbf{A}$  into two parts. The first part is a linear combination of  $\mathbf{A}$  and a low-rank matrix  $\mathbf{X}$ , which contains the hidden information in the original matrix. The second part is noisy data, which is a sparse matrix  $\mathbf{E}$ :

$$\mathbf{A} = \mathbf{A}\mathbf{X} + \mathbf{E}. \quad (3)$$

In the above formula, the solutions of  $\mathbf{X}$  and  $\mathbf{E}$  are infinite, but we want the solutions of  $\mathbf{X}$  to be low rank, so we can convex relax the optimization problem:

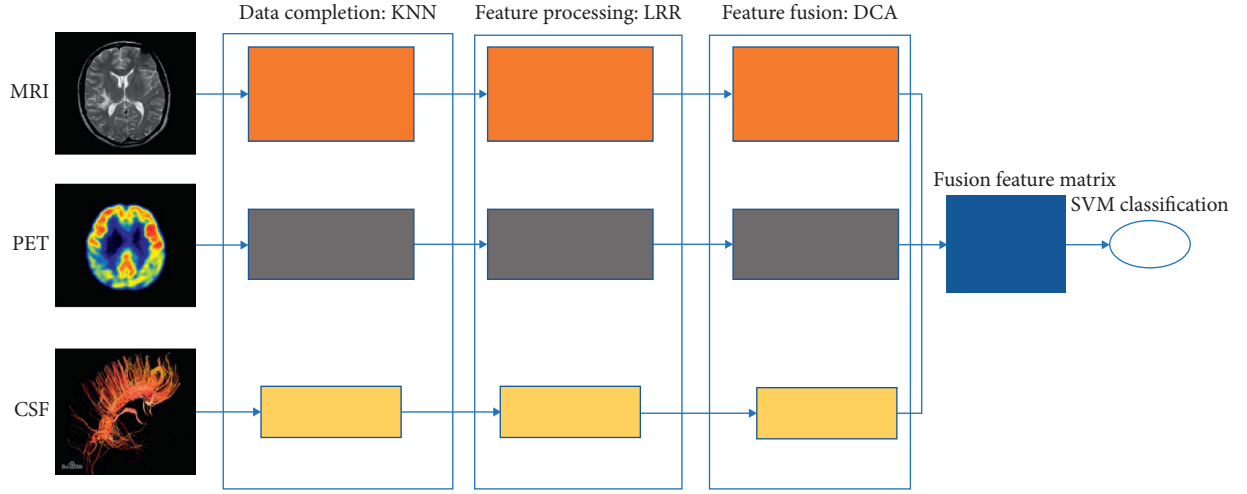


FIGURE 1: Framework of the proposed method LDF.

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_*, \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{A}\mathbf{X}. \end{aligned} \quad (4)$$

We need to extract features from multiple subspaces, and in order to make noise and outliers more robust. Considering joint subspaces, they can be expressed as

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{E}} \quad & \|\mathbf{X}\|_* + \lambda \|\mathbf{E}\|_{2,1}, \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{A}\mathbf{X} + \mathbf{E}, \end{aligned} \quad (5)$$

where  $\|\mathbf{X}\|_* = \sum_i \sigma_i$  ( $\sigma_i$  is the singular value of  $\mathbf{X}$ ) [23],  $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m (\mathbf{E}_{ij})^2}$  is the noise regularization strategy, and  $\lambda$  is a positive free parameter, which is used to balance the weight of the low-rank matrix and the sparse matrix.

**2.3. Feature Fusion.** Discriminant correlation analysis (DCA) [24, 25] is an improved algorithm based on canonical correlation analysis (CCA). The existing feature fusion algorithm [12–15, 26] uses the neural network or sparse representation to jointly represent multimodal data, leading to suppress the relationship between multimodal data. CCA (20–22) can effectively model the relationship between multimodal data, but it cannot deal with the redundant information in the data. To this end, we propose the LDF method which uses DCA for Alzheimer's disease detection based on low-rank representation. The LDF method effectively models the relationship between submodes by maximizing the correlation of similar features. That is to say, on one hand, low-rank representation can remove the noise data existing in original data; on the other hand, DCA can minimize the correlation between different features and remove redundant information.

DCA can be divided into two parts: (1) discriminant analysis of each feature is set through the interclass scatter matrix; (2) correlation analysis between feature sets is driven by the diagonalization of the intergroup covariance matrix. The calculation formula of the scatter matrix  $S_{\text{intera}}$  between classes is

$$\begin{aligned} S_{\text{intera}_{p \times p}} &= \sum_{j=1}^c m(\bar{a}_j - \bar{a})(\bar{a}_j - \bar{a})^T \\ &= \Phi_{\text{intera}} \Phi_{\text{intera}}^T, \end{aligned} \quad (6)$$

where  $\Phi_{\text{intera}} = [\sqrt{m}(\bar{a}_1 - \bar{a}), \sqrt{m}(\bar{a}_2 - \bar{a}), \dots, \sqrt{m}(\bar{a}_c - \bar{a})]$ ,  $\bar{a}_j$  is the mean of the  $j$ th class and  $\bar{a}$  is mean of whole feature set, and  $\Phi_{\text{intera}} \Phi_{\text{intera}}^T$  is a symmetric semidefinite matrix.

The first step of DCA is to project the feature matrix  $\mathbf{A}$  into a new  $r$ -dimensional space  $\mathbf{A}'_{(r \times n)} = \mathbf{W}_{\text{intera}}^T \mathbf{A}_{(p \times n)}$  by finding a new transformation matrix  $\mathbf{W}_{\text{intera}}$  [24]. Our aim is to reduce the correlation between different features by minimizing the correlation between different features. So we can diagonalize [27] it and change the divergence matrix  $S_{\text{intera}}$  between-class into  $S'_{\text{intera}} = \mathbf{W}_{\text{intera}}^T S_{\text{intera}} \mathbf{W}_{\text{intera}} = \Phi'_{\text{intera}} \Phi_{\text{intera}}^T = \mathbf{I}$ . The scattering matrix  $\Phi'_{\text{intera}} \Phi_{\text{intera}}^T$  is a strictly diagonally dominant matrix, in which the diagonal element is close to 1 and the nondiagonal element is close to 0. The way to obtain the transformation matrix  $\mathbf{W}_{\text{intera}}$  is obtained from [24]. Similarly, we can use the same method to solve for the second feature  $\mathbf{B}$  to get the transformation matrix  $\mathbf{W}_{\text{interb}}$  corresponding to different feature subsets, and the updated class scatter matrix  $S'_{\text{interb}} = \mathbf{W}_{\text{interb}}^T S_{\text{interb}} \mathbf{W}_{\text{interb}} = \Phi'_{\text{interb}} \Phi_{\text{interb}}^T = \mathbf{I}$ .

Secondly, in order to maximize the correlation between the two feature sets  $\mathbf{A}$  and  $\mathbf{B}$ , diagonalize the interclass covariance matrix  $S'_{ab} = \mathbf{A}' \mathbf{B}'$  of the transformed feature set, which is diagonalized by SVD:

$$S'_{ab_{(r \times r)}} = \mathbf{U} \Sigma \mathbf{V}^T \Rightarrow \mathbf{U}^T S'_{ab_{(r \times r)}} \mathbf{V} = \Sigma, \quad (7)$$

where  $\Sigma$  is the diagonal matrix made up of nonzero diagonal elements. If  $\mathbf{W}_{ca} = \mathbf{U} \Sigma^{-1/2}$  and  $\mathbf{W}_{cb} = \mathbf{V} \Sigma^{-1/2}$ , then

$$(\mathbf{U} \Sigma^{-1/2})^t S'_{ab} (\mathbf{V} \Sigma^{-1/2}) = \mathbf{I}. \quad (8)$$

Thus,  $\mathbf{W}_{ca}$  and  $\mathbf{W}_{cb}$  are the transformation matrices for  $\mathbf{A}'$  and  $\mathbf{B}'$  and the resulting transformed feature sets are written as

$$\begin{aligned} \mathbf{A}^* &= \mathbf{W}'_{ca} \mathbf{A}' = \mathbf{W}_{ca}^T \mathbf{W}_{intera}^T \mathbf{A} = \mathbf{W}_a \mathbf{A}, \\ \mathbf{B}^* &= \mathbf{W}'_{cb} \mathbf{B}' = \mathbf{W}_{cb}^T \mathbf{W}_{interb}^T \mathbf{B} = \mathbf{W}_b \mathbf{B}. \end{aligned} \quad (9)$$

After getting the transformed features  $\mathbf{A}^*$  and  $\mathbf{B}^*$ , we connect  $\mathbf{A}^*$  and  $\mathbf{B}^*$  in series to get the fusion matrix. The specific flowchart is shown in Figure 2.

### 3. Experiment

**3.1. Data Set and Experimental Environment.** In recent years, using multimodal incomplete heterogeneous data to detect Alzheimer's disease has become a very important clinical and research problem. The ADNI-1 database has been widely used in many studies. The ADNI-1 dataset contains the longitudinal, multisite MRI and PET image data of Alzheimer's disease, mild cognitive impairment, and elderly control patients, describing the longitudinal changes of the brain structure and metabolism, as well as clinical/cognitive and biomarker data. According to MMSE (Mini-Mental State Examination), ADNI-1 is divided into the NC (normal control) group, MCI (mild cognitive impairment) group, and AD (Alzheimer disease) group. There are 805 subjects in the baseline ADNI-1 database. Specifically, 226 subjects are NC, 393 MCI, and 186 AD. All subjects had at least one of the three data modalities: MRI, PET, and CSF. A summary of the ADNI-1 database used in this study is given in Table 1. For a detailed description of the ADNI-1 database, please visit <http://adni.loni.usc.edu>.

All the algorithms are carried out in Matlab2018b on a computer with Intel Core i7-8750H 2.20 GHz CPU and 8.00 GB RAM.

**3.2. Comparison Method.** Feature fusion methods are divided into the pixel-level fusion method, feature-level fusion method, and decision-level fusion method. In this experiment, we select a variety of feature-level and decision-level fusion methods to compare with our methods. The specific methods used are shown in Table 2.

Among them, KNN, SVD, and EM [24] are the three frequently used methods of data completion. KNN, SVD, and EM [24] algorithms are used to complete the missing data, and the completed data are concatenated in series to get the fusion matrix. Among them, the number of iterations of the EM algorithm is 50, and the value of  $K$  in KNN is 5. For SVD, we choose a matrix containing 95% data information. Specifically, In the KNN method, the missing data are completed with the mean of its  $K$ -nearest neighbor columns; in the SVD method, the missing data are iteratively computed using the matrix completion technique with low-rank approximation; in the EM method, the missing data are imputed using the EM algorithm.

CCA [28] is a traditional feature-level multimodal data fusion method, which integrates the correlation between two modes to fuse the multimodal data. Specifically, by analyzing the linear relationship between the original eigenvectors, the CCA feature-level fusion method uses the correlation criterion function to extract the typical correlation components

of the two modal eigenvectors, thus obtaining the final features.

The LMP [29] algorithm uses the low-rank representation to extract the features of all the modal data and then gets the features of each modal and finally assigns different weights to these features. Parameter is the weight of modal data. For the three submodes, the different weights are 0.5, 0.25, and 0.25, respectively. Specifically, the low-rank representation is used to project the data into a low-dimensional space. The score matrix is obtained by using the relationship between the original data and the projection matrix. Different weights are assigned to different modal data according to the order of scores.

**3.3. Analysis of Experimental Results.** In this experiment, due to the missing of data, we used the KNN algorithm to complete the data. We obtained a  $K$  value of 5 (different from the value of  $K$  in the comparison algorithm). We used 10-fold cross-validation strategy to evaluate all comparison methods. Specifically, we first randomly partitioned the whole dataset into 10 subsets, and then selected one subset for testing and used the remaining 9 subsets for training. We repeated the whole process 10 times to avoid the possible bias in dataset partitioning during cross-validation, and then the averaging result was adopted as the final result.

We performed extensive experiments with the datasets demonstrated in Table 1 and the parameter settings demonstrated in Table 2. The obtained results are reported in Tables 3–5 and Figures 2 and 3. In this experiment, we selected ACC (accuracy), SEN (sensitivity), SPE (specificity), BAC (balanced accuracy), PPV (positive predictive value), NPV (negative predictive value), and other indicators as the evaluation criteria to compare our method with other methods. According to these indicators, we can know the prevalence, missed diagnosis rate, and misdiagnosis rate of our method. We choose the average value and standard deviation of each index result in ten experiments as the final output. Due to the limited space, in this paper, we only choose the ROC curve for MCI/NC and the contrast experiment of time complexity, and the ROC curves for AD/NC and MCI/AD are similar to that of MCI/NC.x

Table 3 shows our experimental results of AD/NC. From Table 3, we can see clearly that our method has achieved good results in all aspects compared with other methods. Compared with other methods, our method has improved the accuracy by about 3.5%. At the same time, our method has performed well in sensitivity. Compared with other methods, our method has improved the accuracy by about 6%, and in NPV and BAC by about 5%, which shows that our method can accurately identify Alzheimer's patients.

It is difficult to find Alzheimer's disease in the early stage. Timely discovery is the best way to treat Alzheimer's disease. From Table 4, we can see that our method can more accurately detect the early symptoms of Alzheimer's disease. Our method has made good achievements in this respect. Compared with other methods, our method has improved the accuracy by 5%, the sensitivity by 25% compared with several feature-level fusion methods, and the BAC by 25%



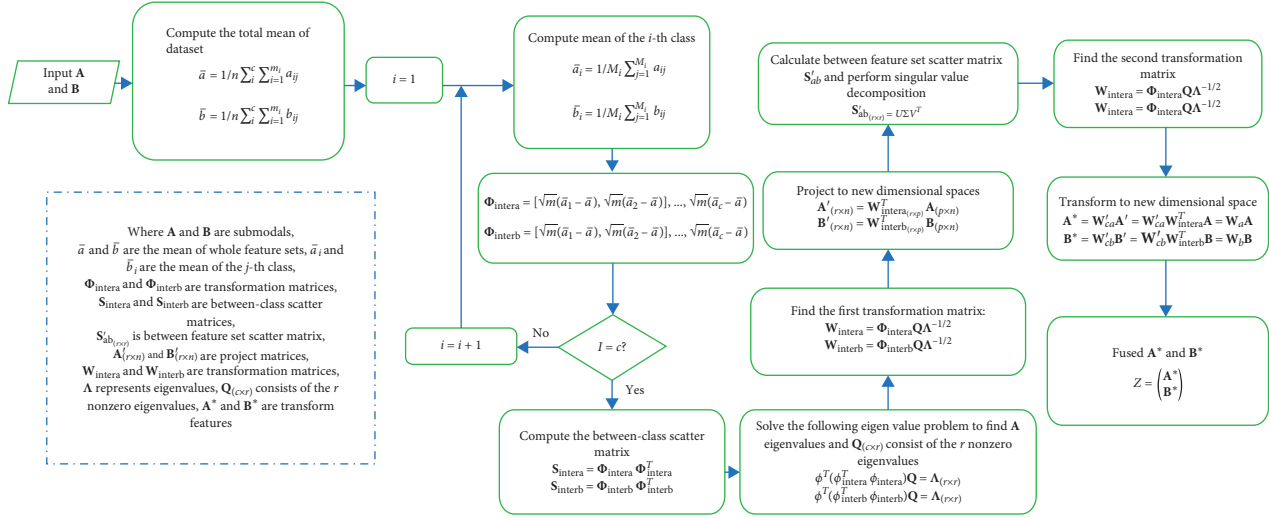


FIGURE 2: Flowchart for computing DCA-based feature-level fusion.

TABLE 1: Summary of the adopted datasets.

	MRI	PET	CSF	Total
Number of features	93	93	3	189
AD subjects	186	93	102	51
MCI subjects	393	201	190	97
NC subjects	226	101	112	52
Total subjects	805	395	404	200

TABLE 2: Description of the compared method.

	Comparison method	Parameter selection
Feature level	EM	50
	KNN	7
	SVD	95%
	CCA	None
Decision level	LRRF	0.5; 0.25; 0.25

TABLE 3: Classification results achieved by 6 different methods for the classification task AD/NC.

	ACC	SEN	SPE	BAC	PPV	NPV
KNN	87.83 ± 3.71	91.47 ± 5.13	85.60 ± 7.38	88.54 ± 2.79	89.95 ± 5.94	88.60 ± 4.73
EM	88.09 ± 4.36	83.96 ± 8.05	91.32 ± 4.11	87.64 ± 4.33	88.35 ± 5.42	88.16 ± 7.05
SVD	87.07 ± 5.52	88.31 ± 6.19	<b>94.73 ± 4.63</b>	83.18 ± 10.70	<b>91.31 ± 8.18</b>	86.08 ± 7.09
CCA	84.71 ± 3.93	89.04 ± 7.16	79.85 ± 5.28	84.44 ± 4.17	83.96 ± 3.10	86.06 ± 9.48
LMP	86.31 ± 5.62	82.74 ± 8.32	93.19 ± 3.28	87.97 ± 4.65	91.26 ± 4.00	86.36 ± 6.99
LDF	<b>90.52 ± 3.69</b>	<b>94.79 ± 3.98</b>	85.92 ± 7.67	<b>91.35 ± 4.15</b>	88.72 ± 7.52	<b>93.17 ± 4.99</b>

TABLE 4: Classification results achieved by 6 different methods for the classification task MCI/NC.

	ACC	SEN	SPE	BAC	PPV	NPV
KNN	66.53 ± 4.19	8.28 ± 5.19	<b>97.58 ± 2.63</b>	53.49 ± 2.24	65.69 ± 4.78	<b>82.72 ± 19.24</b>
EM	63.71 ± 8.34	25.60 ± 9.27	86.46 ± 5.40	56.03 ± 4.39	52.93 ± 12.35	66.05 ± 10.13
SVD	64.34 ± 5.67	26.01 ± 8.90	85.19 ± 8.47	55.59 ± 7.28	49.74 ± 20.72	68.31 ± 5.21
CCA	68.78 ± 4.94	25.63 ± 8.12	91.66 ± 7.68	58.65 ± 4.17	69.95 ± 4.59	68.52 ± 23.53
LMP	69.90 ± 5.63	<b>55.98 ± 11.79</b>	77.11 ± 7.70	66.54 ± 6.66	<b>73.99 ± 8.15</b>	59.04 ± 16.56
LDF	<b>73.49 ± 5.36</b>	48.64 ± 10.30	86.28 ± 6.50	<b>67.76 ± 5.38</b>	67.59 ± 13.20	76.80 ± 7.07

TABLE 5: Classification results achieved by 6 different methods for the classification task AD/MCI.

	ACC	SEN	SPE	BAC	PPV	NPV
KNN	69.48 $\pm$ 5.14	97.65 $\pm$ 2.56	13.87 $\pm$ 7.11	55.76 $\pm$ 3.61	69.13 $\pm$ 5.09	<b>80.00 <math>\pm</math> 20.11</b>
EM	69.14 $\pm$ 3.94	12.67 $\pm$ 4.55	<b>98.51 <math>\pm</math> 2.09</b>	55.61 $\pm$ 3.16	<b>80.83 <math>\pm</math> 27.79</b>	68.40 $\pm$ 4.66
SVD	68.45 $\pm$ 4.81	19.18 $\pm$ 12.60	93.91 $\pm$ 5.00	56.57 $\pm$ 5.70	65.39 $\pm$ 30.62	68.91 $\pm$ 4.69
CCA	67.55 $\pm$ 4.93	<b>98.12 <math>\pm</math> 2.25</b>	3.88 $\pm$ 3.70	52.43 $\pm$ 4.19	67.79 $\pm$ 5.23	59.52 $\pm$ 34.50
LMP	69.96 $\pm$ 6.62	93.98 $\pm$ 3.69	26.36 $\pm$ 8.87	<b>59.97 <math>\pm</math> 5.15</b>	70.13 $\pm$ 7.23	70.17 $\pm$ 17.99
LDF	<b>72.54 <math>\pm</math> 5.78</b>	16.46 $\pm$ 6.40	96.45 $\pm$ 2.84	56.45 $\pm$ 4.13	66.17 $\pm$ 25.34	73.16 $\pm$ 6.59

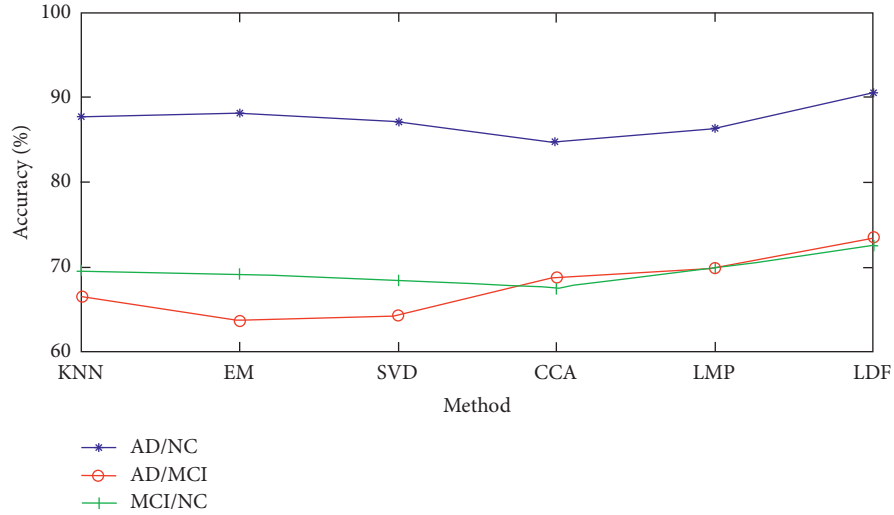


FIGURE 3: Classification result curves of six different methods of three classification tasks.

TABLE 6: Running time achieved by 6 different methods for the classification task MCI/NC.

	KNN (s)	EM (s)	SVD (s)	CCA (s)	LMP (s)	LDF (s)
Running time	3.256	4.362	3.358	0.254	10.508	5.66

compared with other methods The increase of 5% indicates that our method can diagnose mild cognitive impairment more accurately.

Table 5 shows the experimental results of AD/MCI. Compared with other methods, our method has improved the accuracy by about 4%. We can see the effectiveness of our method through the experimental results of AD/NC, AD/MCI, and MCI/NC.

In this experiment, the time complexity of several methods is also analyzed and compared. The results are shown in Table 6. Compared with other methods, our method needs more time. In the future work, we will further optimize it to reduce the time complexity.

In order to analyze the experimental results more clearly, we have carried out visual processing on the experimental results, as shown in Figure 3. In Figure 3, we can clearly see that in this experiment, compared with the decision-making level fusion method, the feature and fusion methods get better results in terms of accuracy, and our method also gets better results in terms of accuracy.

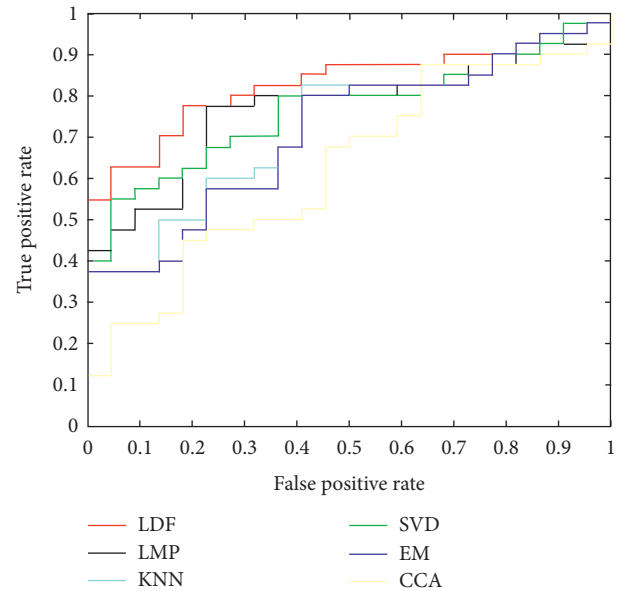


FIGURE 4: ROC curves of six different methods of MCI/NC classification tasks.

