

Biomedical Signal Processing Using Non-Invasive Sensors for Healthcare

Lead Guest Editor: Humaira Nisar

Guest Editors: Dr. G R Sinha and Siti Anom Ahmad





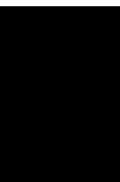
Biomedical Signal Processing Using Non-Invasive Sensors for Healthcare

Journal of Healthcare Engineering

Biomedical Signal Processing Using Non-Invasive Sensors for Healthcare

Lead Guest Editor: Humaira Nisar

Guest Editors: Dr. G R Sinha and Siti Anom Ahmad



Copyright © 2021 Hindawi Limited. All rights reserved.

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

Associate Editors

Xiao-Jun Chen , China
Feng-Huei Lin , Taiwan
Maria Lindén, Sweden

Academic Editors







Cherif Adnen, Tunisia
Saverio Affatato , Italy
Óscar Belmonte Fernández, Spain
Sweta Bhattacharya , India
Prabadevi Boopathy , India
Weiwei Cai, USA
Gin-Shin Chen , Taiwan
Hongwei Chen, USA
Daniel H.K. Chow, Hong Kong
Gianluca Ciardelli , Italy
Olawande Daramola, South Africa
Elena De Momi, Italy
Costantino Del Gaudio , Italy
Ayush Dogra , India
Luobing Dong, China
Daniel Espino , United Kingdom
Sadiq Fareed , China
Mostafa Fatemi, USA
Jesus Favela , Mexico
Jesus Fontecha , Spain
Agostino Forestiero , Italy
Jean-Luc Gennisson, France
Badicu Georgian , Romania
Mehdi Gheisari , China
Luca Giancardo , USA
Antonio Gloria , Italy
Kheng Lim Goh , Singapore
Carlos Gómez , Spain
Philippe Gorce, France
Vincenzo Guarino , Italy
Muhammet Gul, Turkey
Valentina Hartwig , Italy
David Hewson , United Kingdom
Yan Chai Hum, Malaysia
Ernesto Iadanza , Italy
Cosimo Ieracitano, Italy

Giovanni Improta , Italy
Norio Iriguchi , Japan
Mihajlo Jakovljevic , Japan
Rutvij Jhaveri, India
Yizhang Jiang , China
Zhongwei Jiang , Japan
Rajesh Kaluri , India
Venkatachalam Kandasamy , Czech Republic
Pushpendu Kar , India
Rashed Karim , United Kingdom
Pasi A. Karjalainen , Finland
John S. Katsanis, Greece
Smith Khare , United Kingdom
Terry K.K. Koo , USA
Srinivas Koppu, India
Jui-Yang Lai , Taiwan
Kuruva Lakshmanna , India
Xiang Li, USA
Lun-De Liao, Singapore
Qiu-Hua Lin , China
Aiping Liu , China
Zufu Lu , Australia
Basem M. ElHalawany , Egypt
Praveen Kumar Reddy Maddikunta , India
Ilias Maglogiannis, Greece
Saverio Maietta , Italy
M.Sabarimalai Manikandan, India
Mehran Moazen , United Kingdom
Senthilkumar Mohan, India
Sanjay Mohapatra, India
Rafael Morales , Spain
Mehrbakhsh Nilashi , Malaysia
Sharnil Pandya, India
Jialin Peng , China
Vincenzo Positano , Italy
Saeed Mian Qaisar , Saudi Arabia
Alessandro Ramalli , Italy
Alessandro Reali , Italy
Vito Ricotta, Italy
Jose Joaquin Rieta , Spain
Emanuele Rizzuto , Italy

Dinesh Rokaya, Thailand
Sébastien Roth, France
Simo Saarakkala , Finland
Mangal Sain , Republic of Korea
Nadeem Sarwar, Pakistan
Emiliano Schena , Italy
Prof. Asadullah Shaikh, Saudi Arabia
Jiann-Shing Shieh , Taiwan
Tiago H. Silva , Portugal
Sharan Srinivas , USA
Kathiravan Srinivasan , India
Neelakandan Subramani, India
Le Sun, China
Fabrizio Taffoni , Italy
Jinshan Tang, USA
Ioannis G. Tollis, Greece
Ikram Ud Din, Pakistan
Sathishkumar V E , Republic of Korea
Cesare F. Valenti , Italy
Qiang Wang, China
Uche Wejinya, USA
Yuxiang Wu , China
Ying Yang , United Kingdom
Elisabetta Zanetti , Italy
Haihong Zhang, Singapore
Ping Zhou , USA



Contents

Combining Rhythm Information between Heartbeats and BiLSTM-Treg Algorithm for Intelligent Beat Classification of Arrhythmia