Figure 4 shows ROC curves of several methods. It can be clearly seen from the figure that our method can achieve more accurate results than other methods. Through the area under the curve (AUC), we can see that our method is

obviously superior to other methods. At the same time, our method is better than other methods in disease recognition.

## 4. Conclusions

In this paper, we propose a feature fusion method based on low-rank representation and discriminant correlation analysis (LDF). Firstly, we use low-rank representation to extract the features of the original data and then use DCA to fuse the features. The experimental results show that our results are effective. In the future work, we will continue to improve our method. While modeling the relationship between modes, we will ensure that the relationship between contexts within modes will not be affected. At the same time, we will continue to improve and simplify it to obtain good time complexity in case of big data application.

## Data Availability

All the data used in the manuscript can be downloaded from the ADNI website at <http://adni.loni.usc.edu/>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was financially supported by the National Natural Science Foundation of China (61703219 and 61702292).

## References

- [1] M. G. Zhou, H. D. Wang, X. Zeng et al., "Mortality, morbidity, and risk factors in China and its provinces, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 394, no. 10204, pp. 1145–1158, 2019.
- [2] J. Gaugler and B. James, "2019 Alzheimer's disease facts and figures," *Alzheimers & Dementia*, vol. 15, no. 3, pp. 321–387, 2019.
- [3] J. Zhao, J. Chen, S. Ma et al., "Recent developments in multimodality fluorescence imaging probes," *Acta Pharmaceutica Sinica B*, vol. 8, no. 3, pp. 320–338, 2018.
- [4] S. H. Park and K. Han, "Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction," *Radiology*, vol. 286, no. 3, pp. 800–809, 2018.
- [5] M. B. Luis, S. Ramon, B. Paula et al., "Multimodality imaging techniques," *Contrast Media & Molecular Imaging*, vol. 5, no. 4, pp. 180–189, 2010.
- [6] K. M. Stewart, K. L. Horton, and S. O. Kelley, "Cell-penetrating peptides as delivery vehicles for biology and medicine," *Organic & Biomolecular Chemistry*, vol. 6, no. 13, pp. 2242–2252, 2008.
- [7] A. P. James and B. V. Dasarathy, "Medical image fusion: a survey of the state of the art," *Information Fusion*, vol. 19, pp. 4–19, 2014.
- [8] Y. Liu, X. Chen, Z. Wang et al., "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Information Fusion*, vol. 42, pp. 158–173, 2018.
- [9] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: a survey of the state of the art," *Information Fusion*, vol. 33, pp. 100–112, 2017.
- [10] J. Du, W. Li, K. Lu, and B. Xiao, "An overview of multi-modal medical image fusion," *Neurocomputing*, vol. 215, pp. 3–20, 2016.
- [11] M. M. Bronstein, A. M. Bronstein, F. Michel et al., "Data Fusion through Cross-Modality Metric Learning Using Similarity-Sensitive hashing," in *Proceedings of the Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010*, San Francisco, CA, USA, June 2010.
- [12] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 173–183, 2018.
- [13] X. Zhu, H.-I. Suk, S.-W. Lee, and D. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 607–618, 2016.
- [14] S. Liu, S. Liu, W. Cai et al., "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1132–1140, 2016.
- [15] H. Arshad, M. A. Khan, M. Sharif, M. Yasmin, and M. Y. Javed, "Multi-level features fusion and selection for human gait recognition: an optimized framework of Bayesian model and binomial distribution," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 12, pp. 3601–3618, 2019.
- [16] Z. Q. Zhu, M. G. Zheng, G. Qi, D. Wang, and Y. Xiang, "A phase congruency and local laplacian energy based multimodality medical image fusion method in NSCT domain," *IEEE Access*, vol. 7, pp. 20811–20824, 2019.
- [17] K. J. M. Janssen, A. R. T. Donders, F. E. Harrell et al., "Missing covariate data in medical research: to impute is better than to ignore," *Journal of Clinical Epidemiology*, vol. 63, no. 7, pp. 721–727, 2010.
- [18] A. Jadhav, D. Pramod, and K. Ramanathan, "Comparison of performance of data imputation methods for numeric dataset," *Applied Artificial Intelligence*, vol. 33, no. 10, pp. 913–933, 2019.
- [19] R. Ramanathan, J. Wang, M. Su et al., "Missing value imputation approach for mass spectrometry-based metabolomics data," *Scientific Reports*, vol. 8, 2018.
- [20] G. Liu and S. Yan, "Latent Low-rank representation for subspace segmentation and feature extraction," in *Proceedings of the IEEE International Conference on Computer Vision*, Barcelona, Spain, November 2011.
- [21] Y. Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 296–308, 2015.
- [22] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognition*, vol. 48, no. 10, pp. 3102–3112, 2015.
- [23] M. Fazel, *Matrix Rank Minimization with Applications*, Stanford University, Stanford, CA, USA, 2002.
- [24] M. Heng, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984–1996, 2016.

- [25] M. Turk and A. Pentland, "Eigenfaces for recognition," *Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [26] T. Zhou, K. H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Human Brain Mapping*, vol. 40, no. 3, pp. 1001–1016, 2018.
- [27] Y. Zhang, M. Haghighat, and M. Abdel-Mottaleb, "Kernel discriminant correlation analysis: feature level fusion for nonlinear biometric recognition," in *Proceedings of the 24th International Conference on Pattern Recognition (ICPR)*, pp. 3198–3203, Beijing, China, August 2018.
- [28] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.
- [29] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol. 24, no. 13, pp. i232–i240, 2008.

## Research Article

# Validity Prediction of Amplitude-Integrated EEG in Early Neuromotor Development Outcomes in High-Risk Neonates

Jian Guo , Wentao Wang , and Qili Zhou 

*Affiliated Hospital of Chengde Medical College, Chengde, Hebei 067000, China*

Correspondence should be addressed to Jian Guo; [guojian@cdmc.edu.cn](mailto:guojian@cdmc.edu.cn)

Received 31 December 2019; Revised 28 January 2020; Accepted 30 January 2020; Published 11 March 2020

Guest Editor: Yi-Zhang Jiang

Copyright © 2020 Jian Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous advancement of medical technology, the survival rate of high-risk children is increasing year by year, but the developmental problems that have gradually become apparent in the later stages have a serious impact on the quality of life of children. Amplitude-integrated EEG is an EEG monitoring technology developed for clinical use in newborns in recent years. Therefore, to better detect neuromata development in high-risk children, this study explores the validity prediction of amplitude-integrated EEG in early neuromata development in high-risk children. For 100 high-risk children, amplitude-integrated EEG was used for monitoring, and the exercise scale and validity predictors in the Bailey Infant Development Scale were used to assess whether high-risk children had neurobehavioral abnormalities. The experimental results show that the application of amplitude-integrated EEG can make accurate and effective predictions of early neuromata development outcomes in high-risk children. Compared with traditional neurological examination methods, it has higher sensitivity, specificity, positive predictive value, and consistency in predicting the early neuromata development outcomes of high-risk children. It is suitable for application and promotion in China and has a good application value.

## 1. Introduction

With the continuous advancement of medical technology, the survival rate of high-risk children is increasing year by year. However, the developmental problems that have gradually become apparent in the later period have severely affected the quality of life of children, including audio-visual impairment, cerebral palsy, and mental retardation [1]. In recent years, the incidence of high-risk children has continued to rise and has now reached about 8–10% [2]. And with the development of the technology of neonatal intensive care unit (NICU), the survival rate of high-risk children is getting higher and higher, especially the survival rate of very low birth weight children and ultralow birth weight children is significantly improved. Along with this, increased attention has been paid to the issue of nerve injury in high-risk children. Studies have found that almost 25% of high-risk children have potential neurodevelopmental problems [3]. Therefore, we should pay attention to the early detection, correct diagnosis, and early intervention of nerve injury in high-risk children.

Objective diagnosis and evaluation methods of nerve injury and nerve development mainly rely on imaging methods such as craniocerebral ultrasound and MRI. Due to technical limitations, early dynamic monitoring is difficult. In recent years, more and more studies have found that [4, 5] although the electroencephalogram (EEG) is not specific for the diagnosis of nerve damage, it is one of the sensitive indicators for judging the function of the brain. Damage usually precedes damage to the nerve structure. Therefore, EEG can be used as an objective evaluation index of early nerve injury and prognosis. However, due to problems such as electroencephalography and analysis techniques, clinical practice has not been widely available. At present, there are many methods to detect neuromotor development, but the standards of various methods are different. It is important to screen out an effective method for assessing the neuromotor development outcomes of high-risk children. Amplitude-integrated electroencephalogram (aEEG) and general movement (GM) quality assessment is a sensitive assessment system for neural injury



that has emerged in recent years, especially for the assessment of neurodevelopmental outcomes in newborns and infants [6]. Technology aEEG is an EEG analysis method, which is mainly analysed by amplitude waves. It has good consistency with conventional EEG, but its advantages in background waveforms or classification are more significant. It can not only be bedside but also long-term, continuous monitoring of brain function makes it easier to judge brain function [7]. GM is a simple and noninvasive operation. Its quality changes when the nervous system is damaged, and various abnormal features appear. It has been proven that GM quality assessment technology has a good early predictive value for cerebral palsy [8]. However, the quality assessment of GMs has limited ability to predict mild neurological dysfunction, and the prediction of neurodevelopmental outcomes in high-risk children is not as good as aEEG [9]. Compared with EEG and GMs, on the one hand, because there are fewer aEEG electrodes, it is convenient for long-term recording, and it is especially suitable for bedside brain function monitoring of high-risk children in NICU. On the other hand, aEEG is easy to operate, intuitive in graphics, and easy to analyse. Many studies have shown that good consistency can be maintained between aEEG and conventional EEG [10, 11]. Goswami et al. [12] showed that according to the acquisition and classification of background waveforms, aEEG has advantages over conventional EEG. Filippi et al. [13] demonstrated in their experiments that they did not find any complications caused by aEEG recording or the use of needle electrodes, indicating that long-term aEEG recording is safe and feasible.

aEEG is a neonatal brain function monitoring device newly introduced in clinical practice in recent years, which can realize continuous real-time monitoring at the bedside, can reflect changes in EEG background activities and epilepsy-like activities, and is very sensitive to newborn brain development and can be diagnosed as early brain injury has become one of the routine monitoring items in the clinical neonatal care unit [14]. Technology aEEG abnormalities suggest a poor prognosis for neurodevelopmental outcomes and can assess neonatal brain function [15]. For these reasons, aEEG has become a research hot spot in the field of neuro-electro physiology in recent years. To better detect the neuromata development of high-risk infants, the validity of amplitude-integrated aEEG in early neuromata development of high-risk infants was investigated. Amplitude-integrated aEEG was used to monitor 100 high-risk infants. At the age of 1 year, the motor scale and validity predictors in the bailey infant and child development scale were used to assess whether or not high-risk children had neurobehavioral dysplasia. Experimental results show that amplitude-integrated aEEG can accurately and effectively predict the outcome of early neuromata development in high-risk infants. Compared with the traditional neurological examination method, it has higher sensitivity, specificity, positive predictive value, and consistency in predicting the outcome of early neuromata development in high-risk

infants, which is suitable for application and promotion in China and has a better application value.

## 2. Materials and Methods

**2.1. The Research Object.** This article selects 100 cases of high-risk children with neurological developmental defects treated in paediatrics from January 2017 to December 2018 as observation objects. There were 60 males and 40 females. There were 57 high-risk infants with a gestational age less than 37 weeks. There were 35 high-risk infants with a gestational age of 37 to 40 weeks. There were 8 high-risk infants with a gestational age greater than 40 weeks. The study was approved by the hospital ethics committee, and all children's families signed informed consent. There were no significant differences in high-risk infants in terms of gender, gestational age, birth weight, 1-minute Apgar score, and delivery methods (down delivery and caesarean delivery), as shown in Table 1.

The inclusion criteria are as follows:

- (1) All are high-risk children, including high-risk factors such as preterm birth, intracranial haemorrhage, asphyxia, ischemic hypoxic encephalopathy, or purulent meningitis;
- (2) There are neurological complications, such as epilepsy, central nervous system infection, and nervous system bleeding;
- (3) The family members of the children were informed and signed the informed consent.

The exclusion criteria are as follows:

- (1) Congenital nervous system diseases or central nervous system malformations, genetic metabolic diseases, central nervous system infections, neonatal hypoglycaemia, respiratory diseases, and other serious system malformations;
- (2) High-risk children with abnormal chromosomes;
- (3) High-risk children who cannot complete aEEG as planned;
- (4) High-risk children who cannot be followed up till 1 year old.

## 2.2. Monitoring Method in aEEG

**2.2.1. Equipment.** The instrument used in this study is a 24-lead digital video EEG and brain function monitor. The EEG box is biologic, and the software system is XLTEK. Disposable electrodes are used. After being cleaned locally with sterile water before tracing, after preparation of the skin with scrub, Ag/AgCl disc electrodes are placed, and an appropriate amount of conductive glue is added. The breathable application fixes the skin and prevents the electrodes from moving. Turn on the power and calibrate the instrument. The electrodes are placed in the central area on both sides (C3 and C4 of the electrode positions according to the international 10–20 electrode placement system [16]). The distance between the two electrode points is 75 mm. The

TABLE 1: Results of general clinical data for high-risk children.

Group	Number of cases	Male/female	Gestational age	Birth weight (g)	1 min Apgar (minutes)	Method of delivery
High-risk child	100	51/49	32.2 $\pm$ 4.5	1912.9 $\pm$ 0.2	8.9 $\pm$ 1.01	30/38
$\chi^2/t$		1.412	1.281	1.512	1.201	1.541
$P$		0.121	0.135	0.101	0.129	0.125

reference electrode is placed on the forehead midline 25 mm from the centre of the top of the head, refer to Figure 1. Record the C3-C4 original EEG signal, and process the signal through the computer's special amplification and filtering algorithm. Finally, amplitude, time compression, and rectification are performed to form the final aEEG signal. The EEG signal is output on the recording paper in a semi-logarithmic form or displayed digitally (0~100  $\mu\text{V}$ ), and the paper feed speed is 6 cm/h. And use artificial visual qualitative analysis.

**2.2.2. Qualitative Interpretation Criteria in aEEG.** According to different aEEG background activity patterns [17] and sleep-wake cycle (SWC) proposed by previous studies, the classification and interpretation were performed.

(1) *Flat Tracing (FT)*. Continuous low-voltage mode below 5  $\mu\text{V}$ , which is in the state of electrical rest, mainly electrical static. The waveform is shown in Figure 2.

(2) *Continuous Extremely Low Voltage (CLV)*. Continuous background low-voltage mode, mostly fluctuating around 5  $\mu\text{V}$  or below. The waveform is shown in Figure 3.

(3) *Burst-Suppression (BS)*. Discontinuous background mode, EEG mode with high amplitude bursts between extremely low voltages (electrical standstill). The waveform is shown in Figure 4.

(4) *Discontinuous Normal Voltage (DNV)*. Discontinuous background mode, the voltage is mostly higher than 5  $\mu\text{V}$ . The waveform is shown in Figure 5.

(5) *Continuous Normal Voltage (CNV)*. Continuous EEG activity, the lower boundary of voltage is 5~10  $\mu\text{V}$  and the upper boundary is 10~50  $\mu\text{V}$ . The waveform is shown in Figure 6.

According to the clinical significance of aEEG, in the above five images, CNV is normal aEEG, DNV is mildly abnormal aEEG, and the rest are severely abnormal aEEG.

SWC is divided into the following three types: broadband is the sleeping period of the newborn, and narrowband is the awakening period. Mature SWCs show most of the aEEG traces analysed. Its amplitude mainly shows that the sine rhythm changes regularly (relatively stable duration and frequency), and each cycle is  $\geq 20$  minutes. If the trace pattern changes periodically, but does not meet the criteria, the mature SWC diagnostic criteria mentioned above are intermediate SWC. If no cycle-like changes occur, it is recorded as no SWC.

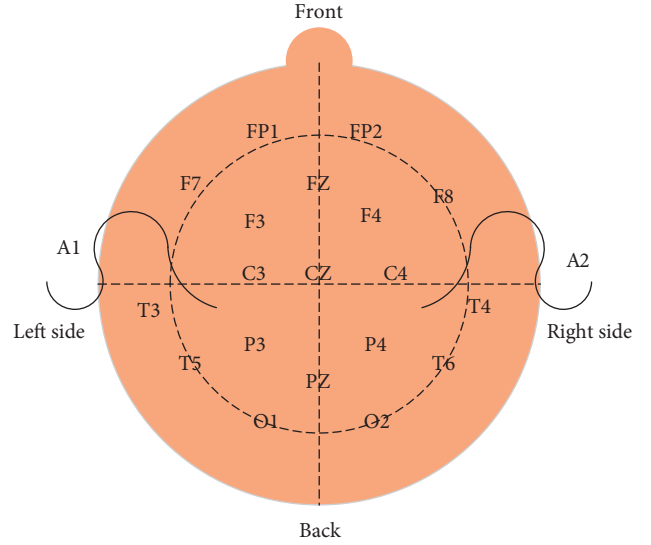
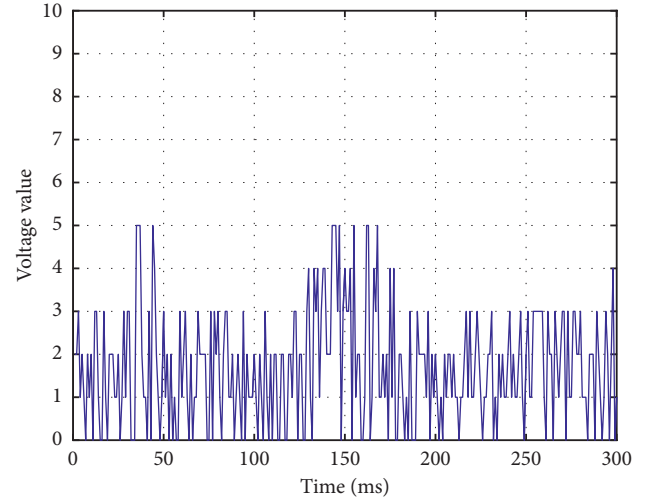


FIGURE 1: Schematic diagram of electrode placement.

FIGURE 2: Continuous low-voltage mode below 5  $\mu\text{V}$ .

**2.2.3. Monitoring Requirements.** Within 1 year after birth, all newborns were examined aEEG by 3 times. The interval between adjacent inspections is at least 3 months. When the two adjacent results are inconsistent, the second result is taken as the final result. All newborns are in a supine position in the crib and dress as little as possible to ensure that the wrist, elbow, knee, and ankle are exposed. The temperature of the room at the time of the check matches the age and clothing of the baby. Try to record when you are awake, and avoid crying, irritability, continuous snoring, or using a soother. At the age of one year, the exercise scale and

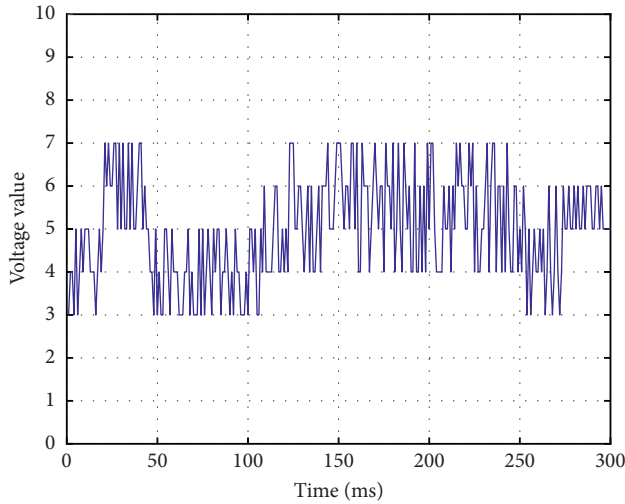
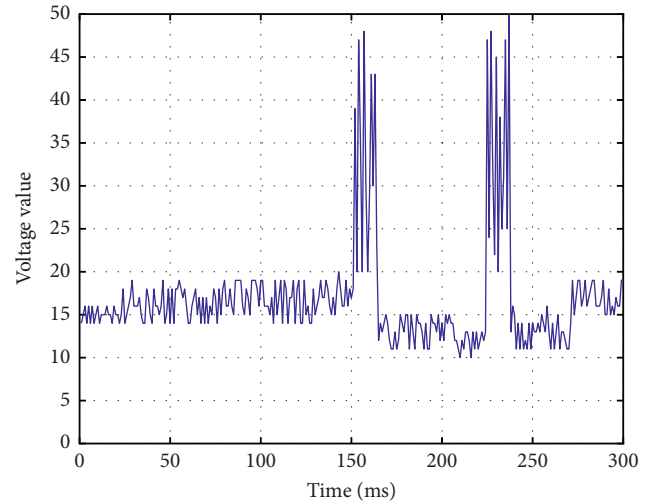
FIGURE 3: Continuous low-voltage mode at around or below  $5\mu v$ .

FIGURE 6: Continuous normal voltage mode.

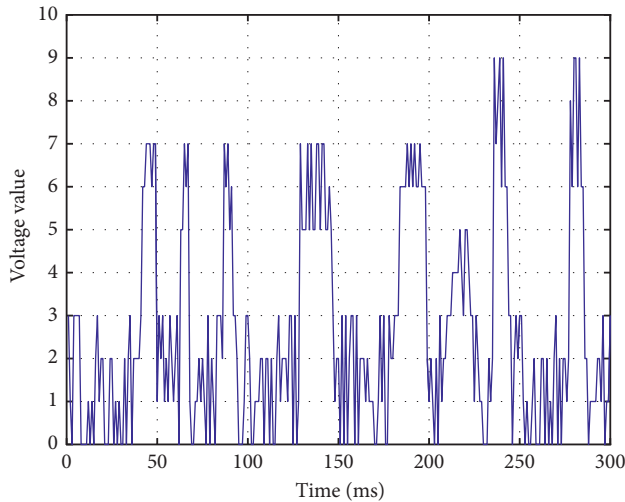
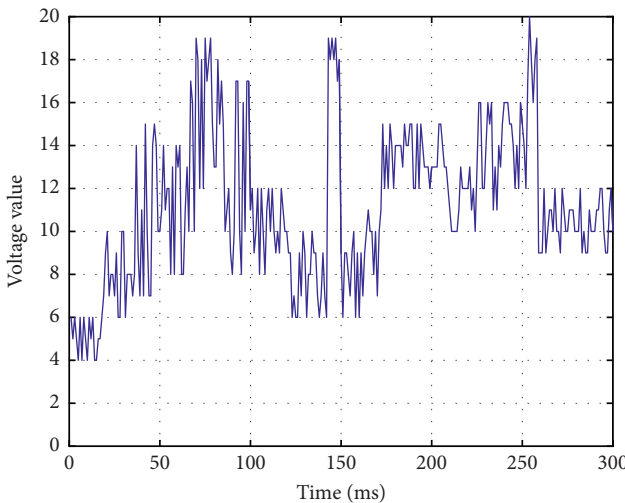


FIGURE 4: EEG patterns between extremely low voltages.

FIGURE 5: Discontinuous voltage mode above  $5\mu v$ .

predictive validity indicators in the bailey infant development scale assess the high-risk newborns for neuro-behavioral abnormalities.

The schematic diagram of aEEG is shown in Figure 7. Electroencephalography is a device that provides a picture of brain activity. When connected to the head, the sensors transmit brain impulses to a machine, which maps them out in lines on an image. Normal, abnormal, or excited brain activity keeps drawing different shapes of lines on the image. Magnetoencephalography is a technique related to electroencephalography, which records electrical impulses in the brain based on changes in the magnetic field in the brain.

#### (1) Bailey infant development scale

High-risk children who entered the follow-up were evaluated at the age of 1 years using the exercise scale in the bailey infant development scale. The exercise scale includes coarse exercise and fine exercise, with a total of 81 items, expressed as psychomotor development index (PDI). PDI is divided into 7 levels: <69 for developmental delay; 70 to 79 for critical status; 80 to 89 for middle and lower; 90 to 109 for middle; 110 to 119 for middle and upper; 120 to 129 for excellent; and >130 points are excellent.

#### (2) Validity prediction indicators

- (a) Sensitivity: the proportion of patients who are positive with a diagnostic test;
- (b) Specificity: the proportion of nonill patients who are negative by diagnostic tests;
- (c) Positive predictive value: of the samples that are positive by the diagnostic test, the proportion of truly sick people is the positive predictive value;
- (d) Negative predictive value: the proportion of truly disease-free children in the samples tested negative by diagnostic tests is the negative predictive value.

#### (3) Statistical analysis

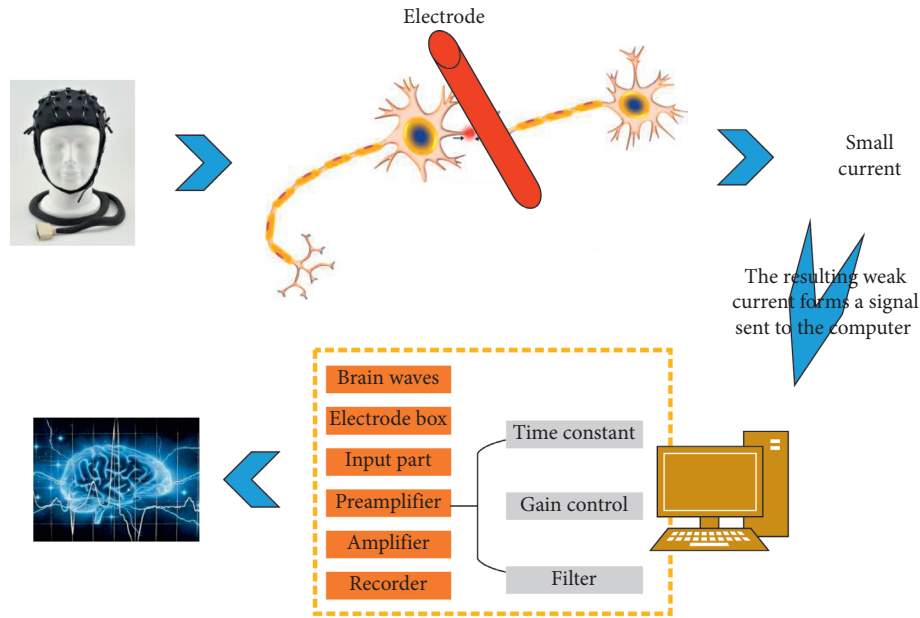


FIGURE 7: The schematic diagram of aEEG.

Data analysis was performed using SPSS 13.0 software. The two assessment methods were compared using sensitivity, specificity, positive predictive value, and negative predictive value. The consistency with the end-stage neurobiological outcome was compared using kappa value for statistical analysis,  $P < 0.05$  indicates that this method is statistically significant.

### 3. Results and Discussion

**3.1. Disease Distribution Monitored by aEEG.** Among the 100 subjects included, the background activity pattern of aEEG was divided into the following: 60 cases were continuous normal activities, 17 cases were mild abnormalities, and 23 cases were severe abnormalities. Of the 23 severe abnormalities, 14 were seizures, 5 were convulsions, 2 were continuous low voltage, and 2 were burst suppression. The results are shown in Figure 8.

A total of 100 high-risk children were included in this article. According to the nature of the case, it can be divided into 23 cases of asphyxia, 21 cases of subventricular haemorrhage, 10 cases of hypoglycaemia, 9 cases of ischemic hypoxic encephalopathy, 20 cases of brain injury, 6 cases of purulent meningitis, 5 cases of intracranial haemorrhage, 3 cases of bilirubin encephalopathy, and 3 cases of neonatal epilepsy, as shown in Figure 9.

**3.2. Evaluation of Exercise Scale in the Bailey Infant Development Scale Monitored by aEEG.** When the study subject was 1 year old, the results evaluated by the exercise scale in the bailey infant development scale were less than 69 points in 3 cases, 70 to 79 points in 8 cases, 80 to 89 points in 16 cases, and 90 to 109 in 21 cases, 110 to 119 points in 35 cases, 120 to 129 points in 12 cases, and >130 points in

5 cases, as shown in Table 2. It can be seen from Table 2 that the lower the score, the more abnormal situations occur.

**3.3. Outcomes and Validity Prediction of Neuromotor Development Monitored by aEEG.** At 12 months of age, 67 cases had normal neuromotor development; 33 cases had delayed neuromotor development. Neuromotor development outcomes are shown in Table 3. In early-risk infants with a gestational age less than 37 weeks, there were 27 abnormal cases of early neuromotor developmental outcomes and 30 normal cases. In high-risk infants with a gestational age between 37 and 40 weeks, there were 3 cases of abnormal neuromotor developmental outcomes. There were 3 cases of abnormal gestational age over 40 weeks. As shown in Figure 10, by plotting the gestational age of neonates and the number of neuromotor development outcomes, it can be seen that the early neuromotor development outcomes of high-risk children are related to the neonatal gestational age. Whether it is a premature infant or a newborn who has been in the mother's body for a long time, there are more abnormalities in early neuromotor developmental outcomes than in newborns born at a normal gestational age.

Table 4 shows the validity prediction results of aEEG monitoring and GMs monitoring for 100 cases. These include sensitivity, specificity, positive predictive value, and negative predictive value.

It can be seen from Table 4 that from the perspective of sensitivity, specificity, positive predictive value, and negative predictive value, the aEEG's validity prediction results are better than GMs. Therefore, the aEEG method for monitoring has better validity; it is predicted that abnormal monitoring of early neuromotor developmental outcomes in high-risk infants can be achieved.

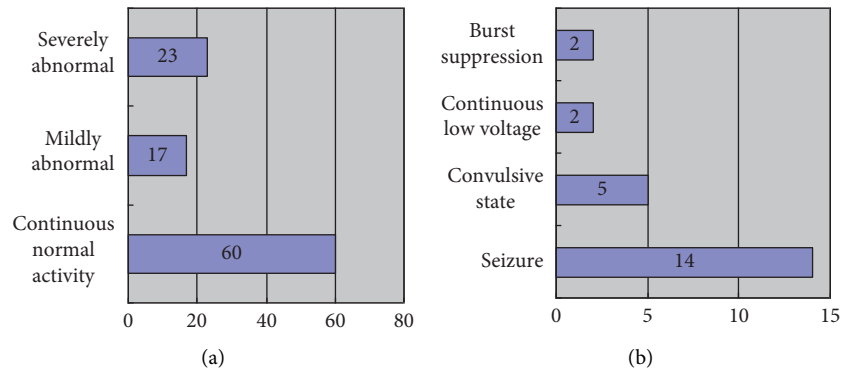


FIGURE 8: aEEG background activity pattern distribution of 100 high-risk neonates. (a) The number of cases with different background activity. (b) The number of cases in different categories of severe abnormalities.

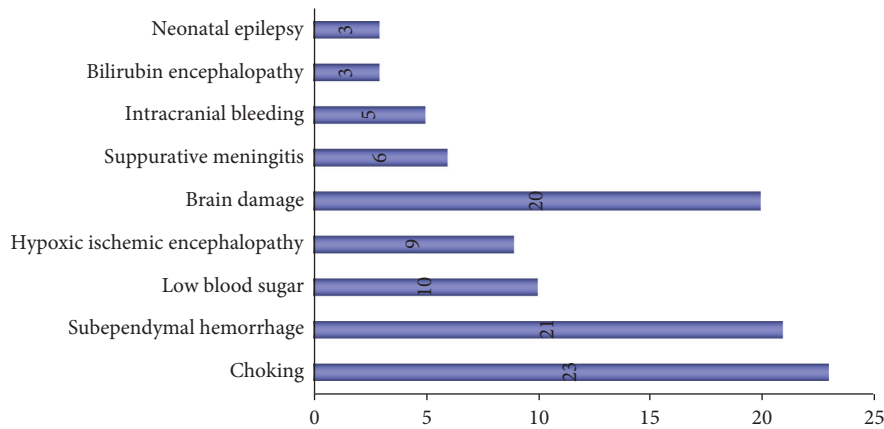


FIGURE 9: Proportion of disease distribution in 100 high-risk neonates.

TABLE 2: Results of 52 neuromotor examination scores and neuromotor development outcomes from 0 to 1 year old.

Score	Number	Torsional phase		Restless movement	
		Abnormal	Normal	Abnormal	Normal
Score <69	3	3	0	3	0
70 <score <79	8	7	1	8	0
80 <score <89	16	11	5	10	5
90 <score <109	21	9	12	10	11
110 <score <119	35	2	33	3	32
120 <score <129	12	1	11	0	12
130 <score	5	0	5	0	5

TABLE 3: Results of neuromotor development outcomes.

	Neuromotor developmental outcomes		Total
	Abnormal	Normal	
Gestational age <37 weeks	27	30	57
37 weeks < gestational age <40 weeks	3	24	35
Gestational age >40 weeks	3	7	8
Total	33	67	100

In the neonatal period, a variety of high-risk factors can cause brain damage, which may cause a variety of neurological sequelae in the growth and development of the newborn, including cerebral palsy, mental retardation, motor retardation, and epilepsy. The early clinical manifestations of brain injury in some newborns are atypical, and early diagnosis is difficult, which often leads to missed diagnosis. When the diagnosis is clear, the best period of treatment and intervention has been missed. Therefore, it is important to determine early brain damage and accurately predict neurodevelopmental outcomes in high-risk neonates.



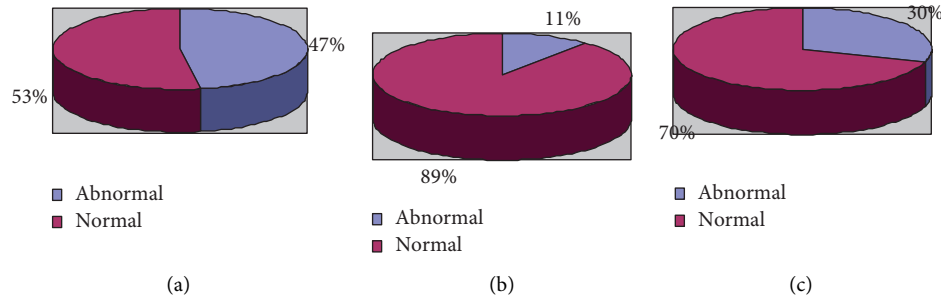


FIGURE 10: Relationship between gestational age and neuromotor development outcomes. (a) Gestational age >37. (b) 37 < gestational age <40. (c) Gestational age <40.

TABLE 4: Results of validity prediction of aEEG and GMs monitoring.

Inspection method	aEEG	GMs
Sensitivity	90.631%	78.381%
Specificity	94.125%	80.187%
Positive predictive degree	87.887%	65.791%
Negative predictive degree	95.529%	88.712%

The results of this study show that aEEG can predict the sensitivity, specificity, positive predictive value, and negative predictive value of high-risk neonatal adverse neurodevelopmental outcomes, which are 90.631%, 94.125%, 87.887%, and 95.529%, respectively. The results are quite higher than those predicted by the GMs quality assessment. This shows that aEEG surveillance has a better predictive value for adverse neuromata development outcomes in high-risk neonates.

#### 4. Conclusions

In recent years, with the gradual improvement of perinatal health care and the advancement of neonatal intensive care and rescue technology, the survival rate of perinatal high-risk neonates has increased significantly. Convulsions, meningitis, respiratory failure, hyperbilirubinemia, and other newborns, even after rescue, whether there are abnormal neurobiological outcomes is worrying. Inspections are becoming increasingly important. In this article, through the study of 100 high-risk children, and by predicting the validity of aEEG monitoring in early neuromotor development outcomes in high-risk children, it is confirmed that there is a correlation between aEEG monitoring and neuromotor development outcomes. When aEEG monitoring is abnormal in high-risk neonates, it is highly suspected that the child's neuromotor developmental outcome is poor, and treatment measures need to be taken as soon as possible. Early intervention reduces the continued damage of the children's brain and nerves and improve the sequelae. The shortcoming of this article is that the follow-up time of the child is short, only corrected to 1 year, and the child needs to be followed up for a longer time to better understand the relationship between aEEG monitoring and neuromotor development outcomes, and to understand the aEEG in predictive value

of monitoring. Our next step is to carry out a longer follow-up to better understand the relationship between aEEG monitoring and neuromata developmental outcomes, and to better understand the predictive value of aEEG monitoring.

#### Data Availability

No data were used to support this study.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest.

#### References

- [1] E. Halstead, N. Ekas, R. P. Hastings, and G. M. Griffith, "Associations between resilience and the well-being of mothers of children with autism spectrum disorder and other developmental disabilities," *Journal of Autism and Developmental Disorders*, vol. 48, no. 4, pp. 1108–1121, 2018.
- [2] T. G. Mazzucchelli, J. Hodges, R. T. Kane et al., "Parenting and family adjustment scales (PAFAS): validation of a brief parent-report measure for use with families who have a child with a developmental disability," *Research in Developmental Disabilities*, vol. 72, pp. 140–151, 2018.
- [3] K. K. Tillman, M. Hakelius, J. Höijer et al., "Increased risk for neurodevelopmental disorders in children with orofacial clefts," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 57, no. 11, pp. 876–883, 2018.
- [4] A. Broncel, R. Bocian, P. Klos-Wojtczak, and J. Konopacki, "Some technical issues of vagal nerve stimulation. An approach using a hippocampal formation theta rhythm," *Brain Research Bulletin*, vol. 140, pp. 402–410, 2018.
- [5] A. R. Clarke, R. J. Barry, S. J. Johnstone, R. McCarthy, and M. Selikowitz, "EEG development in attention deficit hyperactivity disorder: from child to adult," *Clinical Neurophysiology*, vol. 130, no. 8, pp. 1256–1262, 2019.
- [6] L. Hellström-Westas, "Amplitude-integrated electroencephalography for seizure detection in newborn infants," *Seminars in Fetal and Neonatal Medicine*, vol. 23, no. 3, pp. 175–182, 2018.
- [7] L. V. S. Magalhães, M. I. B. Winckler, J. A. Bragatti, R. S. Procianny, and R. C. Silveira, "Early amplitude-integrated electroencephalogram as a predictor of brain injury in newborns with very low birth weight: a cohort study," *Journal of Child Neurology*, vol. 33, no. 10, pp. 659–663, 2018.

- [8] A. K. L. Kwong, T. L. Fitzgerald, L. W. Doyle, J. L. Y. Cheong, and A. J. Spittle, "Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review," *Developmental Medicine & Child Neurology*, vol. 60, no. 5, pp. 480–489, 2018.
- [9] A. J. Spittle, C. Morgan, J. E. Olsen, I. Novak, and J. L. Y. Cheong, "Early diagnosis and treatment of cerebral palsy in children with a history of preterm birth," *Clinics in Perinatology*, vol. 45, no. 3, pp. 409–420, 2018.
- [10] N. Wiwattanadittakul, M. Prust, W. D. Gaillard et al., "The utility of EEG monitoring in neonates with hyperammonemia due to inborn errors of metabolism," *Molecular Genetics and Metabolism*, vol. 125, no. 3, pp. 235–240, 2018.
- [11] W. O. Tatum, G. Rubboli, P. W. Kaplan et al., "Clinical utility of EEG in diagnosing and monitoring epilepsy in adults," *Clinical Neurophysiology*, vol. 129, no. 5, pp. 1056–1082, 2018.
- [12] I. Goswami, L. Bello-Espinosa, J. Buchhalter et al., "Introduction of continuous video EEG monitoring into 2 different NICU models by training neonatal nurses," *Advances in Neonatal Care*, vol. 18, no. 4, pp. 250–259, 2018.
- [13] L. Filippi, P. Fiorini, S. Catarzi et al., "Safety and efficacy of topiramate in neonates with hypoxic ischemic encephalopathy treated with hypothermia (NeoNATI): a feasibility study," *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 31, no. 8, pp. 973–980, 2018.
- [14] S. Backman, I. Rosén, M. Blennow et al., "Swedish consensus reached on recording, interpretation and reporting of neonatal continuous simplified electroencephalography that is supported by amplitude-integrated trend analysis," *Acta Paediatrica*, vol. 107, no. 10, pp. 1702–1709, 2018.
- [15] B. Hüning, T. Storbeck, N. Bruns et al., "Relationship between brain function (aEEG) and brain structure (MRI) and their predictive value for neurodevelopmental outcome of preterm infants," *European Journal of Pediatrics*, vol. 177, no. 8, pp. 1181–1189, 2018.
- [16] B. D. C. F. L. Lopes, C. Sachet, P. Sentenac et al., "Preliminary non-intrusive geophysical electrical resistivity tomography surveys of a mock-up scale monitoring of an engineered barrier system at URL Tournemire," *Geological Society, London, Special Publications*, vol. 482, no. 1, pp. 331–345, 2019.
- [17] E. Shany, N. Taha, E. Benkovich et al., "Association of cerebral activity with MRI scans in infants with neonatal encephalopathy undergoing therapeutic hypothermia," *European Journal of Pediatrics*, vol. 178, no. 6, pp. 851–861, 2019.

## Research Article

# An Informative and Comprehensive Behavioral Characteristics Analysis Methodology of Android Application for Data Security in Brain-Machine Interfacing

Xin Su <sup>1,2</sup>, Qingbo Gong <sup>3</sup>, Yi Zheng <sup>3</sup>, Xuchong Liu <sup>1,2</sup> and Kuan-Ching Li <sup>4</sup>

<sup>1</sup>Hunan Provincial Key Laboratory of Network Investigational Technology, Hunan Police Academy, Changsha, China

<sup>2</sup>Big Data Intelligence Police Hunan Provincial Engineering Research Center, Hunan Police Academy, Changsha, China

<sup>3</sup>Zhejiang Economic Information Center, Hangzhou, China

<sup>4</sup>Department of Computer Science and Information Engineering, Providence University, Taichung, Taiwan

Correspondence should be addressed to Xuchong Liu; liuxuchong@163.com

Received 31 December 2019; Accepted 28 January 2020; Published 10 March 2020

Guest Editor: Chenxi Huang

Copyright © 2020 Xin Su 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.

Recently, brain-machine interfacing is very popular that link humans and artificial devices through brain signals which lead to corresponding mobile application as supplementary. The Android platform has developed rapidly because of its good user experience and openness. Meanwhile, these characteristics of this platform, which cause the amazing pace of Android malware, pose a great threat to this platform and data correction during signal transmission of brain-machine interfacing. Many previous works employ various behavioral characteristics to analyze Android application (or app) and detect Android malware to protect signal data secure. However, with the development of Android app, category of Android app tends to be diverse, and the Android malware behavior tends to be complex. This situation makes existing Android malware detections complicated and inefficient. In this paper, we propose a broad analysis, gathering as many behavior characteristics of an app as possible and compare these behavior characteristics in several metrics. First, we extract static and dynamic behavioral characteristic from Android app in an automatic manner. Second, we explain the decision we made in each kind of behavioral characteristic we choose for Android app analysis and Android malware detection. Third, we design a detailed experiment, which compare the efficiency of each kind of behavior characteristic in different aspects. The results of experiment also show Android malware detection performance of these behavior characteristics combine with well-known machine learning algorithms.

## 1. Introduction

Brain-machine interfaces (BMIs) are a communication technology that link humans and artificial devices through brain signals. Many mobile applications (or apps) are designed as an assistant tool to improve BMIs availability. Android is currently the most popular smart-mobile device platform in the world, occupying about 85% of market share. As of now, there are nearly 3 million Android applications available for downloading from Android official market. The rapid growth of Android and its apps have led to an increased amount of malware. Android malware has resulted in immense losses and security threats of

individuals and businesses, such as data leak during transmission [1]. Developing techniques to analyze and detect Android malware and protect the signal data do not falsify during transmission has become an emergent critical issue.

Traditional Android malware detection approaches focus on signature matching, which is simple and efficiency on detecting known Android malware. However, more and more zero-day malware sprang up all over popular Android markets, which make this approach disable to detect them with high accuracy. Some researchers employ behavioral characteristic with machine learning algorithm to solve limitations of signature

matching approach. These kinds of works can divide into two categories: static behavioral characteristic analysis [2–8] and dynamic behavioral characteristic analysis [9, 10]. For example, Li et al. [2], DroidAPIMiner [3], APK Auditor [4], and Drebin [5] focus on extracting static behavioral characteristics from Android app and then combine them with well-known classification algorithms to detect Android malicious app. However, this kind of approach cannot detect malware with code obfuscation and also cannot capture behavioral characteristics when running the app. Therefore, some researchers employ dynamic behavioral characteristics with machine learning algorithm to detect Android malware. For example, Enck et al. [9] define several privacy information source simultaneously employing dynamic tainting analysis. Su et al. [10] propose an approach which capture HTTP traffic of Android app running to detect Android malware. However, this kind of approach needs to modify code with Android OS version update which lack of scalability.

Based on these previous research works, many kinds of behavioral characteristics have been used to analyze and detect Android malware. However, these are three problems still not solved. First, these works choose several kinds of behavioral characteristics to analyze Android malware but do not explain decisions on these behavioral characteristics. Second, these works do not explain whether behavioral characteristics of Android malware change over time. Third, which kind of behavioral characteristics plays important role during Android malware detection is not illustrated in these works.

To address these problems, we propose a broad and efficient Android malware behavioral characteristics analysis approach under a real-world dataset. First, we extract 11 different static behavioral characteristics and 12 different dynamic behavioral characteristics from each Android app. Second, the proposed method provides an explainable detection, and we explain the reason why we select these kinds of behavioral characteristics. Third, we fed these kinds of behavioral characteristics into several well-known machine learning models to observe which kind of behavioral characteristics would be efficient during Android malware detection. Finally, we compare behavioral characteristics of Android malware change over time.

Based on above description, we summarize the main contributions of this paper as follows:

- (i) We introduce an analysis approach combining static and dynamic behavioral characteristics that is capable of depicting Android malware with comprehensive and accuracy, which is able to provide a high-quality behavioral characteristic dataset for Android malware detection.
- (ii) We provide an explainable analysis for each kind of behavioral characteristics we extracted. This analysis could explicit illustrate each kind of behavioral characteristics playing the role of Android malware detection.
- (iii) In the experiment, we analyze the detection results of each kind of behavioral characteristic and combination of behavioral characteristics. Moreover, we also compare behavioral characteristics of Android malware collected from different periods of time and observe whether behavioral characteristics would change with time change.

The reminder of this paper is organized as follows: the related works have been described in Section 2. We explain reason for each kind of extracted behavioral characteristics in Section 3. The experiment of this paper is shown in Section 4. Finally, we summarize the proposed approach in Section 5.

## 2. Related Work

There are many previous research works paying attention on Android malware detection and behavioral characteristic analysis. In this section, we divide these approaches into two categories, namely, static analysis and dynamic analysis.

**2.1. Static Analysis.** The first category of approach for detecting Android malware was inspired by static program analysis. Several methods have been proposed that statically inspect Android apps and disassemble their code. For example, static approach includes analyzing permission requests for app installation [11] and signature-based detection [12].

The static analysis approach focuses on Android malware analysis based on the static behavioral characteristic, which disassembles the Android app installation file (.apk file) into configuration file and source code. Enck et al. [13] decompile a set of popular Android apps back into dex file and then identify untrusted problems. Yang et al. [14] present a static analysis method called AppContext, which is able to distinguish Android benign app and malware. This work uses contexts that trigger security-sensitive to classify the Android apps. Zhu et al. [15] present an approach that collects requested permission, sensitive API, and system event as behavioral characteristic and use the classification algorithm to build a model for Android malware detection. Mehtab et al. [16] proposes a framework, AdDroid, for analyzing and detecting malicious behavior in Android applications based on various combinations of dynamic behaviors such as accessing network, uploading a file to a remote server, or installing another package on the device. AdDroid employs an ensemble-based machine learning technique where Adaboost is combined with traditional classifiers in order to train a model found on static analysis of Android applications that is capable of recognizing malicious applications. Xie and Li [17] proposed an Android malware detection model based on Bagging-based SVM and static features (permission, intent, and component) extracted from *AndroidManifest.xml*. This work first proposed the *IG-Relief hybrid* selection algorithm to reduce the dimension of the dataset and then used a Bagging-based SVM ensemble classifier trained by the multiple balanced datasets to detect Android malware.