Jinliang Yao , Runchuan Li , Shengya Shen, Wenzhi Zhang , Yan Peng , Gang Chen , and Zongmin Wang 

Research Article (16 pages), Article ID 8642576, Volume 2021 (2021)

The Ensemble Machine Learning-Based Classification of Motor Imagery Tasks in Brain-Computer Interface

Abdulhamit Subasi  and Saeed Mian Qaisar 








Research Article (12 pages), Article ID 1970769, Volume 2021 (2021)

Interrater and Intrarater Reliability of Electrical Impedance Myography: A Comparison between Large and Small Handheld Electrode Arrays

Huijing Hu , Wai Leung Ambrose Lo , Xiaoyun Wang, Le Li , and Ping Zhou 

Research Article (8 pages), Article ID 7296322, Volume 2021 (2021)

Persistent Homology-Based Topological Analysis on the Gestalt Patterns during Human Brain Cognition Process

Zaisheng Liu , Fei Ni , Rongpeng Li , Honggang Zhang , Chang Liu , Jiefang Zhang , and Songyun Xie 


Research Article (11 pages), Article ID 2334332, Volume 2021 (2021)

Investigating Feature Ranking Methods for Sub-Band and Relative Power Features in Motor Imagery Task Classification

Samrudhi Mohdiwale , Mridu Sahu , G. R. Sinha , and Humaira Nisar 




Research Article (11 pages), Article ID 3928470, Volume 2021 (2021)

Complexity and Entropy Analysis to Improve Gender Identification from Emotional-Based EEGs

Noor Kamal Al-Qazzaz , Mohannad K. Sabir, Sawal Hamid Bin Mohd Ali, Siti Anom Ahmad, and Karl Grammer






Research Article (17 pages), Article ID 8537000, Volume 2021 (2021)

Stress Classification by Multimodal Physiological Signals Using Variational Mode Decomposition and Machine Learning

Nilima Salankar , Deepika Koundal , and Saeed Mian Qaisar 

Research Article (12 pages), Article ID 2146369, Volume 2021 (2021)

Assessment of Acoustic Features and Machine Learning for Parkinson's Detection

Moumita Pramanik , Ratika Pradhan , Parvati Nandy , Saeed Mian Qaisar , and Akash Kumar Bhoi 

Research Article (13 pages), Article ID 9957132, Volume 2021 (2021)

Kinematic Evaluation via Inertial Measurement Unit Associated with Upper Extremity Motor Function in Subacute Stroke: A Cross-Sectional Study






Ze-Jian Chen , Chang He, Ming-Hui Gu, Jiang Xu , and Xiao-Lin Huang 

Research Article (7 pages), Article ID 4071645, Volume 2021 (2021)


Initial Geometrical Templates with Parameter Sets for Active Contour on Skin Cancer Boundary Segmentation

Prachya Bumrungkun, Kosin Chamnongthai , and Wisarn Patchoo 
Research Article (30 pages), Article ID 9528460, Volume 2021 (2021)

Multiple Sclerosis Lesion Segmentation in Brain MRI Using Inception Modules Embedded in a Convolutional Neural Network

Shahab U. Ansari , Kamran Javed , Saeed Mian Qaisar , Rashad Jillani , and Usman Haider 
Research Article (10 pages), Article ID 4138137, Volume 2021 (2021)

ECG Signal Modeling Using Volatility Properties: Its Application in Sleep Apnea Syndrome

Maryam Faal  and Farshad Almasganj 
Research Article (12 pages), Article ID 4894501, Volume 2021 (2021)





EEG-Based Closed-Loop Neurofeedback for Attention Monitoring and Training in Young Adults

Bingbing Wang , Zeju Xu , Tong Luo , and Jiahui Pan 
Research Article (13 pages), Article ID 5535810, Volume 2021 (2021)






Classification of Mental Stress Using CNN-LSTM Algorithms with Electrocardiogram Signals

Mingu Kang, Siho Shin, Jaehyo Jung , and Youn Tae Kim 
Research Article (11 pages), Article ID 9951905, Volume 2021 (2021)

Reduce Surface Electromyography Channels for Gesture Recognition by Multitask Sparse Representation and Minimum Redundancy Maximum Relevance

Yali Qu , Haoyan Shang , Jing Li , and Shenghua Teng 
Research Article (9 pages), Article ID 9929684, Volume 2021 (2021)

Optimizing Residual Networks and VGG for Classification of EEG Signals: Identifying Ideal Channels for Emotion Recognition

Kit Hwa Cheah , Humaira Nisar , Vooi Voon Yap , Chen-Yi Lee , and G. R. Sinha 
Research Article (14 pages), Article ID 5599615, Volume 2021 (2021)