The proposed approach of this paper has been related to the above works; however, our work is different from them.

First, we extract far more behavioral characteristics to depict Android benign app and malware comprehensively. Second, we explain reasons for the extracted behavioral characteristics for the purpose of Android malware detection.

**2.2. Dynamic Analysis.** Unlike static analysis approach, this kind of analysis approach focuses on capturing runtime behavior from the Android app while running on an Android emulator or real devices, for example, dynamic taint analysis [18] and dynamic behavior-based detection [19, 20].

DroidScope [18] allows dynamically monitoring several apps in a protected environment, where the former focuses on taint analysis and the latter enables introspection at different layers of the platform. While both systems provide detailed information about the behavior of apps, they are technically too involved to be deployed on smartphones and detect malware directly. Saracino et al. [19] propose a framework (MADAM) for Android malware detection which monitors apps at the kernel and user level. The MADAM detects system calls at the kernel level and user activity/idleness at the user level to capture the app behavior. Crowdroid [20] is another behavior-based malware detection approach for Android that uses system calls and machine learning techniques. Crowdroid collects information about system calls through a community of users. A lightweight app, installed in the user's devices, monitors system calls (frequency) of running apps and sends them to a centralized server. The server produces feature vectors and applies a K-means clustering to classify the apps as malware or benign app. Hou et al. [21] propose a novel dynamic analysis method named component traversal that can automatically execute the code routines of each given Android app as completely as possible. Ali et al. [22] introduce the use of Fuzzy C-means clustering combined with the generated network traffic in Android malware detection. The features selected in this work were extracted from the network traffic and then were used in Fuzzy C-means clustering algorithm.

However, dynamic taint analysis approaches need to modify OS code, which lacks scalability. Other dynamic analysis approaches focus on few behavioral characteristics, such as network traffic and system call, and cannot depict the comprehensive dynamic behavior. In our work, we extract 12 dynamic behavioral characteristics including network traffic and system call for analysis and we not need modify OS code to track data flow during Android app running.

### 3. Behavioral Characteristic Analysis

In this section, we divide behavioral characteristics of Android app into static and dynamic and extract 11 different kinds of them, respectively. Then, we give a deep analysis of each kind of extracted behavioral characteristic and explain them in detail.

#### 3.1. Behavioral Characteristic Extraction

**3.1.1. Static Behavioral Characteristic Extraction.** Our static behavioral characteristic extraction focuses on

*AndroidManifest.xml* and *class.dex*, and the Android apps were systematically profiled into six types: requested permission, used permission, sensitive API call, action, app component, and intent. Therefore, we extract these types of features from the crawled Android apps, and these types of behavioral characteristics are comprehensive yet a unique representation of Android apps that help determine the typical indications of malicious activity. The detailed extraction process is shown in Figure 1.

To obtain *AndroidManifest.xml* and *class.dex*, we need to decompress the .APK files by *apktool* which is a tool for reverse engineering Android apk files [23]. After decompressing, we start to extracting behavioral characteristic from *AndroidManifest.xml* and *class.dex*. The *AndroidManifest.xml* file contains several configuration information of Android app, such as requested permission and app component. We parse this file by employing *AXMLPrinter2* and *TinyXml* and then extract the behavioral characteristic of requested permission, intent, action, and app component. The *class.dex* file is responsible for storing *Dalvik byte code* which can be converted to *smali code* [24] for better behavioral characteristic extraction. Used permission and sensitive API call can be extracted from *smali code*. In [25], API call and requested permission are matched to discover which permission is used. Therefore, we can obtain the used permission by extracting the API call. We define several customized extraction rules in xml files. For instance, we focus on the APIs provided by the Android framework, so that we can define a rule as List 1 shows.

In Listing 1, the value of regex is defined in regular expression. According to this rule, the decoder can extract APIs such as *android.telephony.TelephonyManager.getSimSerialNumber* from *smali code*. Other behavioral characteristics can be extracted by the similar rules. The node *multiMatch* indicates whether we want the regex to be matched more than once or not.

#### 3.1.2. Dynamic Behavioral Characteristic Extraction.

Unlike static behavioral characteristics, dynamic behavioral characteristics are able to depict running behaviors of the Android app. To extract this kind of behavioral characteristic, we need to install an Android app and operate it in an Android smartphone or emulator. Therefore, we design an automatic tool to install and operate the Android app.

From Figure 2, we find that the proposed approach consists of three modules: first is app execution, which automatically executes Android apps on several real smartphones or Android emulators and outputs the captured dynamic behavioral characteristics (e.g. network traffic and system call) that are generated during execution. This part is also configurable by choosing different rules, e.g., composition of Android apps, execution duration, and execution behavior. The second part is responsible for receiving the captured dynamic behavioral characteristic as input, and then extracting the packet, flow, and system call. The third part is the dynamic behavioral characteristic generator which reads configurable file and generates a dynamic behavioral characteristic for high-level study.



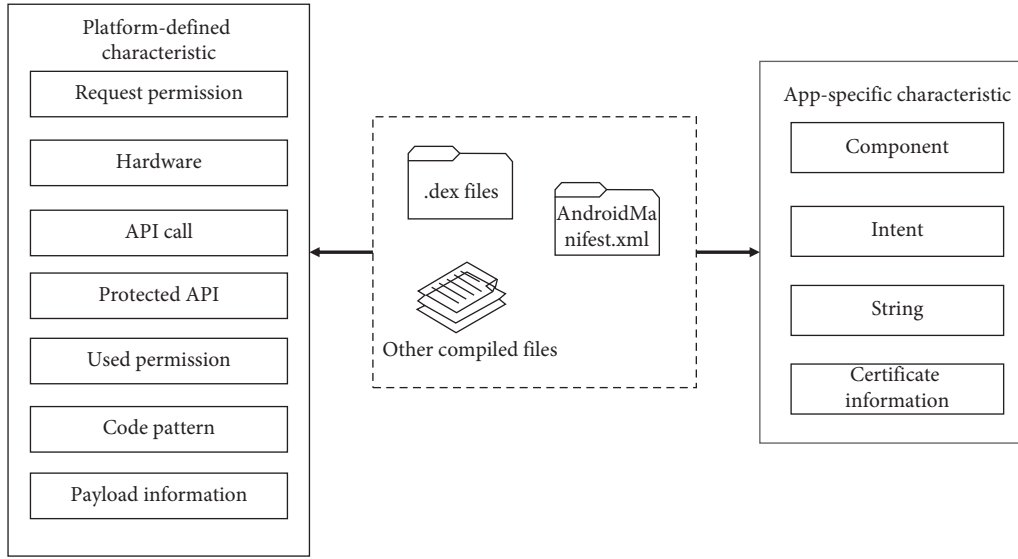


FIGURE 1: Architecture of the static behavioral characteristic extraction.

```

<rule>
  <id>5</id>
  <category>FrameworkAPI</category>
  <description>Extract APIs of Framework</description>
  <regex>Landroid(/w+)+;->w+|Landroid(/w+)+\w+</regex>
  <targetfile>smali</targetfile>
  <multiMatch>true</multiMatch>
</rule>

```

LISTING 1: An example of feature extraction rule.

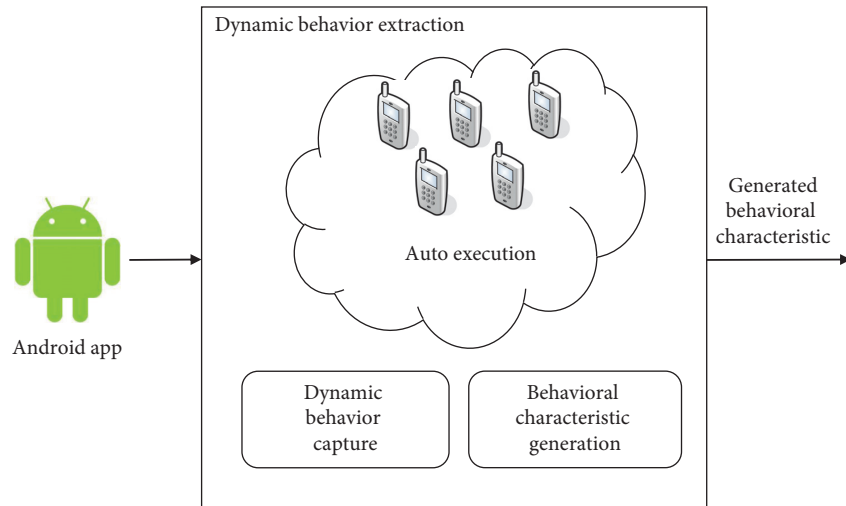


FIGURE 2: Architecture of dynamic behavioral characteristic extraction.

### 3.2. Behavioral Characteristic Analysis

**3.2.1. Static Analysis.** In this section, we explain the behavioral characteristic of the Android platform we mainly used in profiling applications by static analysis. Our static

analysis consists of several parts which focus on the Android manifest file and disassembled dex code of the app. Every app developed for Android must include a manifest file called `AndroidManifest.xml` which provides data supporting the installation and later execution of the application. This

kind of file contains requested permission, app component, intent, and action.

- (i) App component: a typical Android app contains *Activity*, *Service*, *Content provider*, and *Broadcast receiver* and defines these components in the Android configuration file for development purpose. Some malware in the same family may share the same name of the app component.
- (ii) Intent: this component is a lightweight message delivery mechanism, which can delivery messages among different components inside an Android app and between different apps. We found that some specific intents are more frequently defined in the Android malicious app than in the benign one, because Android malware need to delivery particular messages during the launch of malicious activities.
- (iii) Requested permission: an Android app would require predefined permission to access corresponding resources. For example, some of malicious apps would request the *READ\_SMS* permission to read SMS messages in background. Therefore, requested permission is a useful feature for distinguishing Android malware.
- (iv) Hardware: smartphone contains several hardware, such as camera and Bluetooth. If an app wants to request these kinds of hardware, the developer needs to declare them in the Android configuration file. Some malwares would utilize this feature to request hardware to monitor the user, such as recording monitor.

The second part is byte-code reverse from the Android app. This kind of file includes API call, used permission, code pattern, and string. We will illustrate these behavioral characteristics in details as follows.

- (i) API call: application programming interface (or API) is a set of system interface for application call. The API call is able to depict behaviors of Android app for identifying sensitive operations. Android malware may frequently invoke some certain functions to obtain sensitive data from device. Therefore, extracting this kind of feature could be helpful to detect Android malware.
- (ii) Protected API: some API could access the sensitive data or resource of smartphone. We define such kind of API as protected API, which is important for smartphone data security. Based on our observation, a majority of Android malware would invoke such kind of API to leak sensitive data from smartphone. For example, *getSubscriberId* and *getDeviceId* may read *IMSI* and *IMEI* from the smartphone.
- (iii) Used permission: this kind of feature reflects the permission used when Android app running, which can depict the behavior more precisely than *Requested Permission*. Based on the mapping

relationship of API-permission and Intent-permission, we can obtain used permission from API calls and intent information [26].

- (iv) Code pattern: this kind of feature can reflect whether an Android app loads external executable files when it is running. We extract this feature from disassemble code of Android app, such as dynamic load dex files and Linux commands. For example, some Android malware would obtain permission to operate files by executing the *chmod 777* command.
- (v) String: this kind of feature contains URL, IP address, and file path extracted from the disassembled code. Some malwares would leak privacy data by network. For example, a malware family called *Basebridge* leaked sensitive information to <http://b4.7755.org:8088> [27].

Except Android app configuration file and disassembled code, we also extract behavioral characteristic from certification and file suffix.

- (i) Certificate information: this behavioral characteristic indicates the author of the Android app. The app developer uses a secret key to sign the apk file when the app is released. The certificate information contains several developer information such as country, e-mail, and organization, which could differentiate the developers. Kang et al. detect and classify Android malware using this information [28].
- (ii) Payload information: this feature indicates the file category contained in the apk file. Some malwares may contain extra dex files in their install file and load them when the app is running.

After 11 kinds of static behavioral characteristics have been explained, we conduct a depth analysis of these characteristics. We found that the extracted characteristics can be divided into two categories based on their providers. The first category is the *platform-defined characteristic*, which represents these kinds of behavioral characteristics provided by the Android platform. Another category is the *app-specific characteristic*, which represents behavioral characteristic defined by the Android app developer. Based on this categorization, we can depict the behavioral characteristic of Android malware in a fine-grained manner. Tables 1 and 2 show several examples for this categorization. The Android app data set we extract and category behavioral characteristic are described in Section 4.

**3.2.2. Dynamic Analysis.** Dynamic analysis represents behavioral characteristic captured during the Android app running time, which mainly includes network traffic and system call. Network traffic mainly depicts behaviors when the Android app accesses Internet. We divide this kind of behavior into three kinds of behavioral characteristics:

TABLE 1: Some behavioral characteristic instances in *platform-defined characteristic*.

Behavioral characteristic	Instance
Requested permission	ACCESS_GPS, GET_TASK, WAKE_LOCK
Hardware	CAMERA, NFC, GPS
API call	util.log.w, Dialog.show, Uri.prase
Protected API	getDeviceId, sendSMS, getWifiState
Used permission	INTERNET, SEND_SMS, READ_CONTACT
Code pattern	MessageDigest, loadLibrary, pathClassLoader
Payload information	.MF, .RSA, .jpg

- (i) Quantitative behavioral characteristic: this category measures and compares the volumes of traffic across malware and benign apps. When the malware communicates with the malicious servers, they request update commands and leak private information with a fixed format. Also, the malware do not generate large traffic volumes to avoid detection by antivirus scanners or intrusion detection systems. Therefore, a malware trace might contain many flows with similar traffic size.
- (ii) Timing-based behavioral characteristic: the second type of feature category is the time-based behavioral characteristics, which try to capture the duration of activity of the Android app.
- (iii) Semantic behavioral characteristic: more than 90% of apps run over the HTTP protocol and that 93% of malware samples use HTTP to receive commands from their C&C servers that can be found in the collected Android apps. Thus, considering this scenario, the network behavior can be correlated to the semantics of the different HTTP requests and responses. The network behavior changes with respect to the HTTP method, contacted hosts, URL paths or queries, and so on.

The Android Operation System is based on Linux kernel. In Linux, a system call is how a program requests a service from the operating systems kernel. System calls provide useful functions to application programs like network communication, file management, or process-related operations. When an app from the user space makes a request to the Operation System, the request goes through glibc library, system call interface, kernel, and finally to the hardware. Functions like getpid(), open(), read(), chmod(), and socket() are some of the functions that glibc provides for apps to invoke a system call.

## 4. Experiment

In this section, we focus on evaluating efficiency of the extracted behavioral characteristic. To achieve this, we conduct the following experiments:

- (i) Performance of single behavioral characteristic: in this experiment, we evaluate the performance of

each kind of behavioral characteristic and observe which kind of behavioral characteristic combined with machine learning algorithm could be more useful in Android malware detection.

- (ii) Performance of behavioral characteristic combination: in this experiment, we combine several behavioral characteristics into different groups based on their provider. Then, which combination could be more efficient during Android malware detection is evaluated.
- (iii) Behavioral characteristic persistence: in this experiment, we evaluate the persistence of static and dynamic behavioral characteristics on classification performance with two datasets collected in different time periods.

**4.1. Dataset.** Before evaluating our proposed approach, we first introduce the dataset we used. The dataset consists of three parts: first dataset consists of Android app obtained from the official market and covers the most popular Android app from each category. We consider this kind of dataset as a benign app because of strict audit mechanism of the official market. Second dataset consists of several well-known malware dataset, such as Drebin [5], Android Malware Genome Project [29], and the Contagio Community [30]. The third dataset consists of Android apps downloaded from several unofficial markets, and we consider this kind of dataset as an unknown type of Android app. At last, we totally get 3,986 Android benign apps, 3,986 Android malicious apps, and 1,515 unknown Android apps, and extract their behavioral characteristics storage in the database.

**4.2. Performance of Single Behavioral Characteristic.** In Section 3, we extract 11 static behavioral characteristics and 12 dynamic behavioral characteristics from the Android app. However, single dynamic behavioral characteristics are common for the Android benign app and malware. Therefore, we choose single behavioral characteristics to classify the Android benign app and malware. Table 3 shows the performance of Android malware classification based on a single kind of static behavioral characteristics.

**4.2.1. Static Behavioral Characteristic.** From Table 3, we find most of single behavioral characteristic cannot achieve high performance, and the performance is differentiated by combining with different machine learning algorithms. *Requested permission* combined with *Random forest* is able to achieve the highest performance among all kinds of behavioral characteristics. In summary, the results of this experiment demonstrate that single behavioral characteristic cannot classify Android benign app and malware well.

**4.2.2. Dynamic Behavioral Characteristic.** Next, we will compare each dynamic behavioral characteristic between Android benign app and malware. The first category is the

TABLE 2: Some behavioral characteristic instances in *app-specific characteristic*.

Behavioral characteristic	Instance
App component	com.google.search, com.eguan.state, com.google.update
Intent	PHONE_STATE, MAIN, SIG_STR
String	map.Google.com, http://www.umeng.com, media. admob.com
Certificate information	2b7172a335b66873dc793af3fe5c3fc6d8. . . , 5fb16d12bc8a36b9071907bc6e042840c2. . .

TABLE 3: Performance of single behavioral characteristic.

Behavioral characteristic	Algorithm	TP	FP	Precision	Recall	F-measure	ROC area
Code	SVM	0.685	0.318	0.686	0.685	0.685	0.684
	KNN	0.692	0.324	0.692	0.692	0.689	0.734
	Random forest	0.694	0.323	0.694	0.694	0.69	0.735
Hardware	SVM	0.548	0.535	0.563	0.548	0.411	0.507
	KNN	0.548	0.534	0.564	0.548	0.413	0.543
	Random forest	0.548	0.534	0.564	0.548	0.413	0.547
Indent	SVM	0.689	0.489	0.762	0.689	0.723	0.69
	KNN	0.836	0.184	0.849	0.836	0.833	0.885
	Random forest	0.84	0.181	0.854	0.84	0.837	0.892
Requested permission	SVM	0.873	0.137	0.876	0.873	0.872	0.868
	KNN	0.936	0.064	0.936	0.939	0.936	0.972
	Random forest	0.937	0.063	0.938	0.937	0.937	0.975
Suspicious API	SVM	0.753	0.0257	0.752	0.753	0.752	0.749
	KNN	0.782	0.228	0.782	0.782	0.781	0.861
	Random forest	0.779	0.231	0.779	0.779	0.777	0.861
Used permission	SVM	0.837	0.177	0.841	0.837	0.835	0.83
	KNN	0.864	0.138	0.864	0.864	0.864	0.942
	Random forest	0.867	0.133	0.867	0.867	0.867	0.943
API	SVM	0.843	0.125	0.859	0.843	0.851	0.905
	KNN	0.831	0.134	0.895	0.831	0.862	0.919
	Random forest	0.904	0.092	0.916	0.904	0.91	0.927
Payload	SVM	0.555	0.531	0.755	0.555	0.407	0.512
	KNN	0.699	0.313	0.698	0.699	0.697	0.774
	Random forest	0.7	0.311	0.699	0.7	0.698	0.779
Cert	SVM	0.744	0.344	0.796	0.744	0.769	0.8
	KNN	0.764	0.319	0.758	0.764	0.761	0.786
	Random forest	0.807	0.183	0.806	0.807	0.807	0.834
String	SVM	0.724	0.269	0.731	0.724	0.727	0.76
	KNN	0.544	0.544	0.544	0.544	0.544	0.544
	Random forest	0.763	0.32	0.736	0.762	0.763	0.784
Component	SVM	0.839	0.164	0.84	0.839	0.839	0.851
	KNN	0.85	0.192	0.863	0.85	0.856	0.872
	Random forest	0.893	0.117	0.893	0.847	0.87	0.875

*quantitative behavioral characteristic* which includes the *number of packets*, *number of bytes*, *number of received packets*, *average bytes of received packets*, *average size of packets*, and *in/out ratio*. The detail results of comparison are shown in Figures 3–5.

In Figure 3(a), 80% of malware flows containing about 10 packets or less have been observed; only 30% of benign app flows achieve this number. In Figure 3(b), 80% of malware flow size reaching 1000 bytes or less has been observed; only 30% benign app flow size achieves this number. Because benign apps have rich functionality, their network activities include, text chat, videos, and image downloading. Therefore, these network activities are expected to have a variable number of packets due to the variable size of the data involved. On the contrary, malware focuses on sending out

private data out, which is usually in standard size regardless of the smartphone in use, and hence, it is expected that the number of packets per flow is similar across the multiple malicious apps.

In Figure 4(a), benign app flows contain more received packets per flow than malware. Also, Figure 4(b) shows the average packet size in each flow. The relative difference in packet sizes is clear from these results. Because benign apps may receive large size files from servers, due to the limitation of a packet length, the large size file is split into several segments, thereby increasing the number of received packets; whereas in malware apps, the received command packets are usually small-sized packets.

In Figure 5(a), almost 50% of malware packet sizes are in the range from 101 bytes to 200 byte, and less than 30% of

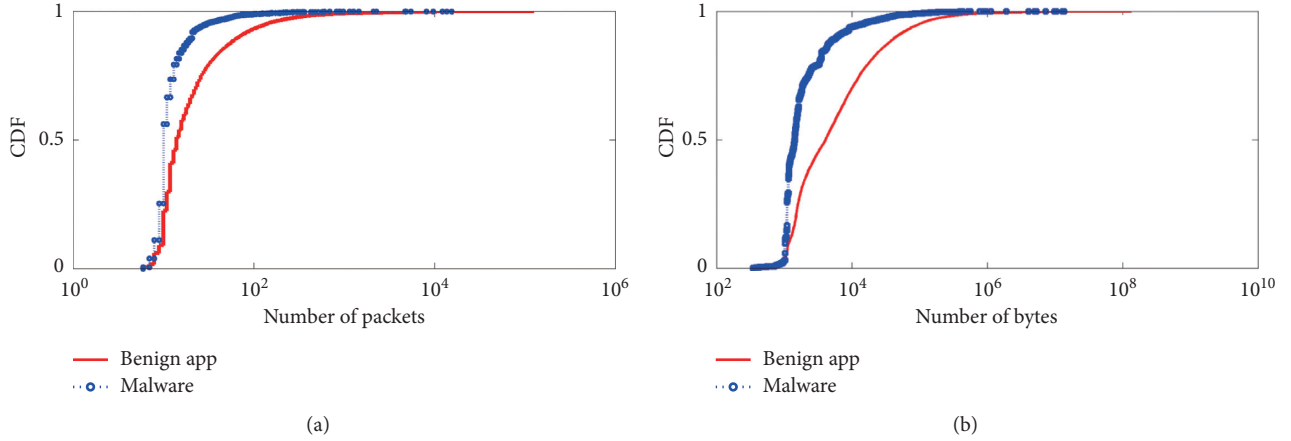


FIGURE 3: Comparison of the numeric aggregates of transfer size. (a) Number of packets. (b) Number of bytes.

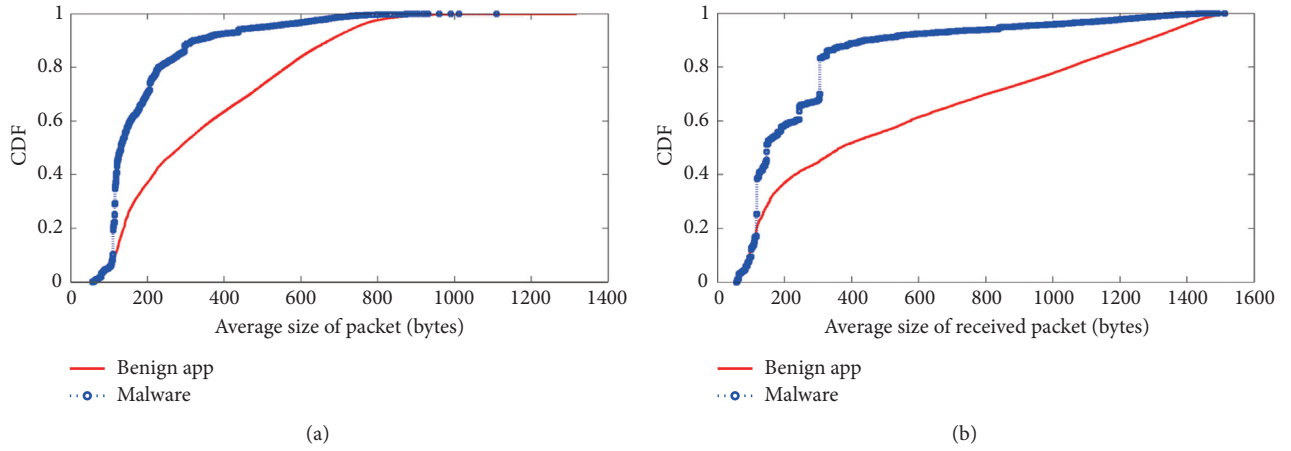


FIGURE 4: Comparison of the numeric aggregates of inward flow. (a) Number of received packets. (b) Average size of received packets.

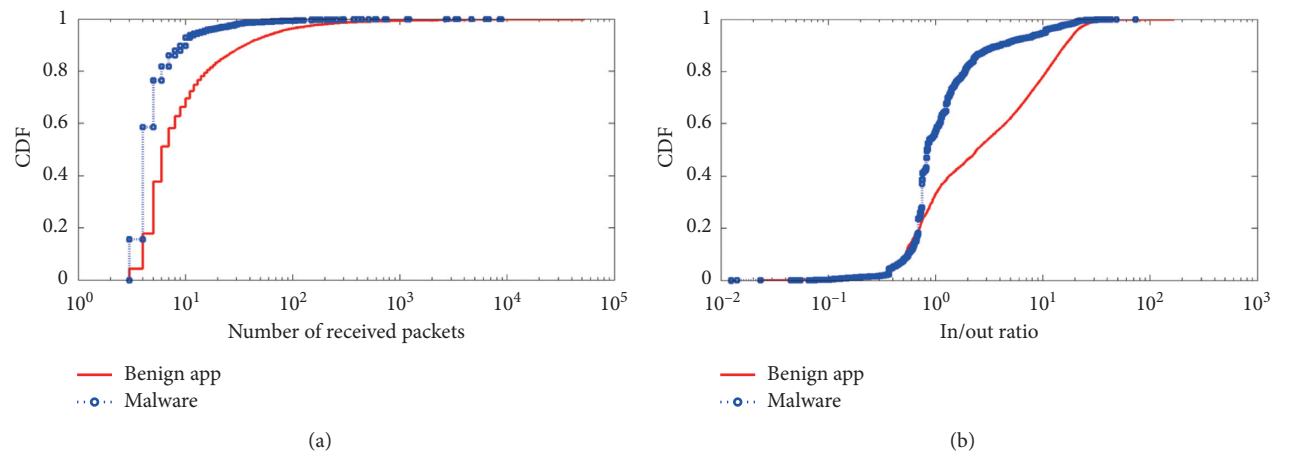


FIGURE 5: Data sizes and relative traffic ratios of flows. (a) Average size of packets. (b) In/out ratio.

benign app packets sizes fall in this range. In Figure 5(b), for benign apps, note that about 20% of the ratios of the incoming and outgoing traffic are lesser than 1, which shows that this traffic contains more sent data than received data. In

benign apps, the packet size is not constrained as the user can download or upload data of any size. For malware apps, the command packets typically have a smaller size due to the compact nature of the Botnet protocol communication.



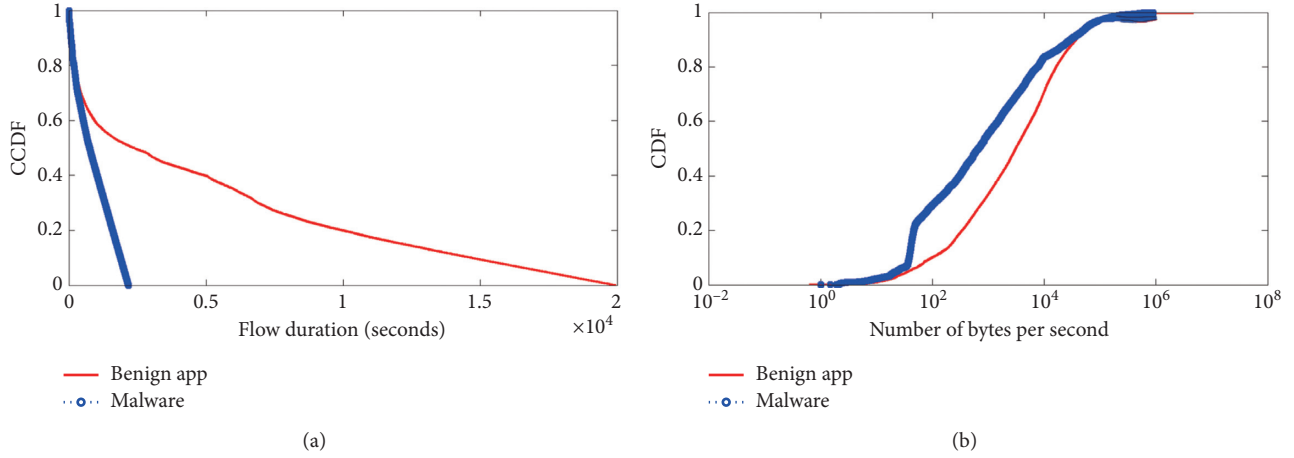


FIGURE 6: Flow duration comparison. (a) Flow duration. (b) Number of bytes per second.

The second category of this kind behavioral characteristic is the *Timing-Based behavioral characteristic* which includes *flow duration* and *the number of bytes per second*. The detail results of comparison are shown in Figure 6.

The duration of flow is typically the TCP session length, which represents the amount of time an app requires to conduct its network functions with its destination server. Figure 6(a) shows a CCDF plot of the HTTP flow duration in benign apps, malware apps and ad libraries. We notice that, for benign app and ad libraries, more than 40% flows have a duration shorter than 2 seconds. This is because many flows in benign apps and ad libraries only transfer small data like text or small image files for which the duration is short. This figure also shows that benign apps account for a larger proportion of long duration flows. Figure 6(b) shows the CCDF of the number of bytes per second in the benign app, malware app, and ad library. There is a clear gap that can be observed, i.e., 70 bytes/s as compared with 1200 bytes/s, between the benign app and malware, respectively, which demonstrates that the malware communication is lightweight, stealthy, and ends in a short time period.

The third category is the *semantic behavioral characteristic* which includes *length of URI per GET/POST request*, *length of page per GET/POST request*, and *length of parameter per GET/POST request*. The detail results of comparison are shown in Figures 7–9.

The length of URI per GET/POST request shows the number of resources requested by the app. Benign apps may request various kinds of files, whereas malware usually request commands like update or leak private data out in a fixed format.

The length of page represents the paths visited by the app to obtain the resources, and typically, the same HTTP request contains more than one resource path. Benign apps usually request multiple resources as they try to maximize the user experience, and on the other hand, malicious apps request a small number of resources.

The GET/POST parameter is a query string, which is the part of a uniform resource locator (URL) that contains data to be passed to servers. Because benign apps have various

types, and they may send requests to servers with variable parameter formats. However, the malware asks commands like update and leak private data out with fixed parameter format, and usually the parameter lengths are fixed within a small statistical threshold.

In summary, single behavioral characteristic of network traffic is able to distinguish the Android benign app and malware partly. However, the Android malware can forge such single behavioral characteristic easy to avoid detection. Therefore, we conduct the second experiment and utilize the characteristic combination to depict Android malware behaviors.

#### 4.3. Performance of Behavioral Characteristic Combination.

We have evaluated a single kind of behavioral characteristic that cannot obtain good performance during the Android malware classification in Section 4.2. In this experiment, we try to combine several kinds of behavioral characteristics based on their provider to evaluate their performance. To achieve this goal, we divide the 11 static behavioral characteristics into two sets, namely, *platform-defined* and *app-specific*. *Platform-defined* represents the behavioral characteristic defined by the Android platform, and *app-specific* represents appSpecific defined by each Android app. Meanwhile, we also divide 12 dynamic behavioral characteristics into 4 sets, namely, *quantitative behavioral characteristic*, *timing-based behavioral characteristic*, *semantic behavioral characteristic*, and *system call*. Moreover, we combine all static and dynamic behavioral characteristics into two behavioral characteristic sets to evaluate its performance, respectively. The detailed results are shown in Tables 4 and 5.

From Table 4, we can find that employing the behavioral characteristic of *app-specific* can obtain better performance than *platform-defined*, because behavioral characteristic of *app-specific* is defined by the Android app which can depict the unique behavioral characteristic. Moreover, combining both kinds of behavioral characteristics can obtain the best performance among the three kinds of behavioral

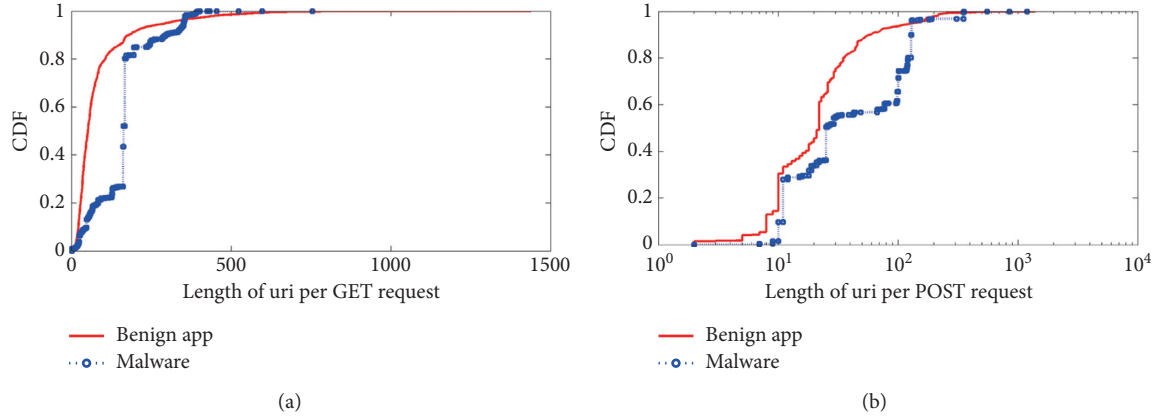


FIGURE 7: URI length of HTTP request. The length of URI per (a) GET request and (b) POST request.

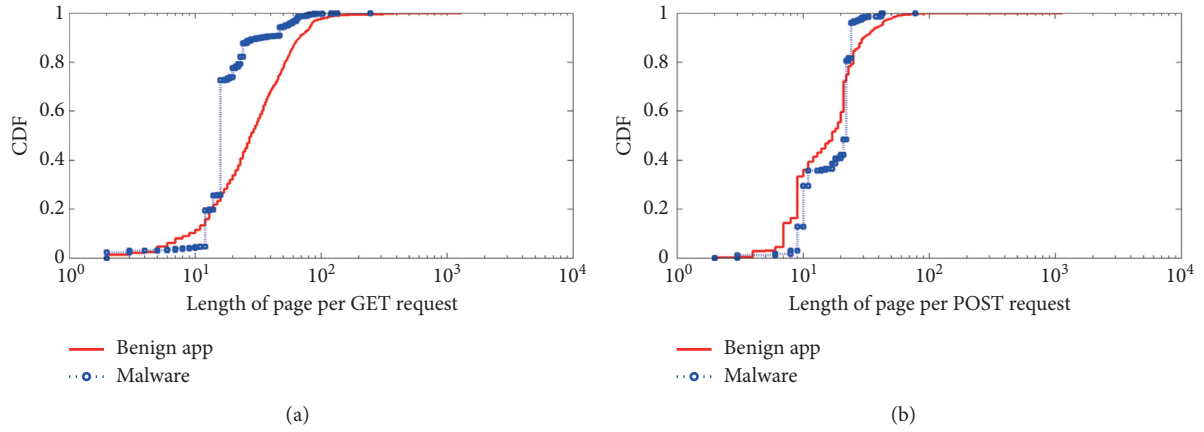


FIGURE 8: Page length of HTTP request. The length of page per (a) GET request and (b) POST request.

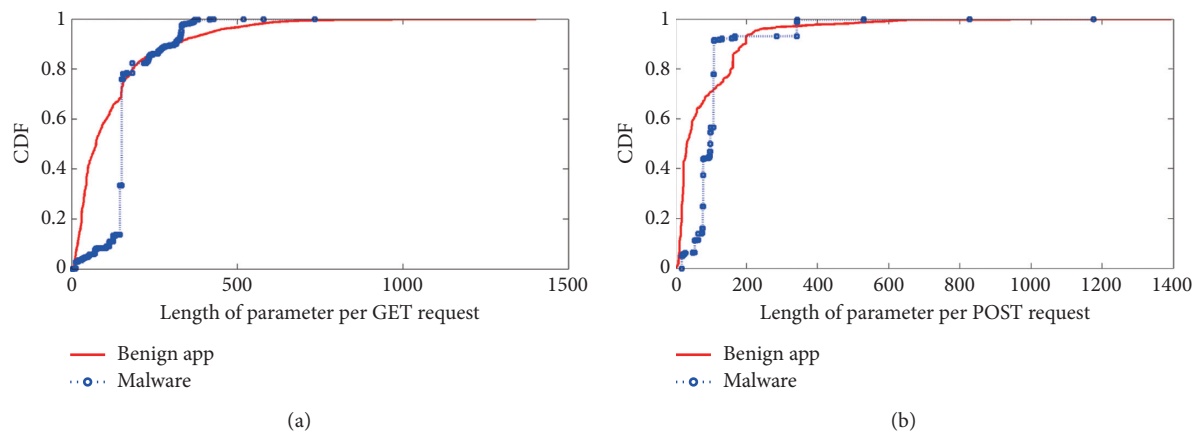


FIGURE 9: Parameter length of HTTP request. Length of parameter per (a) GET request and (b) POST request.

characteristics. This result demonstrates that the extract comprehensive behavioral characteristic can be helpful for the Android malware classification.

From Table 5, we can find that any kind of dynamic behavioral characteristic cannot obtain a good

performance, because these kinds of behavioral characteristics may be overlapped by the Android benign app and malware. However, combining these kinds of behavioral characteristics can achieve better performance, due to its uniqueness.

TABLE 4: Performance of static behavioral characteristic combination.

Behavioral characteristic set	Algorithm	TP	FP	Precision	Recall	F-Measure	ROC area
Platform-defined	SVM	0.887	0.123	0.892	0.887	0.886	0.88
	KNN	0.94	0.06	0.94	0.94	0.94	0.975
	Random forest	<b>0.942</b>	<b>0.058</b>	<b>0.942</b>	<b>0.942</b>	<b>0.942</b>	<b>0.981</b>
App-specific	SVM	0.898	0.102	0.899	0.898	0.898	0.895
	KNN	0.961	0.039	0.961	0.961	0.961	0.981
	Random forest	<b>0.961</b>	<b>0.039</b>	<b>0.961</b>	<b>0.961</b>	<b>0.961</b>	<b>0.99</b>
All	SVM	0.905	0.095	0.909	0.905	0.904	0.899
	KNN	0.962	0.038	0.962	0.962	0.962	0.981
	Random forest	<b>0.971</b>	<b>0.029</b>	<b>0.971</b>	<b>0.97</b>	<b>0.971</b>	<b>0.992</b>

TABLE 5: Performance of dynamic behavioral characteristic combination.

Behavioral characteristic set	Algorithm	TP	FP	Precision	Recall	F-Measure	ROC area
Quantitative	SVM	0.914	0.086	0.914	0.914	0.914	0.905
	KNN	0.909	0.091	0.908	0.928	0.908	0.889
	Random forest	<b>0.927</b>	<b>0.073</b>	<b>0.927</b>	<b>0.927</b>	<b>0.927</b>	<b>0.916</b>
Timing	SVM	0.875	0.125	0.858	0.875	0.866	0.5
	KNN	0.875	0.125	0.858	0.875	0.866	0.5
	Random forest	<b>0.881</b>	<b>0.119</b>	<b>0.881</b>	<b>0.881</b>	<b>0.881</b>	<b>0.59</b>
Semantic	SVM	0.91	0.09	0.91	0.91	0.91	0.899
	KNN	0.902	0.092	0.902	0.902	0.902	0.881
	Random forest	<b>0.917</b>	<b>0.093</b>	<b>0.917</b>	<b>0.917</b>	<b>0.917</b>	<b>0.903</b>
System call	SVM	0.855	0.145	0.856	0.856	0.856	0.879
	KNN	0.812	0.188	0.815	0.815	0.815	0.846
	Random forest	<b>0.871</b>	<b>0.129</b>	<b>0.871</b>	<b>0.87</b>	<b>0.871</b>	<b>0.892</b>
All	SVM	0.945	0.055	0.945	0.938	0.948	0.966
	KNN	0.947	0.053	0.947	0.942	0.945	0.961
	Random forest	<b>0.965</b>	<b>0.035</b>	<b>0.979</b>	<b>0.977</b>	<b>0.979</b>	<b>0.986</b>

TABLE 6: Performance of the dynamic behavioral characteristic combination.

Dataset	Description
Pass training set	4,403 Android benign apps from Google play
Pass testing set	3,982 Android malware (from Drebin, Android Genome project)
New training set	4,000 Android apps from third-party markets
3,948 Android malware	4,455 Android benign apps from Google play
New testing set	4,000 Android apps from third party markets

**4.4. Behavioral Characteristic Persistence.** For evaluating the behavioral characteristic persistence of Android app over time, this experiment first divides our Android app dataset into two categories based on different periods of time. The detailed information is listed in Table 6.

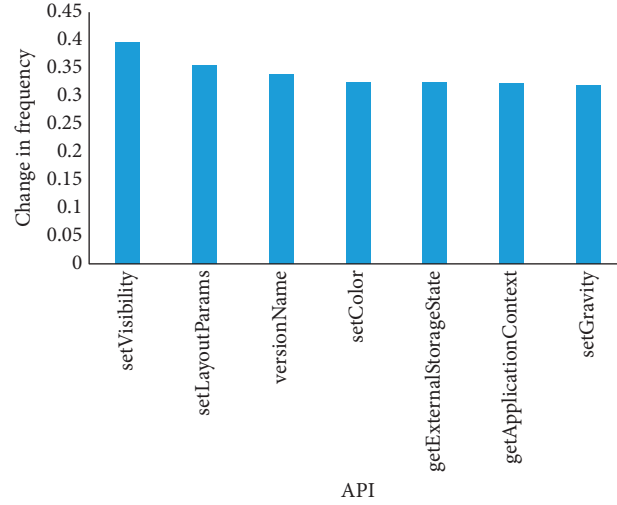
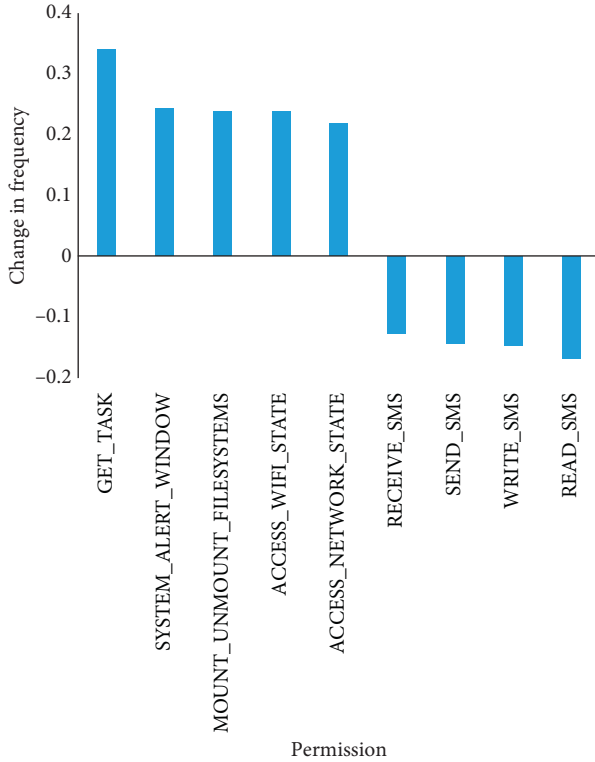
From Table 6, the kind of *Pass* dataset represents these apps released before 2015, and *New* dataset represents these apps released after 2015. In this experiment, we choose three kinds of static behavioral characteristics as an example to evaluate persistence of behavioral characteristics. The behavioral characteristic of *API* persistence is shown in Figure 10.

From this figure, we find that Android malware from *New* dataset calls more APIs of UI elements than the *Pass* dataset, such as visibility setting, layout parameter setting, and color. Because majority of Android malwares embed malicious code into the benign app, these benign apps

contain more UI elements with change of time. Another reason for this situation is that Android malware needs to avoid detection and pretends as a benign app by design similar to UI. Figure 11 shows the variation of permission.

From Figure 11, we find three main differences between *Pass* and *New* datasets. First, permissions for accessing Internet and phone state are the most requested permissions. Second, comparing Android apps from the *New* dataset with those from the *Pass* dataset, we find that the frequency of SMS-related permission reduces. Third, the frequency of related permissions of the device from Android apps of the *New* dataset are more than that of the *Pass* dataset, such as *SYSTEM\_ALERT\_WINDOW*. Figure 12 shows the variation of intent.

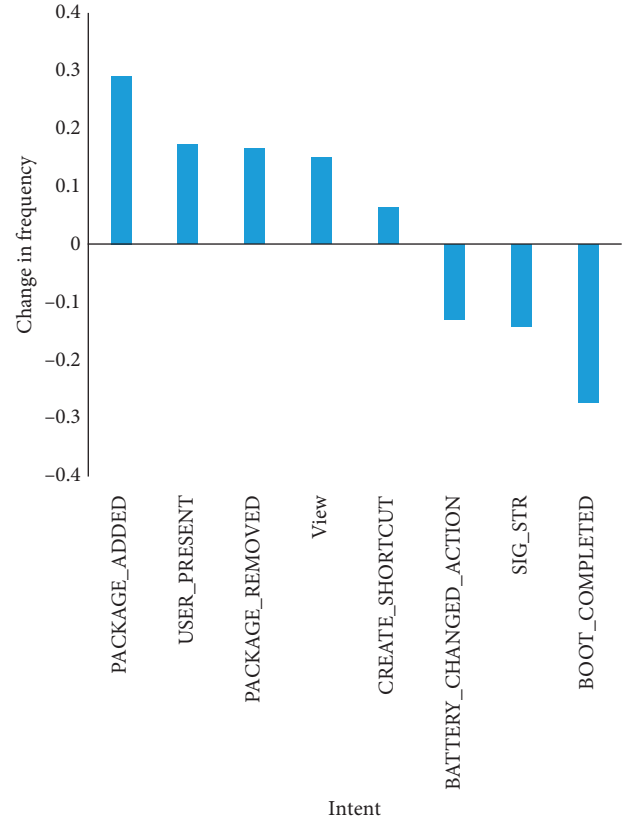
From the change of intent behavioral characteristics between *New* and *Pass* datasets, we obtain two main conclusions. First, the frequency of intent of monitor phone

FIGURE 10: API call variation frequencies between *Pass* and *New* datasets.FIGURE 11: Permission variation frequencies between *Pass* and *New* datasets.

signal and battery has been reduced. Second, the Android malware from the *New* dataset employs more intent of visibility setting.