Research Article

Combining Rhythm Information between Heartbeats and BiLSTM-Treg Algorithm for Intelligent Beat Classification of Arrhythmia

Jinliang Yao ^{1,2}, Runchuan Li ^{1,2}, Shengya Shen,^{1,3} Wenzhi Zhang ^{1,2}, Yan Peng ^{1,2},
Gang Chen ^{1,2} and Zongmin Wang ^{1,2}

¹School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China

²Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou 450000, China

³Zhengzhou College of Economics and Business, Zhengzhou, Henan 450000, China

Correspondence should be addressed to Runchuan Li; rcli@ha.edu.cn and Gang Chen; gchen@ha.edu.cn

Received 17 August 2021; Accepted 23 November 2021; Published 13 December 2021

Academic Editor: G R Sinha

Copyright © 2021 Jinliang 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.

Arrhythmia is a cardiovascular disease that seriously affects human health. The identification and diagnosis of arrhythmia is an effective means of preventing most heart diseases. In this paper, a BiLSTM-Treg algorithm that integrates rhythm information is proposed to realize the automatic classification of arrhythmia. Firstly, the discrete wavelet transform is used to denoise the ECG signal, based on which we performed heartbeat segmentation and preserved the timing relationship between heartbeats. Then, different heartbeat segment lengths and the BiLSTM network model are used to conduct multiple experiments to select the optimal heartbeat segment length. Finally, the tree regularization method is used to optimize the BiLSTM network model to improve classification accuracy. And the interpretability of the neural network model is analyzed by analyzing the simulated decision tree generated in the tree regularization method. This method divides the heartbeat into five categories (nonectopic (N), supraventricular ectopic (S), ventricular ectopic (V), fused heartbeats (F), and unknown heartbeats (Q)) and is validated on the MIT-BIH arrhythmia database. The results show that the overall classification accuracy of the algorithm is 99.32%. Compared with other methods of classifying heartbeat, the BiLSTM-Treg network model algorithm proposed in this paper not only improves the classification accuracy and obtains higher sensitivity and positive predictive value but also has higher interpretability.

1. Introduction

With the improvement of people's living standards, the incidence and mortality of cardiovascular diseases are increasing year by year and are accompanied by a younger trend [1]. Arrhythmia is a common cardiovascular disease, which may endanger people's lives in serious cases [2]. Therefore, the accurate detection of arrhythmia to prevent heart disease has a very important significance. Electrocardiogram (ECG), as a comprehensive expression of cardiac electrical activity on the body surface, contains a wealth of physiological and pathological information reflecting cardiac rhythm and electrical conduction and is one of the important bases for diagnosis of heart disease and evaluation of cardiac function [3]. Different types of arrhythmias can be identified and diagnosed by

analyzing the ECG waveform. Traditional ECG waveform analysis is performed manually by medical personnel, who need to give a diagnosis based on cardiovascular disease diagnosis rules and personal experience. Due to the individual differences of patients and the complexity of diseases, there are many types of ECG. In addition, some arrhythmias occur only occasionally in the daily life of the patients, and the ECG data need to be recorded for a long time. Therefore, the amount of collected ECG data is huge, which brings a heavy burden to doctors. Under the circumstances, mistakes, missed inspections, or misdetections easily occur. With the rapid development of computer technology and electronic information technology, the computer has become an indispensable and important tool of medical modernization, and computer-aided medical treatment has penetrated into every

corner of medical service [4]. In recent years, increasing attention has been paid to the study of computer-aided analysis algorithms for electrocardiography, particularly those that can accurately and rapidly identify and diagnose arrhythmias. The automatic classification and diagnosis algorithm of ECG signals can save doctors' time by helping them better judge the symptoms of arrhythmia quickly. In addition, it can provide good healthcare in areas where medical resources are scarce.

This paper presents a beat classification method based on the time-series network, which integrates the interheartbeat rhythm information. This method is based on tree regularization constraints and the BiLSTM neural network model. This method improves the accuracy of heartbeat classification. And the interpretability of the proposed algorithm is analyzed by tree regularization constraints and feature analysis. The main contributions of this work are as follows:

- (1) A time-series BiLSTM-Treg algorithm was designed to classify the beats, which combined the information of the beats so that the deep neural network could learn more rhythm information between heartbeats.
- (2) A tree regularization method for the heartbeat classification model is proposed to optimize the BiLSTM-Treg algorithm and improve the generalization ability of the neural network model.
- (3) By analyzing the key nodes of the simulated decision tree in tree regularization, the concerns in the learning process of the BiLSTM-Treg algorithm are analyzed, and the interpretability of the model is analyzed to a certain extent.
- (4) Compared with other deep learning methods, the proposed BiLSTM-Treg algorithm improves the accuracy of heartbeat classification and reduces doctor's misdiagnosis rate to a certain extent

2. Related Work

The diagnosis of early arrhythmia is mainly the doctor's manual analysis of ECG waveform, which requires the doctor to have a professional medical theoretical basis and rich clinical experience. Because of the diversity of arrhythmia and the complexity of the ECG waveform, this method cannot meet the needs of patients. With the development of artificial intelligence, the classification of arrhythmia using intelligent processing technology has become a hot topic in recent years.