## 5. Conclusion

In this paper, we introduce an informative and comprehensive Android malware behavioral characteristics analysis methodology, which aims to detect malicious activities during the data transmission of brain-machine interfacing.

FIGURE 12: Intent variation frequencies between *Pass* and *New* datasets.

To achieve this goal, we first extract two categories of behavioral characteristics from the Android app, namely, static behavioral characteristics and dynamic behavioral characteristics. These kinds of extracted behavioral characteristics are able to cover a majority of Android app behaviors. Then, we explain the role of these kinds of behavioral characteristics play in the Android app analysis and Android malware detection. In the experiment, we design three kinds of

evaluations which aim to verify the performance and persistence of extracted behavioral characteristics.

## Data Availability

The data used to support the findings of this study are available from the author upon request (suxin@hnu.edu.cn).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the Research Foundation of Education Bureau of Hunan Province, China (no. 18B548), in part by the Science and Technology Project of Hunan Province of China (no. 2017SK1040), by the Science and Technology Projects of Hunan Province (no. 2018JJ2108), and by the Open Research Fund of Hunan Provincial Key Laboratory of Network Investigational Technology (no. 2017WLZC006).

## References

- [1] W. Liang, M. Tang, J. Long, J. Xu, and K. C. Li, "A secure FaBric blockchain based data transmission technique for industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 9, 2018.
- [2] J. Li, L. Sun, Q. Yan, Z. Li, W. Srisa-an, and H. Ye, "Significant permission identification for machine-learning-based android malware detection," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3216–3225, 2018.
- [3] Y. Aafer, W. L. Du, and H. Yin, "DroidAPIMiner: mining api-level features for robust malware detection in android," *SecureComm*, vol. 127, 2013.
- [4] K. A. Talha, D. I. Alper, and C. Aydin, "Apk auditor: permission-based android malware detection system," *Digital Investigation*, vol. 13, pp. 1–14, 2015.
- [5] D. Arp, M. Spreitzenbarth, M. Hubner, H. Gascon, and K. Rieck, "Drebin: effective and explainable detection of android malware in your pocket," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, San Diego, CA, USA, February 2014.
- [6] N. Xie, X. Wang, W. Wang, and J. Liu, "Fingerprinting Android malware families," *Frontiers of Computer Science*, vol. 13, no. 3, pp. 637–646, 2018.
- [7] W. Liang, K. C. Li, J. Long, X. Kui, and A. Zomaya, "An industrial network intrusion detection algorithm based on multi-characteristic data clustering optimization model," *IEEE Transactions on Industrial Informatics*, vol. 10, 2019.
- [8] Y. Peng and B.-L. Lu, "Discriminative extreme learning machine with supervised sparsity preserving for image classification," *Neurocomputing*, vol. 261, pp. 242–252, 2017.
- [9] W. Enck, P. Gilbert, B. Chun et al., "An information-flow tracking system for realtime privacy monitoring on smartphones," *ACM Transactions on Computer Systems*, vol. 32, no. 2, 2014.
- [10] X. Su, X. Liu, J. Lin, S. He, Z. Fu, and W. Li, "De-cloaking malicious activities in smartphones using HTTP flow mining," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 6, pp. 3230–3253, 2017.
- [11] A. Gorla, I. Tavecchia, F. Gross, and A. Zeller, "Checking app behavior against app descriptions," in *Proceedings of the 36th International Conference on Software Engineering (ICSE)*, pp. 1025–2035, Hyderabad, India, May 2014.
- [12] F. Yu, A. Saswat, D. Isil, and A. Alex, "Apposcopy: semantics-based detection of android malware through static analysis," in *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering (FSE)*, pp. 576–587, Hong Kong, China, November 2014.
- [13] W. Enck, D. Octeau, P. McDaniel, and S. Chaudhuri, "A study of android application security," in *Proceedings of the 20th USENIX Conference on Security*, pp. 21–36, San Francisco, CA, USA, August 2011.
- [14] W. Yang, X. S. Xiao, B. Andow, S. H. Li, T. Xie, and W. Enck, "Appcontext: Differentiating malicious and benign mobile app behaviors using context," in *Proceedings of the 37th International Conference on Software Engineering (ICSE)*, pp. 303–313, Florence, Italy, May 2015.
- [15] H.-J. Zhu, Z.-H. You, Z.-X. Zhu, W.-L. Shi, X. Chen, and L. Cheng, "DroidDet: effective and robust detection of android malware using static analysis along with rotation forest model," *Neurocomputing*, vol. 272, pp. 638–646, 2018.
- [16] A. Mehtab, W. B. Shahid, T. Yaqoob, H. Abbas, H. Afzal, and M. N. Saqib, "AdDroid: rule-based machine learning framework for android malware analysis," *Mobile Network and Application*, vol. 24, no. 128, pp. 1–13, 2019.
- [17] L. Xie and S. Li, "Android malware detection model based on bagging-SVM," *Journal of Computer Applications*, vol. 38, no. 3, pp. 818–823, 2018.
- [18] L. K. Yan, "Droidscape: seamlessly reconstructing the os and dalvik semantic views for dynamic android malware analysis," in *Proceedings of the 21st USENIX Conference on Security Symposium*, p. 29, Berkeley, CA, USA, 2012.
- [19] A. Saracino, D. Sgandurra, G. Dini, and F. Martinelli, "MADAM: effective and efficient behavior-based android malware detection and prevention," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 1, pp. 83–97, 2018.
- [20] I. Burguera, U. Zurutuza, and S. Nadjm-Tehrani, "Crowdroid: behavior-based malware detection system for android," in *Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pp. 15–26, Chicago, IL, USA, 2011.
- [21] S. Hou, A. Saas, L. Chen, and Y. Ye, "Deep4MalDroid: a deep learning framework for android malware detection based on linux kernel system call graphs," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence Workshops*, pp. 104–111, WIW), Omaha, NE, USA, October 2017.
- [22] F. Ali, A. N. Badrul, and S. Rosli, "Evaluation of network traffic analysis using fuzzy C-means clustering algorithm in mobile malware detection," *Advanced Science Letters*, vol. 24, no. 2, pp. 929–932, 2018.
- [23] APKTOOL, <http://ibotpeaches.github.io/Apktool/>.
- [24] An disassembler for androids dex format, <http://code.google.com/p/smali/>.
- [25] A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner, "Android permissions demystified," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, pp. 627–638, Chicago, IL, USA, October 2011.
- [26] K. W. Y. Au, Y. Zhou, and Z. Huang, "Pscout: analyzing the android permission specification," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ACM, Raleigh, NC, USA, October 2012.



- [27] Android.basebridge, [http://www.symantec.com/securityresponse/writeup.jsp?docid=2011-060915-4938-99\\$&\\$tabid=2](http://www.symantec.com/securityresponse/writeup.jsp?docid=2011-060915-4938-99$&$tabid=2).
- [28] H. Kang, J. Jang, and A. Mohaisen, "Detecting and classifying android malware using static analysis along with creator information," *International Journal of Distributed Sensor Networks*, vol. 11, no. 6, pp. 1–9, 2015.
- [29] Y. Zhou and X. Jiang, "Dissecting android malware: characterization and evolution," in *Proceedings of the 2012 IEEE Symposium on Security and Privacy (S&P)*, pp. 95–109, San Francisco, CA, USA, May 2012.
- [30] H. L. Thanh, "Analysis of malware families on android mobiles: detection characteristics recognizable by ordinary phone users and how to fix it," *Journal of Information Security*, vol. 4, no. 4, pp. 213–224, 2013.

## Research Article

# An Assessment of the Relationship between Structural and Functional Imaging of Cerebrovascular Disease and Cognition-Related Fibers

Xiaoping Tang <sup>1</sup>, Xinlan Xiao <sup>1</sup>, Jianhua Yin <sup>1</sup>, Ting Yang <sup>1</sup>, and Bingliang Zeng <sup>2</sup>

<sup>1</sup>Department of Radiology, Second Affiliated Hospital of Nanchang University, Nanchang 330006, Jiangxi, China

<sup>2</sup>Department of Radiology, Jiangxi Provincial People's Hospital Affiliated to Nanchang University, Nanchang 330006, Jiangxi, China

Correspondence should be addressed to Bingliang Zeng; [zengbingliang@sina.com](mailto:zengbingliang@sina.com)

Received 19 September 2019; Revised 5 December 2019; Accepted 23 December 2019; Published 20 January 2020

Guest Editor: Kaijian Xia

Copyright © 2020 Xiaoping Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to assess the relationship between structural and functional imaging of cerebrovascular disease and cognition-related fibers, this paper chooses a total of 120 patients who underwent cerebral small vessel disease (CSVD) treatment at a designated hospital by this study from June 2013 to June 2018 and divides them into 3 groups according to the random number table method: vascular dementia (VaD) group, vascular cognitive impairment no dementia (VCIND) group, and noncognition impairment (NCI) group with 40 cases of patients in each group. Cognitive function measurement and imaging examination were performed for these 3 groups of patients, and the observation indicators of cognitive state examination (CSE), mental assessment scale (MAS), clock drawing test (CDT), adult intelligence scale (AIS), frontal assessment battery (FAB), verbal fluency test (VFT), trail making test (TMT), cognitive index (CI), white matter lesions (WML), third ventricle width (TVW), and frontal horn index (FHI) were tested, respectively. The results shows that the average scores of CSE, MAS, AIS, and VFT in the VaD and VCIND group are lower than those of the NCI group and the differences are statistically significant ( $P < 0.05$ ); the average scores of FAB, TMT, and CI in the VaD group are higher than those of the VCIND group and the differences are also statistically significant ( $P < 0.05$ ); the average scores of FHI and TVW in the VaD group are lower than those of the VCIND and NCI group with statistically significant differences ( $P < 0.05$ ); the average scores of WML, CDT, and AIS in the VaD group are higher than those of the VCIND and NCI group with statistically significant differences ( $P < 0.05$ ). Therefore, it is believed that the structural and functional imaging features of cerebrovascular disease are closely related to cognition-related fibers, and the incidence of white matter lesions is closely related to the degree of lesions and cognitive dysfunction of cerebral small vessel disease, in which a major risk factor for cognitive dysfunction in patients with small blood vessels is the severity of white matter lesions; brain imaging and neuropsychiatric function assessment can better understand the relationship between cerebrovascular disease and cognitive impairment. The results of this study provide a reference for the further research studies on the relationship between structural and functional imaging of cerebrovascular disease and cognition-related fibers.

## 1. Introduction

Cerebral small vessel disease (CSVD) refers to a group of lesions in intracranial arterioles, arterioles, anterior capillaries, and venules due to different causes and their clinical manifestations mainly include focal cerebral infarction, various lacunar syndromes, and cognitive dysfunction with imaging findings of vascular dementia (VaD), lacunar

infarction (LI), cerebral white matter lesion (WML), and cerebral microbleeds (CMB). The vascular dementia (VaD) is mainly caused by a series of brain vascular risk factors, which lead to dementia syndrome and clinical manifestations of cognitive dysfunction [1]. For subcortical ischemic VaD, it is a subtype of VaD, which is better in homogeneity, and is also a type of dementia in which cognitive function is impaired due to severe white matter lesions under the cortex

[2]. In neuroimaging, white matter lesions around the lateral ventricle are typical, but the cortex is rarely involved. Subcortical ischemic vascular disease (SIVD) is caused by small blood vessel lesions leading to lacunar infarction and white matter lesions, and the pathogenesis of VaD remains to be further elucidated [3]. Therefore, some modern research techniques and methods, such as brain imaging and neuropsychiatric functional assessment, can better understand the relationship between cerebrovascular disease and cognitive impairment [4].

Although structural magnetic resonance imaging (MRI) has been a very useful tool to help physicians to accurately detect gross anatomical changes associated with a range of clinical disorders and typical developmental processes, while functional magnetic resonance imaging enables researchers to map different cognitive structures of specific brain regions in conscious and task-seeking individuals. Functional magnetic resonance is a relatively new technology, and it is clear that this technology may drive the rapid advancement of psychology and neuroscience [5]. Clinically, the ADC value of patients with cerebrovascular disease will increase to varying degrees when MRI is used; the ADC value mainly reflects the diffusion of water molecules in the direction of the gradient magnetic field, which cannot fully reflect specificity between different tissues. Functional magnetic resonance not only has excellent spatial resolution, but it also provides researchers with a window that allows researchers to observe brain changes in subjects performing cognitive tasks. Although the voxel-based morphology (VBM) analysis method can be used not only to analyze the difference in the gray matter structure between Alzheimer's disease (AD) patients and normal individuals but also to suggest changes in white matter structure, but whether white matter atrophy exists and its degree of damage cannot pass [6]. This method is completely deterministic, suggesting that macroscopic white matter damage detected by voxel-based morphology (VBM) does not fully reflect microscopic white matter integrity changes [7]. Diffusion tensor imaging (DTI) is a more sensitive imaging method for detecting white matter microscopic tissue damage, and it has been confirmed that DTI parameters are more closely related to MRI image volume and cognition [8].

In order to assess the relationship between structural and functional imaging of cerebrovascular disease and cognition-related fibers, this paper chooses a total of 120 patients who underwent cerebral small vessel disease treatment at a designated hospital by this study from June 2013 to June 2018 and divides them into 3 groups according to the random number table method: vascular dementia group, vascular cognitive impairment no dementia group, and noncognition impairment group with 40 cases of patients in each group. Cognitive function measurement and imaging examination were performed for these 3 groups of patients, and the observation indicators of cognitive state examination, mental assessment scale, clock drawing test, adult intelligence scale, frontal assessment battery, verbal fluency test, trail making test, cognitive index, white matter lesions, third ventricle width, and frontal horn index were tested, respectively. The detailed chapters are organized as follows:

Section 2 presents research materials and methods; Section 3 performs results analysis; Section 4 assesses the relationship between structural and functional imaging of cerebrovascular disease and cognition-related fibers; Section 5 is discussion; and Section 6 is conclusion.

## 2. Materials and Methods

**2.1. General Materials.** A total of 120 patients who underwent SIVD treatment at a designated hospital by this study from June 2013 to June 2018 were selected as study subjects and were divided into 3 groups according to the random number table method: vascular dementia (VaD) group, vascular cognitive impairment no dementia (VCIND) group, and noncognition impairment (NCI) group with 40 cases of patients in each group. The VaD group contains 23 males and 17 females; their ages are in 55–83 years old with an average age of  $(66.27 \pm 5.67)$  years old; their daily living ability scale (ADL) scores are in 18–45 with an average score of  $(28.17 \pm 1.59)$ ; their cognitive state examination (CSE) scores are in 10–23 with an average score of  $(20.06 \pm 3.84)$ ; their mental assessment scale (MAS) scores are in 11–26 with an average score of  $(24.52 \pm 2.71)$ ; their educational experiences include 2 cases of primary school and below, 22 cases of junior high school education, and 16 cases of high school education and above (see Table 1). The VCIND group contains 24 males and 16 females; their age are 52–85 years old with an average age of  $(66.32 \pm 5.64)$  years old; their ADL scores are 16–44 with an average score of  $(28.04 \pm 1.62)$ ; their CSE scores are 11–23 with an average score of  $(19.79 \pm 3.71)$ ; their MAS scores are 10–25 with an average score of  $(24.57 \pm 2.63)$ ; their educational experiences include 5 cases of primary school and below, 17 cases of junior high school education, and 28 cases of high school education and above (see Table 1). The NCI group contains 20 males and 20 females; their ages are 53–86 years old with an average age of  $(66.48 \pm 5.67)$  years old; their ADL scores are 18–44 with an average score of  $(27.83 \pm 1.67)$ ; their CSE scores are 12–25 with an average score of  $(20.14 \pm 3.49)$ ; their MAS scores are 11–25 with an average score of  $(24.33 \pm 2.57)$ ; their educational experiences include 7 cases of primary school and below, 15 cases of junior high school education, and 28 cases of high school and above (see Table 1). There are no significant differences in the general data of the three groups ( $P < 0.05$ ), and the data are comparable. This study was reviewed and approved by the hospital medical ethics committee.

**2.2. Inclusion and Exclusion Criteria.** Inclusion criteria: (1) patients who meet SIVD MRI diagnostic criteria and SIVD diagnostic criteria developed in the expert consensus on vascular cognitive impairment; (2) patients with acute ischemic cerebrovascular disease more than 1 time and more than 3 months; (3) patients whose brain MRI diagnosis shows white matter damage and lacunar infarction prominent; (4) patients with cognitive impairment; (5) patients who were passed by the hospital ethics committee and themselves or their family members were informed and voluntarily signed informed consent.

TABLE 1: General information of SIVD patients in VaD, VCIND, and NCI group ( $\bar{x} \pm s$ ).

Group	Average age (years old)	Gender (male/female)	ADL	CSE	MAS	Education level		
						PSB	JHS	HSA
VaD	66.27 $\pm$ 5.67	23/17	28.17 $\pm$ 1.59	20.06 $\pm$ 3.84	24.52 $\pm$ 3.84	2	22	16
VCIND	66.32 $\pm$ 5.64	24/16	28.04 $\pm$ 1.62	19.79 $\pm$ 3.71	24.57 $\pm$ 2.63	5	17	28
NCI	66.48 $\pm$ 5.67	20/20	24.33 $\pm$ 2.57	84.97 $\pm$ 5.31	24.33 $\pm$ 2.57	7	15	28
<i>P</i>	>0.05	>0.05	>0.05	>0.05	>0.05	>0.05		

PSB: primary school and below; JHS: junior high school; HAS: high school and above.

Exclusion criteria: (1) patients with severe visual and hearing impairment, severe physical weakness, or aphasia affecting the examination or failing to cooperate; (2) patients with mental illness or disturbance of consciousness; (3) patients who had already presented cognitive dysfunction and their assessment scores exceeding 2 points or a short caregiver questionnaire totalling more than 56 points; (4) patients who have pseudocognitive impairment or dementia caused by depression, anxiety, or AD; (5) patients whose cognitive dysfunction caused by other diseases, including infection, cancer, poisoning, metabolic diseases, and congenital mental retardation; (6) patients with severe heart, liver, lung, kidney and other organ failure, and therefore the examination cannot be completed; (7) patients who have MRI scan contraindications or are unable to scan; (8) patients whose imaging examination indicates the presence of other intracranial diseases, including cerebral hemorrhage and subarachnoid hemorrhage; (9) patients or their family are not willing to participate in the experiment.

**2.3. Cognitive Function Determination Methods.** All patients were enrolled with basic information such as name, age, gender, and education level. The cognitive function rating scale is mainly divided into four parts: (1) simple mental state examination CSE; the table consists of 40 questions, a total of 40 items; each answer is correct 1 point, the answer is wrong or the answer does not know the score 0; the total score of the scale is 0–30 points; (2) the meaning of the word is composed of 40 questions, each score is 1 or 0 points, out of 40 points (3) digital recitation is a subtest consisting of two parts: the back and back of the number, each score is 2, 1, and 0, and the total score is the sum of the two, with a score of 40; (4) line connection test is divided into two parts: A and B; A is 40 numbers in sequence and B puts 40 numbers in square and round and requires two patterns to be alternately and sequentially connected and the time spent in two parts are calculated. Pearson correlation analysis software was used to analyze the relationship between imaging findings and cognitive dysfunction of cerebral small vessel disease.

**2.4. Imaging Examination Methods.** All patients were examined by magnetic resonance imaging to obtain diffusion coefficient (ADC) parameters; after admission, the MRI was performed with a Philips 3.0T magnetic resonance instrument. The relevant parameters were set according to the actual situation of each patient; T1-weighted image (T1W1): time of repetition (TR) is 1800 ms, echo time (TE) is 30 ms; T2-weighted image (T2W1): TR is 2000 ms, TE is 100 ms;

fluid attenuated inversion recovery (FLAIR): TR is 10000 ms, TE is 125 ms, TE is 100 ms; matrix 120  $\times$  124, layer thickness is 4 mm, and pitch is 0.3 mm with continuous 80 layer scans. Image processing: the obtained route is processed by Philips proprietary software, and then the image is subjected to the dispersion alignment software to obtain ADC parameters.

**2.5. Observation Indicators.** Observation indicators include cognitive state examination (CSE), mental assessment scale (MAS), clock drawing test (CDT), adult intelligence scale (AIS), frontal assessment battery (FAB), verbal fluency test (VFT), color word test (CWT), trail making test (TMT), cognitive index (CI), white matter lesions (WML), third ventricle width (TVW), and frontal horn index (FHI).

**2.6. Indicator Examination Methods.** The main indicator examination workflow is shown in Figure 1 and is described as follows:

TMT: the project is a subtest of the neuropsychological test; part A connects 40 Arabic numerals in sequence, and part B alternates numbers and letters; the test only performs part A, recording the completion time.

FAB: it is a trial of the frontal dysfunction including similarity, lexical fluency, motion sequence testing, inconsistency instructions, inhibitory control and gripping behaviour and six subtests; the score for each item is between 0 and 4 points, with a maximum total score of 20 points; a higher score indicates better performance.

CDT: on a blank sheet of paper, the subject draws a dial and is asked to write all the numbers; the tester indicates the pointer at 11:10; the tester evaluates the subject, including whether the contour of the watch is complete and whether the pointer is correct and the total score is 3 points, and the correction and rubbing are not deducted.

CWT: a set of three test cards and the test contents of the three cards are arranged in a 8  $\times$  8 matrix, and the time (s) used to read the three cards is recorded separately; the A card consists of red dots, yellow dots, blue dots, and green dots, and reads the dot color; the B card is composed of four characters and the characters are red; the four colours of yellow, blue and green are printed in monochrome, and the color of characters is required to be read; the C card is four characters of red, yellow, blue, and green and the printing colours are

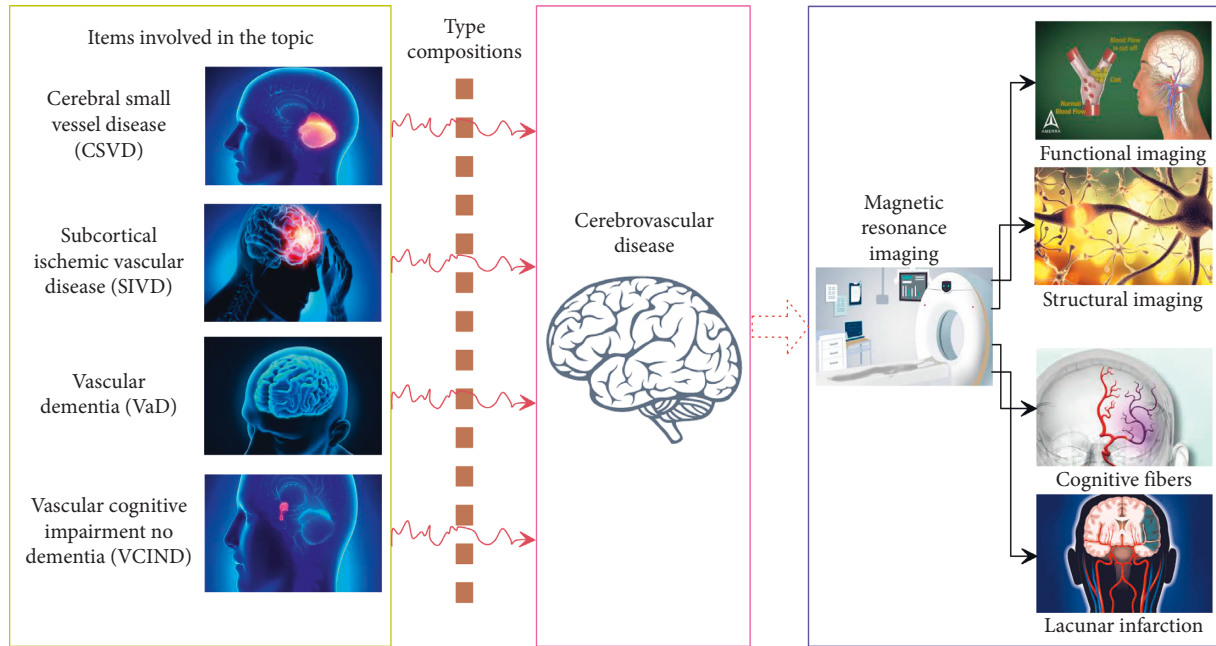


FIGURE 1: Framework of detecting the relationship between structural and functional imaging of cerebrovascular disease and cognition-related fibers.

these four colours, and the printing colours are different from the meaning of the characters.

**2.7. Statistical Methods.** Data analysis was performed using SPSS 19.0 statistical software. The data of the measurement data were expressed as mean  $\pm$  standard deviation ( $\bar{x} \pm s$ ) and the  $t$  test was used to compare the two groups. The statistical analysis was performed after the normal distribution was not consistent with the normal distribution and one-way analysis of variance was used for comparison between groups; the LSD- $t$  test was used for comparison between groups.  $P < 0.05$  was considered statistically significant.

### 3. Result Analysis

**3.1. Comparison of Overall Cognitive Function in Three Groups of Patients.** At present, there is a view that the number of lesions in the lacunar infarct is more significant than the number and volume of lesions. This study also found a correlation between cognitive function and lacunar infarction and there was a statistically significant difference in the number of subcortical white matter lacunar infarction between the three groups which may be due to lacunar infarction in the white matter (see Table 2). The lesion cuts off the fibrous relationship between certain structures under the cortex and the prefrontal cortex, the frontal cortex, and the cingulate gyrus, thereby inhibiting some of the functions associated with cognition in the frontal cortex and further leading to dementia. Patients with congestive heart failure are at high risk for cognitive dysfunction, suggesting that a decrease in cardiac output may affect the blood and nutritional supply of the brain and patients with low left ventricular dysfunction have a marked decline in cognitive

function. This study did not find that cognitive impairment was associated with the number of lacunar infarct lesions in the putamen, caudate nucleus, and globus pallidus; while the number of lacunar infarct lesions in the thalamus between the three groups was different, but no statistically significant speculation may be related to the small number of cases in this study. In this study, compared with the noncognitive disorder group, the WML scores of the mild cognitive impairment group and the vascular dementia group generally showed an increasing trend, suggesting that there is a correlation between the degree of cognitive impairment and the degree of total WML lesions.

There are many terms describing white matter high signal, such as leukoaraiosis, white matter lesions, white matter high signal, and leukoencephalopathy. The MRI of leukoaraiosis is defined as high signal in sequence, often difficult to find in sequence or mildly low signal, mainly distributed around the ventricles, bilateral white hemisphere deep white matter, basal ganglia, pons, and occasionally in the cerebellum white matter areas such as brainstem. The leukoaraiosis is closely related to the occurrence of dementia and has been found that leukoaraiosis can accelerate the occurrence of dementia through ischemic injury of the cortex and directly aggravate the pathological changes of AD. It is hypothesized that the angiogenic white matter high signal excludes white matter lesions caused by other non-angiogenic diseases, such as multiple sclerosis and white matter malnutrition [5]. The imaging feature is a high signal with a point-like, patchy, or convergent weighting and the weighted signal is equal or low, and the signal depends on the image sequence parameters and the degree of lesion. Some scholars have pointed out that hypothetical angiogenic white matter high signal has a high heritability, making it an intermediate marker for clinical research and screening for



TABLE 2: Comparison of TVW, AIS, VFT, and TMT in the VaD, VCIND, and NCI patients ( $\bar{x} \pm s$ ).

Group	Treatment time	TVW	AIS	VFT	TMT
VaD ( $n = 40$ )	Before treatment	$0.96 \pm 0.23$	$1.52 \pm 0.49$	$10.27 \pm 3.17$	$123.63 \pm 57.73$
	After 6 months of treatment	$0.93 \pm 0.21$	$1.36 \pm 0.38$	$12.46 \pm 2.19$	$122.15 \pm 45.38$
	After 12 months of treatment	$0.84 \pm 0.19$	$1.63 \pm 0.11$	$15.13 \pm 2.14$	$120.77 \pm 54.69$
VCIND ( $n = 40$ )	Before treatment	$0.91 \pm 0.22$	$1.68 \pm 0.39$	$11.14 \pm 3.12$	$123.55 \pm 56.87$
	After 6 months of treatment	$0.88 \pm 0.21$	$1.57 \pm 0.27$	$13.03 \pm 2.06$	$121.89 \pm 45.44$
	After 12 months of treatment	$0.82 \pm 0.18$	$1.21 \pm 0.63$	$16.44 \pm 2.16$	$120.32 \pm 54.76$
NCI ( $n = 40$ )	Before treatment	$0.80 \pm 0.19$	$1.79 \pm 0.41$	$15.03 \pm 3.06$	$123.46 \pm 46.35$
	After 6 months of treatment	$0.81 \pm 0.17$	$1.80 \pm 0.24$	$15.75 \pm 2.04$	$115.57 \pm 55.01$
	After 12 months of treatment	$0.79 \pm 0.16$	$1.87 \pm 0.36$	$16.72 \pm 2.13$	$111.47 \pm 45.32$
F/P VaD group value		0.163/0.028	0.325/0.003	0.754/0.003	0.855/0.016
F/P VCIND group value		0.174/0.005	0.542/0.018	0.584/0.027	0.585/0.027
F/P NCI group value		0.164/0.002	0.234/0.003	0.234/0.015	0.584/0.005
t/P 6 months treatment group value		0.471/0.013	0.422/0.014	0.742/0.003	0.234/0.023
t/P 12 months treatment group value		0.315/0.001	0.512/0.032	0.284/0.002	0.679/0.015

new risk factors for stroke or dementia. Clinical case resources and rapid development of imaging techniques study its pathogenesis and combine with clinical risk factors.

Brain atrophy can manifest as symmetric or asymmetrical, extensive or localized atrophy, and may be tissue selective. The imaging manifestation is that the brain volume becomes smaller, and it has nothing to do with the local trauma and infarction of the brain visible to the naked eye. The pathological changes of brain atrophy are heterogeneous, which represents neuronal loss, cortical thinning, reduced white matter sparseness, arteriosclerosis, and secondary neuronal degeneration [9]. The brain atrophy may be associated with cognitive impairment, and studies have found that the volume of nucleus accumbens, amygdala, caudate nucleus, thalamus, and brainstem is closely related to cognition. Several patients with transient ischemic attack or suspected stroke participated in the survey and several patients had atrophy of medial temporal lobe, and the study found that medial temporal lobe atrophy was associated with memory, naming, sensory ability, executive ability and speed, and attention. Cerebral microinfarction refers to small infarction caused by ischemia, and high-resolution magnetic resonance can detect large microinfarction. Based on population and autopsy studies, brain microinfarction is associated with cognitive decline and the mechanism of cognitive influence is not clear, which may be related to the destruction of the brain contact fiber network and brain microinfarction is an important bridge linking cerebrovascular disease and dementia.

**3.2. Comparison of Imaging Findings of Three Groups of Patients.** Previous studies have indicated that patients with a more uniform reduction in cortical density include the primary motor cortex, frontal lobe, temporal lobe, marginal lobe, and lateral central lobules. The results of this study suggest that compared with the healthy control group, the gray density of the bilateral central anterior gyrus is significantly lower than that of the normal control group and

the gray matter density of other brain regions is not significantly abnormal (see Table 3). The bilateral central anterior gyrus is the motor area of the cerebral cortex and this part of the cortical dysfunction can lead to dyskinesia in patients; this result supports the core pathological changes of amyotrophic lateral sclerosis. Lacunar infarction, white matter lesions, and basal ganglia lesions are closely related to stroke and transient ischemic attacks; they are closely related to the presence of gait disorders in patients. The study sample size is small, the patients included in the early stage of the disease, the magnetic resonance scanning parameters used in previous studies, and the limitations of the magnetic resonance processing technology itself may cause the analysis results of this study to be inconsistent with previous reports. The enrolment of patients in the early stages of the disease is the main reason for the large difference between the results of this study and the past. However, cross-sectional studies grouped by disease severity or subsequent longitudinal follow-up studies of this subject are needed to confirm this reasoning. The thickness of the cortex can reflect the neuropathological characteristics of the disease more sensitively than the gray matter density or volume which is an imaging index that studies the morphological changes of the cerebral cortex with high reliability.

Conventional MRI imaging showed that mild to moderate WML had no significant effect on cognitive function. The cognitive function of mild to moderate WML depends mainly on the total number of basal and thalamic regions LI; the basal ganglia and thalamus are key sites in the development of LI, which can better predict cognitive decline. LI distributed in the white matter of the brain extensively interferes with the structure of the white matter, and chronic inflammatory processes, loss of myelin, and axonal degeneration may cause more severe cognitive decline due to such interference. The pathogenesis of WML progresses from the subcortical white matter to the white matter adjacent to the ventricles and cystic LI may be involved in extensive WML under the cortex, and the number of LI in subcortical WML cannot be fully recognized. Analysis of the MRI results of LI

TABLE 3: Comparison of CI, FAB, WML, and FHI in the VaD, VCIND, and NCI patients ( $\bar{x} \pm s$ ).

Group	Treatment time	CI	FAB	WML	FHI
VaD ( $n = 40$ )	Before treatment	$5.92 \pm 0.89$	$11.41 \pm 2.34$	$14.15 \pm 3.23$	$3.43 \pm 0.63$
	After 6 months of treatment	$5.58 \pm 0.81$	$11.57 \pm 2.53$	$9.54 \pm 2.35$	$3.75 \pm 0.58$
	After 12 months of treatment	$5.14 \pm 0.69$	$11.64 \pm 2.77$	$3.63 \pm 2.47$	$3.95 \pm 0.29$
VCIND ( $n = 40$ )	Before treatment	$5.28 \pm 0.84$	$11.45 \pm 2.44$	$11.34 \pm 3.16$	$3.55 \pm 0.87$
	After 6 months of treatment	$5.16 \pm 0.81$	$11.53 \pm 2.34$	$7.68 \pm 2.33$	$3.84 \pm 0.64$
	After 12 months of treatment	$5.09 \pm 0.78$	$11.57 \pm 2.70$	$3.55 \pm 2.45$	$4.01 \pm 0.46$
NCI ( $n = 40$ )	Before treatment	$5.05 \pm 0.88$	$11.44 \pm 2.44$	$3.10 \pm 3.55$	$4.06 \pm 0.35$
	After 6 months of treatment	$5.03 \pm 0.67$	$11.56 \pm 2.83$	$2.34 \pm 2.27$	$4.07 \pm 0.51$
	After 12 months of treatment	$5.02 \pm 0.66$	$11.63 \pm 2.54$	$2.67 \pm 2.68$	$4.09 \pm 0.32$
F/P VaD group value		0.758/0.001	0.660/0.014	0.116/0.016	0.855/0.001
F/P VCIND group value		0.431/0.011	0.357/0.006	0.775/0.007	0.465/0.015
F/P NCI group value		0.297/0.024	0.486/0.001	0.248/0.001	0.784/0.020
t/P 6 months treatment group value		0.644/0.001	0.355/0.022	0.351/0.014	0.694/0.010
t/P 12 months treatment group value		0.586/0.014	0.622/0.001	0.896/0.001	0.579/0.001

and WML in the basal and thalamic regions showed that the severity of WML and the number of basal and thalamic regions LI were related to the extent of cognitive impairment. This study found that severe WML or basal ganglia and multiple lymphoid in the thalamus area can cause cognitive decline. Probably because cognitive function depends on neural network function, both WML and LI have a phenomenon of neural network interruption, the so-called separation syndrome, which has a decline in cognitive function [10].

Perivascular spaces (PVS) refer to the filling of fluid around the cerebral perforating arterioles and venules, and its diameter is usually less than 3 mm. The imaging findings are T1W1 low signal, T2W1 high signal, and FLAIR low signal. It plays an important role in the formation of a discharge channel for metabolic waste and liquid in the brain. Whether or not PVS have clinical significance is controversial and some studies suggest that these gaps do not represent damage, and some studies have shown that PVS is associated with decreased cognitive function. Previous studies found that there was no independent association between PVS and cognitive impairment in people with ischemic stroke and transient ischemic attack large PVS is associated with subcortical infarction, increased volume, and microbleeds. It indicates that there may be a correlation between PVS and AD, and the study found that PVS in the basal ganglia were associated with high levels of cerebrospinal fluid; PVS in the semioval centre and PVS in the basal ganglia were different, but were associated with cerebral amyloid angiopathy and hypertensive vasculopathy. The former is more associated with AD than the PVS in the white matter region and the basal ganglia PVS.

#### 4. Relationship between Cerebrovascular Disease Imaging and Cognition-Related Fibers

4.1. Relationship between Structural Imaging of Cerebrovascular Disease and Cognition-Related Fibers. The results

suggest that the medial temporal lobe memory system, especially the integrity of hippocampus and cortical-associated fibers, is closely related to cognitive function. The medial temporal lobe memory system, also known as the hippocampus memory system or the marginal memory system, is located in the ventral aspect of the temporal lobe and is composed of the hippocampus formation and adjacent cortex. Nearly two-third of the afferent fibers of the hippocampus are from the parasitic cortex and the hippocampus cortex (see Figure 2(a)). The parasitic cortex and the hippocampus cortex have extensive bidirectional fiber connections with the frontal, temporal, and parietal cortex. The information in the hippocampus is the main pathway, and the entorhinal cortex is a temporary integration zone in the cortical liaison zone. In the course of AD, severe pathological changes block the transmission of information in different brain regions, the damage to the circuit is different in different cortex, and the two major cortical systems that play an important role in cognitive memory are particularly selective [11]. One is the fiber connection between the hippocampus and the adjacent cortical structure in the temporal lobe, which is reflected in the atrophy of the parasitic cortex and the expansion of the hook spacing representing the hippocampus formation and atrophy of the hippocampus cortex (see Figure 2(b)). The second is the basal forebrain energy system. These circuits show severe degenerative changes in the early stages of AD, which may be the cause of significant memory loss in the early stages of AD.

The integrity of cerebral blood circulation is closely related to the occurrence and prognosis of cerebral infarction, especially in the event of internal carotid artery stenosis or occlusion. In patients with severe stenosis or occlusion of the unilateral internal carotid artery, anterior communicating artery opening is a predictor of hemodynamic preservation, whereas no arterial flow is a sign of hemodynamic damage after collateral circulation or only blood flow. Secondary collateral circulation occurs when the primary collateral circulation is not developed or is still unable to maintain normal perfusion, and is also important, including pia joint anastomosis and ophthalmic artery. As

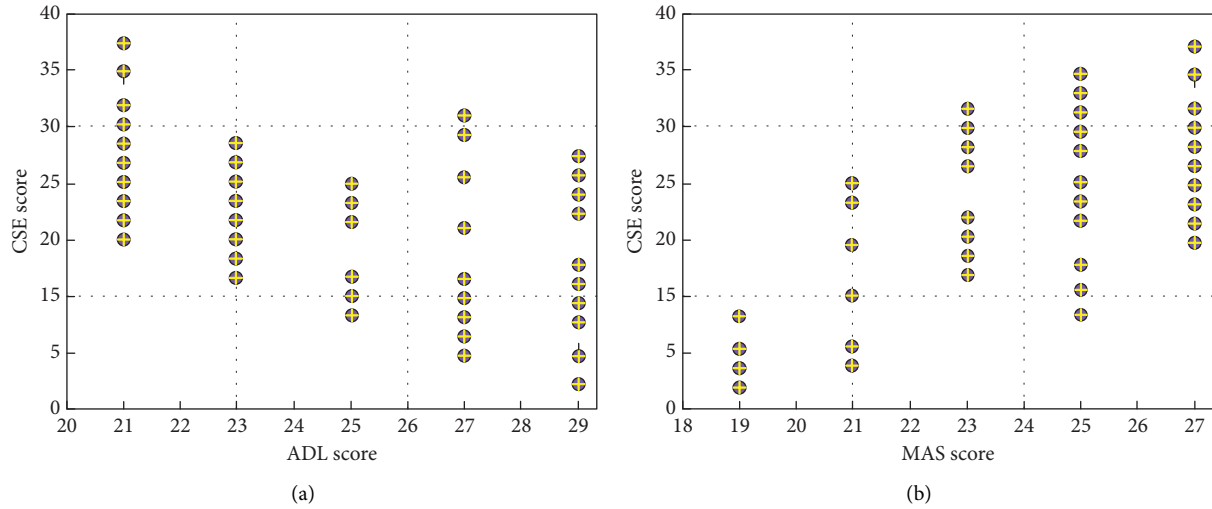


FIGURE 2: Analysis results of the relationship between structural imaging of cerebrovascular disease and cognition-related fibers. (a) The relationship between CSE and ADL score in the VaD group of patients. (b) The relationship between CSE and MAS score in the VaD group of patients.

the pia matter undulates on the surface of the sulci, it belongs to the secondary collateral circulation, and the diameter and number of anastomotic vessels are extremely large variability in which blood flow can be bidirectional with hemodynamic changes and metabolic needs in the anastomotic region. Clinically, the infarction, stenosis or occlusion caused by secondary branch lesions occurs in the anterior and posterior cerebral arteries, and the circulation is difficult to compensate. The opening of the pial-macular anastomosis will play a major role, and the opening of the pia matter anastomosis is a reliable predictor of the good outcome of ischemic cerebrovascular disease.

As a common type of cerebral infarction, LI is mainly caused by hypertension, diabetes, and other factors leading to changes in the structure of small arteries in the brain, and the cerebral blood flow automatic regulation mechanism is impaired, eventually leading to cerebral infarction. It is characterized by small infarct size and multiple occurrences, and obvious cognitive dysfunction can occur without obvious signs of the nervous system, for example, the lesion in the posterior horn of the sac is associated with impaired executive function. The lesions in the frontal lobe are associated with memory impairment and executive function, coding, and management information dysfunction; occipital cavity and visual occlusion lesions are related; temporal lobe lesions are associated with memory, language, hearing, and affective disorders; lesions in the anterior thalamus are associated with amnesia. Studies have shown that the number, volume, and location of LI are independent predictors of treatment speed and performance impairment in symptomatic stroke. This study found that the CSE score and MAS score in the simple LI group were lower than those in the control group, indicating that the occurrence of LI does have a certain effect on cognitive dysfunction in the elderly and the plasma cognitive fiber level in the vascular dementia group was significantly higher; the increase in cognitive fiber was significantly related to the patient's oral memory disorder. The possible pathogenesis is LI destruction

and information processing, speed, memory, and executive function related to the frontal-cortical loop [12].

**4.2. Relationship between Functional Imaging of Cerebrovascular Disease and Cognition-Related Fibers.** Patients with cerebral small vessel disease are often accompanied by imaging changes and strengthening MRI can predict the degree of cerebral small vessel injury and guide clinical treatment. MR diffusion imaging is a commonly used imaging method in patients with cerebral small vessel disease; this method uses the flow of water molecules in the body of MRI to map the internal structure of the brain, which not only clearly shows the small changes in the brain, but also more accurate. It describes the changes in the blood flow of intracranial cerebrospinal fluid and small blood vessels while conventional magnetic resonance cannot reflect whether the diseased axon is damaged or not, and the MR can relatively determine whether the white matter fiber bundle is damaged or not and can assess the severity of the disease. Studies have shown that patients with cerebral small vessel disease can cause neuronal axon loss, resulting in the proliferation of glial cells and clinically, and have different degrees of ADC values when using MRI. The reason for this phenomenon is multifaceted because the ADC value mainly reflects the dispersion of water molecules in the direction of the gradient magnetic field; it cannot completely react to the specificity between different tissues (see Figure 3).

In the central nervous system, the diffusion of water molecules is affected by cell structures such as cell membranes, axonal membranes, and cytoskeleton and is also affected by natural barriers such as myelin, white matter fiber bundles, and protein macromolecules. There is a significant difference in the dispersion of the various parts of the central nervous system, and this difference is also the basis of diffusion tensor imaging. It is precisely because the dispersion of water molecules is a spontaneous physical

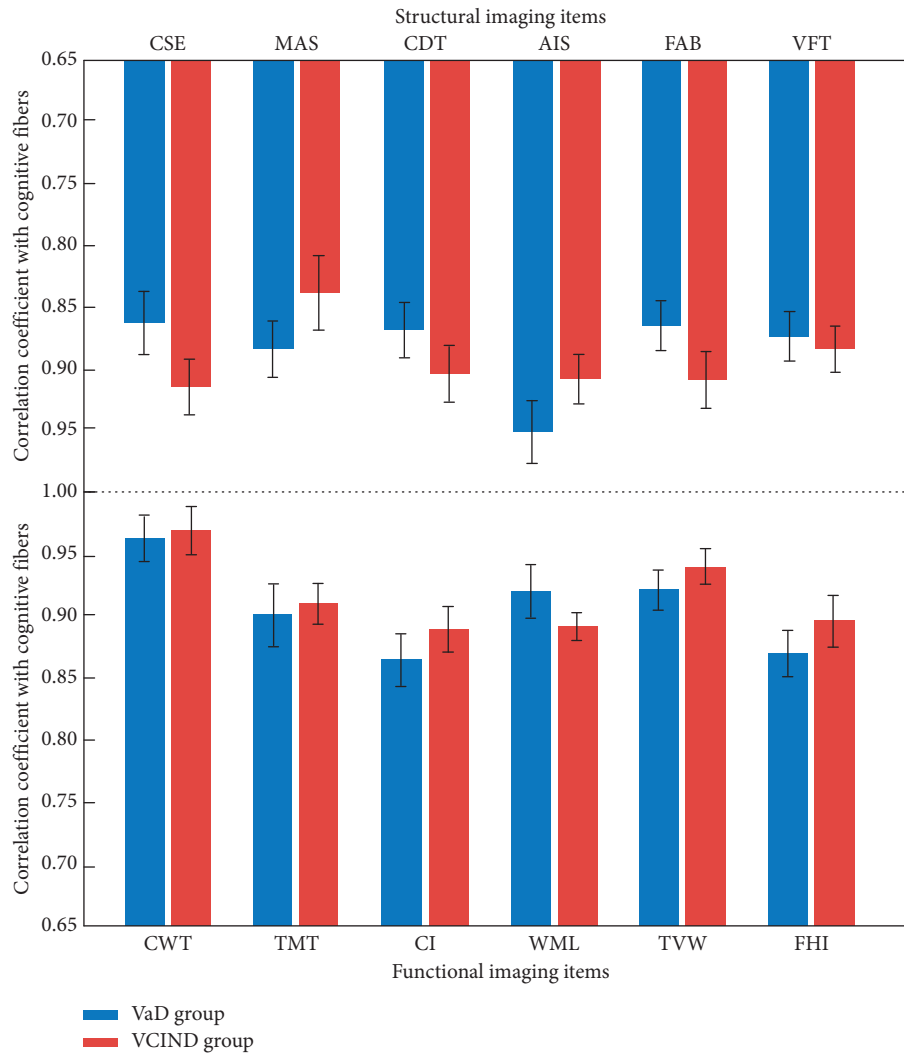


FIGURE 3: Result analysis of the relationship between functional imaging of cerebrovascular disease and cognition-related fibers (for the abbreviations, refer to the context).

phenomenon, so the scanning of the diffusion tensor does not require a specific cognitive task, and the data obtained by the subjects with or without stimulation is relatively stable. Diffusion tensor is currently the only noninvasive method for effectively observing and tracking white matter fiber bundles, which can show the anatomical and pathological processes of white matter fibers and indirectly evaluate the integrity of white matter fibers [13]. Compared with the traditional sequence of magnetic resonance imaging, the diffusion tensor has many advantages, such as scanning time period, more information obtained, better image quality, sensitive to changes in microstructure and function, and no need to use radioactivity, tracer, and high security.

Cognitive fibers and vascular lesions caused by atherosclerosis are not difficult to identify, and atherosclerosis is mainly seen in the elderly. The lesions are mainly located in the proximal end of the artery or in the bifurcation of the arteries, and there are many risk factors for cerebrovascular disease such as diabetes, hypertension, and hyperlipidemia found in young women; the lesions involve the middle and distal end of the artery, and there are many risk factors for

cerebrovascular disease. The clinical symptoms of the patient are related to the degree of stenosis of the affected artery and also to the location of cognitive fiber lesions, involving the carotid or vertebral artery cognitive fibers, can be expressed as black, hemiplegia, and cranial neuropathy without any clinical symptoms. The affected blood vessels expand in a saclike manner, and the blood flow is slow or eddies in the local flow. Platelets, red blood cells, etc. form aggregates to form thrombus, and emboli fall off to cause cerebral embolism; severe stenosis or occlusion of the affected blood vessels can cause the blood flow in the distal cerebral tissue insufficient perfusion, causing hypoperfusion cerebral infarction. The lesions can involve aneurysms caused by intracranial blood vessels, subarachnoid hemorrhage caused by aneurysm rupture, and it causing stroke episodes combined with intracranial vascular dissection [14].

## 5. Discussion

*5.1. Correlation between the Scores of Neuropsychological Scales and the Degree of White Matter Lesions.* The influence

TABLE 4: Correlation analysis between neuropsychological scale score and degree of white matter lesions.

Indicators	VaD	VCIND	NCI	<i>t</i>	<i>P</i>
CSE	0.43	0.24	-0.35	17.78	0.26
MAS	-0.35	0.37	-0.59	10.64	0.13
CDT	0.71	0.68	0.66	46.66	0.27
DST	-0.29	0.55	0.19	23.17	0.09
FAB	0.32	0.46	0.22	19.35	0.01
VFT	-0.67	-0.74	-0.45	12.55	0.05
CWT	-0.54	-0.62	-0.84	10.28	0.25
RT	-0.39	-0.56	-0.75	11.45	0.06
TE	-0.21	-0.63	-0.53	14.46	0.13
T1W1	0.38	-0.30	0.74	16.85	0.04
T2W2	0.46	-0.28	0.11	13.57	0.06
DWI	0.57	-0.76	-0.54	36.68	0.03
FVO	-0.11	0.36	-0.64	44.58	0.12
TMT	-0.57	0.56	-0.78	15.43	0.01
CI	-0.33	0.76	0.63	14.66	0.31
WML	0.79	0.44	0.37	17.28	0.12
TVW	0.43	-0.12	0.45	15.69	0.09
FHI	0.29	-0.49	0.52	18.73	0.02

of the location of lacunar infarction on cognitive function has been widely concerned; infarction in some key areas is more likely to lead to severe cognitive dysfunction which is recognized by most researchers. Case reports of a single cavity lesion in the strategic site have increased significantly, and cognitive impairment has occurred in the acute phase of stroke. Table 4 shows the correlation analysis between the neuropsychological scale score and degree of white matter lesions. Brain activity as a whole, imaging features are high signals that are spotted, patchy, or fused on a weighted basis, and different behaviours are associated with different brain functional parts. For example, when the executive function is compromised, it is often suggested that the site of destruction is located in the subcortical loop associated with the frontal lobe [15]. The loop inside the dorsal prefrontal lobes is most important for performing functions and the medial cortex of the prefrontal cortex serves as the starting point of this loop. The nerve fibers that are emitted are first connected to the dorsal medial side of the caudate nucleus and then projected to the old striatum, reaching the nucleus of the thalamus, and finally returning, the cerebral cortex to the dorsal medial side of the prefrontal cortex. However, in this loop, small penetrating arteries supply blood to the white matter and gray matter structure, so ischemic diseases are easily produced in this loop.

Related studies have found that cardiovascular events, carotid plaque formation, and peripheral vascular arteriosclerosis are independent risk factors for cognitive impairment. The study found that ischemic heart disease is an important factor in the aggravation and death of AD, and long-term chronic arrhythmia and cardiac insufficiency can cause dementia caused by ischemia and hypoxia. Some studies have shown that patients with congestive heart failure have a high risk of cognitive dysfunction, suggesting that decreased cardiac output may affect the blood and nutrient supply of the brain; another study suggests that cognitive function is significantly reduced in patients with

low left ventricular function (see Figure 4). There is no clear conclusion about the direct relationship between coronary heart disease and cognitive function and the effect of stroke on cognitive function is obvious. The decline in cognitive function caused by brain tissue loss due to the formation of intracerebral arterial infarction is called multi-infarct dementia or vascular dementia. Previous studies on vascular dementia have focused on cerebrovascular disease, and studies on cognitive impairment caused by other vascular lesions are relatively lacking. Because the detection of vascular lesions in the brain is relatively difficult and changes in coronary heart disease or peripheral arterial occlusive disease are easily identified, it can be used as a risk factor for cerebrovascular disease.

Secondary collateral circulation occurs when the primary collateral circulation is poorly compensated in patients with severe internal carotid artery stenosis or occlusion, and secondary collateral circulation and open pial obstruction is an important collateral compensatory pathway. The presence of secondary collateral circulation did not reduce the incidence of infarct events, but cognitive impairment of the internal carotid artery stenosis or occlusion in the presence of infarction was more pronounced than in patients without infarction. The study analyzed the risk factors of severe internal stenosis or occlusion of the internal carotid artery and used neuroimaging to observe the damage of brain tissue structure, collateral circulation, and cerebral blood flow changes and evaluate cognitive function to explore the severe internal stenosis or occlusion of the internal carotid artery. The mechanism of circulatory compensation, cerebral blood perfusion, and cognitive dysfunction provides a basis for establishing an early clinical neuroimaging diagnostic evaluation system, suggesting that clinicians are not only in the early stage of patients with severe internal stenosis or occlusion of the internal carotid artery. The evaluation of the brain tissue structure should pay more attention to collateral circulation compensation, cerebral blood perfusion, and cognitive function changes, so as to formulate standardized individual treatment plans, and should conduct early screening and reasonable vascular cognitive dysfunction and intervention to improve the quality of life of patients [16].

*5.2. Association Mechanism of Cerebrovascular Disease Imaging and Cognition-Related Fibers.* Cerebral small vessel disease is a common type of disease in vascular cognitive impairment, and clinical studies have consistently shown that frontal white matter lesions and lacunar infarction are leading to further aggravation of patients, progressive cognitive impairment, and the gait is an important factor in abnormality, and the occipital leukoencephalopathy and lacunar infarction are very likely to cause symptoms such as convulsions and incontinence in patients with cerebral small vessel disease (see Figure 5). Related studies have shown that lacunar infarction, white matter lesions, and basal ganglia lesions are closely related to stroke and transient ischemic attack, and they are closely related to patients having gait disorders [17]. Cerebral small vessel disease is a disease that



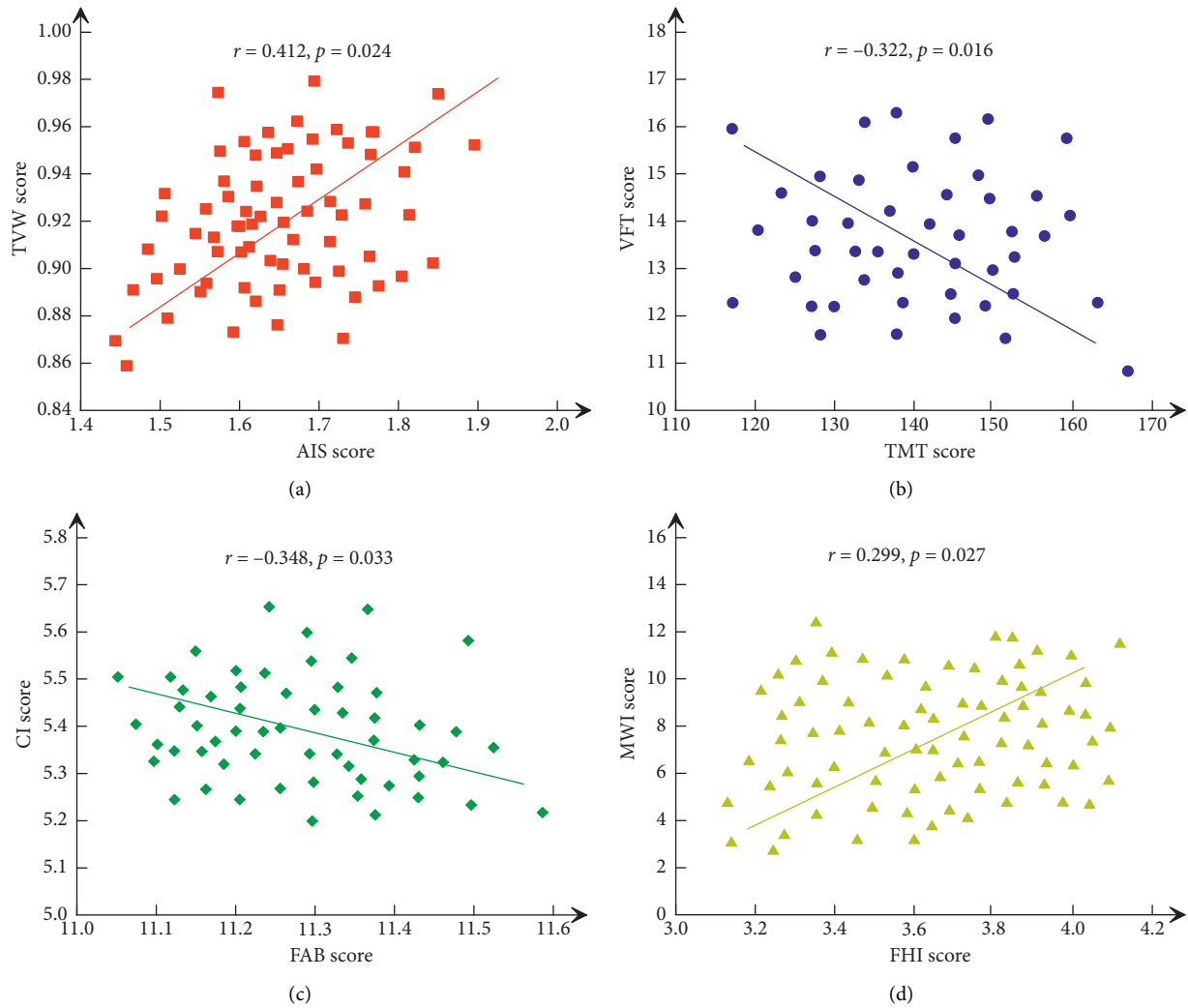


FIGURE 4: Correlation between the scores of neuropsychological scales and the degree of white matter lesions. (a) The relationship between TVM and AIS score. (b) The relationship between VFT and TMT score. (c) The relationship between CI and FAB score. (d) The relationship between MWL and FHI score. “ $r$ ” and “ $p$ ” were correlation coefficient and degree of fit, respectively.

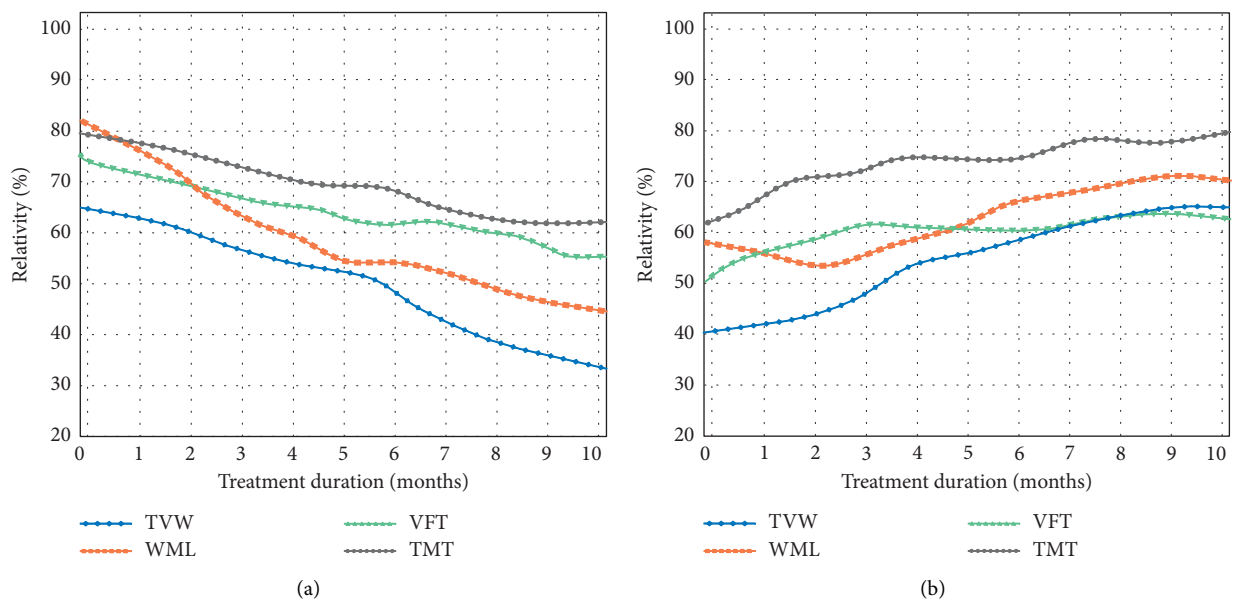


FIGURE 5: Evolution of key indicator relativity with treatment duration. (a) VaD group. (b) VCIND group.

is seriously harmful to the brain function and systemic health of patients and is the main cause of stroke and cognitive decline in patients. White matter and lacunar infarction are the most important imaging manifestations of the disease, which is beneficial to the clinical combination of characteristics to understand the disease and the brain damage of patients, and then select the corresponding drugs for patients and improve the functional damage.

A meta-analysis showed a positive correlation between cognitive fiber levels and the incidence of vascular dementia; high cognitive fiber levels promote cognitive dysfunction in patients with existing white matter lesions. Compared with the AD patients, plasma cognitive fiber levels were significantly increased in the vascular dementia group; cognitive fiber elevation was significantly associated with oral memory impairment in patients. The patients with mild cognitive impairment with high cognitive fiber levels are more likely to develop dementia, suggesting that controlling plasma cognitive fiber levels in patients with mild cognitive impairment may help delay cognitive decline. The specific mechanisms by which cognitive fibers cause cognitive dysfunction are not yet clear. In the animal model of AD, it is found that the regional cognitive fiber level of axonal dystrophy is elevated, and the cognitive fiber level is moderately reduced, which can reduce neuronal damage and amyloid deposition in the brain. Elevated cognitive fibers can also promote changes in blood coagulation function, which in turn affect platelet aggregation and vascular endothelial cell function, increase leukoaraiosis, and cause cognitive decline. In addition, cognitive fibers can promote platelet aggregation, increase blood viscosity, and promote thrombosis. Increased cognitive fibers can aggravate the degree of intracranial and extracranial stenosis and increase the recurrence rate of cerebral infarction, thus increasing cognitive dysfunction [18].

White matter lesions have long been a widespread concern as a common brain imaging change and they usually coexist with stroke and neurodegenerative diseases. A large number of imaging studies have confirmed that most patients with AD have varying degrees of white matter lesions, and more importantly, more and more studies in recent years have shown that white matter lesions can be used as mild cognitive impairment or a neuroimaging marker of AD. However, severe white matter lesions are almost impossible to reverse or repair, and early recognition of mild white matter lesions is particularly important for the brain. As the pia matter undulates and spreads on the surface of the gyrus, it belongs to the secondary collateral circulation; the effects of early white matter lesions on cognition, brain structure and function, and their relationship in order to provide a basis for clinicians to better understand their pathogenic mechanisms [19]. In the past, some studies have pointed out that no significant cognitive impairment was found in mild white matter lesions and one of the reasons may be that multiple factors that are beneficial and unfavourable for maintaining normal cognitive function have multiple effects on the decline of neuroprotective mechanisms. Therefore, different degrees of cognitive decline occur in different stages of white matter lesions from mild to severe.

## 6. Conclusions

In order to assess the relationship between structural and functional imaging of cerebrovascular disease and cognition-related fibers, this paper chooses a total of 120 patients who underwent cerebral small vessel disease treatment at a designated hospital by this study from June 2013 to June 2018 and divides them into 3 groups according to the random number table method: vascular dementia group, vascular cognitive impairment no dementia group, and noncognition impairment group with 40 cases of patients in each group. Cognitive function measurement and imaging examination were performed for these 3 groups of patients, and the observation indicators of CSE, MAS, CDT, AIS, FAB, VFT, TMT, CI, WML, TVW, and FHI were tested, perceptively. All patients were examined by magnetic resonance imaging to obtain diffusion coefficient parameters; after admission, the MRI was performed with magnetic resonance spectroscopy. The relevant parameters were set according to the actual situation of each patient and the statistics were used and the data were analyzed by software SPSS 19.0; the data of the measurement data were expressed by the mean  $\pm$  standard deviation and the *t* test was used for comparison between the two groups. The statistical analysis was performed after the normal distribution was not converted into a normal distribution, and the one-way analysis was used for comparison between groups. The LSD-*t* test was used to compare the two groups and the difference was statistically significant at  $P < 0.05$ .

The results shows that the average scores of CSE, MAS, AIS, and VTF in the VaD and VCIND group are lower than those of the NCI group and the differences are statistically significant ( $P < 0.05$ ); the average scores of FAB, TMT, and CI in the VaD group are higher than those of the VCIND group and the differences are also statistically significant ( $P < 0.05$ ); the average scores of FHI and TVW in the VaD group are lower than those of the VCIND and NCI group with statistically significant differences ( $P < 0.05$ ); the average scores of WML, CDT, and AIS in the VaD group are higher than those of the VCIND and NCI group with statistically significant differences ( $P < 0.05$ ). Therefore, it is believed that the structural and functional imaging features of cerebrovascular disease are closely related to cognition-related fibers, and the incidence of white matter lesions is closely related to the degree of lesions and cognitive dysfunction of cerebral small vessel disease, in which a major risk factor for cognitive dysfunction in patients with small blood vessels is the severity of white matter lesions. Imaging features are high signals that are spotted, patchy, or fused on a weighted basis, equal or low signals on a weighted basis, and their signals depend on the image sequence parameters and the degree of lesions. The integrity of cerebral blood circulation is closely related to the occurrence and prognosis of cerebral infarction, especially in the internal carotid artery stenosis or occlusion; in patients with severe stenosis or occlusion of the unilateral internal carotid artery, anterior communicating artery opening is a predictor of hemodynamics. The results of this study provide a reference for further research on the relationship between structural and

functional imaging and cognition-related fiber relationship in cerebrovascular disease.

## Data Availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Consent

Informed consent was obtained from all individual participants included in the study references.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This study was supported by the Natural Science Foundation of Jiangxi Province, China (20192BAB205085), and the applied research and cultivation program of Science and Technology Department of Jiangxi Province, China (20181BBG78065).

## References

- [1] J.-H. Park, S. W. Seo, C. Kim et al., "Effects of cerebrovascular disease and amyloid beta burden on cognition in subjects with subcortical vascular cognitive impairment," *Neurobiology of Aging*, vol. 35, no. 1, pp. 254–260, 2014.
- [2] B. S. Ye, S. W. Seo, G. H. Kim et al., "Amyloid burden, cerebrovascular disease, brain atrophy, and cognition in cognitively impaired patients," *Alzheimer's & Dementia*, vol. 11, no. 5, pp. 494–503, 2015.
- [3] B. Friedler, J. Crapser, and L. McCullough, "One is the deadliest number: the detrimental effects of social isolation on cerebrovascular diseases and cognition," *Acta Neuropathologica*, vol. 129, no. 4, pp. 493–509, 2015.
- [4] M. L. Alosco, J. Duskin, L. M. Besser et al., "Modeling the relationships among late-life body mass index, cerebrovascular disease, and Alzheimer's disease neuropathology in an autopsy sample of 1,421 subjects from the national Alzheimer's coordinating center data set," *Journal of Alzheimer's Disease*, vol. 57, no. 3, pp. 953–968, 2017.
- [5] B. Haring, A. Omidpanah, A. M. Suchy-Dicey et al., "Left ventricular mass, brain magnetic resonance imaging, and cognitive performance," *Hypertension*, vol. 70, no. 5, pp. 964–971, 2017.
- [6] X. Xu, S. Hilal, S. L. Collinson et al., "Validation of the total cerebrovascular disease burden scale in a community sample," *Journal of Alzheimer's Disease*, vol. 52, no. 3, pp. 1021–1028, 2016.
- [7] D. R. Jorgensen, C. Rosano, and E. M. Novelli, "Can neuroimaging markers of vascular pathology explain cognitive performance in adults with sickle cell anemia? A review of the literature," *Hemoglobin*, vol. 40, no. 6, pp. 381–387, 2016.
- [8] I. Bos, F. R. Verhey, I. H. G. B. Ramakers et al., "Cerebrovascular and amyloid pathology in predementia stages: the relationship with neurodegeneration and cognitive decline," *Alzheimer's Research and Therapy*, vol. 9, no. 1, p. 101, 2017.
- [9] R. M. Lazar, M. A. Pavol, T. Bormann et al., "Neurocognition and cerebral lesion burden in high-risk patients before undergoing transcatheter aortic valve replacement," *JACC: Cardiovascular Interventions*, vol. 11, no. 4, pp. 384–392, 2018.
- [10] Q. Dong, S.-W. Teng, Y. Wang et al., "Sitagliptin protects the cognition function of the Alzheimer's disease mice through activating glucagon-like peptide-1 and BDNF-TrkB signalings," *Neuroscience Letters*, vol. 696, pp. 184–190, 2019.
- [11] F. Prada, M. Y. S. Kalani, K. Yagmurlu et al., "Applications of focused ultrasound in cerebrovascular diseases and brain tumors," *Neurotherapeutics*, vol. 16, no. 1, pp. 67–87, 2019.
- [12] S.-T. Chan, K. C. Evans, B. R. Rosen, T.-Y. Song, and K. K. Kwong, "A case study of magnetic resonance imaging of cerebrovascular reactivity: a powerful imaging marker for mild traumatic brain injury," *Brain Injury*, vol. 29, no. 3, pp. 403–407, 2015.
- [13] D. Kashiwazaki, N. Akioka, N. Kuwayama, K. Noguchi, K. Tanaka, and S. Kuroda, "Pathophysiology of acute cerebrovascular syndrome in patients with carotid artery stenosis," *Neurosurgery*, vol. 76, no. 4, pp. 427–434, 2015.
- [14] W. D. Heiss, "Contribution of neuro-imaging for prediction of functional recovery after ischemic stroke," *Cerebrovascular Diseases*, vol. 44, no. 5-6, pp. 266–276, 2017.
- [15] S. Griffiths, S. Wood, R. Reniers, and M. Birchwood, "O8.6. The relationship between cognition and functional improvement in the context of a psychosocial intervention targeting social disability in first episode psychosis," *Schizophrenia Bulletin*, vol. 44, no. suppl\_1, p. S98, 2018.
- [16] G. Özsoy, "An investigation of the relationship between metacognition and mathematics achievement," *Asia Pacific Education Review*, vol. 12, no. 2, pp. 227–235, 2011.
- [17] D. Mungas, B. R. Reed, W. J. Jagust et al., "Volumetric MRI predicts rate of cognitive decline related to AD and cerebrovascular disease," *Neurology*, vol. 59, no. 6, pp. 867–873, 2002.
- [18] K. Kantarci, R. C. Petersen, S. A. Przybelski et al., "Hippocampal volumes, proton magnetic resonance spectroscopy metabolites, and cerebrovascular disease in mild cognitive impairment subtypes," *Archives of Neurology*, vol. 65, no. 65, pp. 1621–1628, 2008.
- [19] M. Katan and M. S. Elkind, "Infectious burden and its role in cerebrovascular disease and cognitive impairment," *Future Virology*, vol. 8, no. 9, pp. 833–836, 2013.