In the 1950s, the automatic analysis technology of ECG signals has appeared in the field of ECG research. At first, only ECG filtering processing technology developed relatively mature. Later, with the continuous development of technology, automatic detection and diagnosis of arrhythmia disease also began to be continuously explored by researchers. In the past decades, domestic and foreign ECG researchers have proposed a variety of heartbeat classification methods. These methods can be divided into two categories from the perspective of whether manual feature

extraction of ECG signals is needed: feature engineering-based classification methods and deep learning-based methods [5]. Traditional rule-based and machine-learning-based heartbeat classification methods both require manual feature extraction.

2.1. Heartbeat Classification Method Based on Feature Engineering. Feature engineering is to process a series of original data and extract the features as the input of the model to improve the performance of the model. Feature engineering mainly includes three aspects: feature selection, feature extraction, and feature construction. Feature extraction is the key step of ECG signal classification and recognition, and the extracted feature quality will affect the accuracy of ECG signal classification and recognition [6]. Generally, the features of ECG signals extracted by researchers mainly include morphological features [7], interphase features [8, 9], wavelet transform features [10], higher-order statistics (HOS) [9, 11], Hermite basis function (HBF) [12], QRS amplitude vector [13], and QRS composite wave area [14]. Then machine-learning algorithms are used for classification, such as the KNN algorithm [15], support vector machine (SVM) [7], and random forest [9]. Zhu et al. [7] extracted the ECG morphological features and used the SVM algorithm to classify the heartbeat, achieving a high classification accuracy. Yang et al. [9] extracted a variety of features, including RR interval, wavelet coefficient, and high-order statistics, and then used the random forest classifier based on an extreme learning machine to detect arrhythmias. Ji et al. [15] proposed a multifeature combination and stacked DWKNN algorithm to classify arrhythmias. The effects of different characteristic combinations on the classification of the heartbeat were analyzed.

Although this method based on feature engineering can also achieve relatively high classification accuracy, because of the complex waveform and poor anti-interference ability of ECG signal, the features extracted by hand often produce the human error. And the features of the manual design are very dependent on the prior knowledge of the researcher. Deep learning has the advantage of automatically extracting features and classification, which well solves a series of problems caused by manual feature extraction.

2.2. Heartbeat Classification Method Based on Deep Learning. The deep learning model has become a common model for ECG data classification [16]. Compared with the feature engineering-based ECG classification method, the deep learning method, which uses original data rather than manually extracted features as input, can achieve better classification performance. In the deep learning method, researchers use the nonlinear transformation of hidden layers in the network to automatically obtain effective features and transform the original features into different new feature spaces by changing the structure of hidden layers in the network and the way of stacking [17], so as to make full use of the rich hidden information in the data and improve the classification accuracy.

Recently, some researchers [18, 19] have used a deep neural network model for automatic classification of ECG signals. Ji et al. [20] proposed an ECG classification system based on Faster R-CNN. One-dimensional ECG signal is converted into two-dimensional image as the input of neural network to realize the classification of arrhythmias. Akarya et al. [21] proposed a 9-layer deep convolutional neural network (CNN) for automatic recognition of ECG signals. The original ECG signal and the ECG signal filtered out the high-frequency noise were used to classify the heartbeat, and the accuracy rates were 94.03% and 93.47%, respectively. Khan et al. [22] used the long short-term memory network (LSTM) to automatically identify 16 different types of arrhythmias. Wu et al. [23] proposed a heartbeat classification algorithm that integrated CNN and BiLSTM deep learning models and extracted the morphological and temporal features of heartbeat, respectively, by using CNN and BiLSTM. Li et al. [24] proposed a BiLSTM-Attention Network model to distinguish different types of arrhythmias. Pandey et al. [25] applied the extracted features of wavelet, RR interval, morphology, and high-order statistics to BiLSTM to achieve the automatic classification of the heartbeat. Yildirim et al. [26] proposed a heartbeat classification model based on wavelet transform and BiLSTM network, which used wavelet to decompose ECG signals into signals of different frequency scales and used the signals as the input sequence of the BiLSTM model.

The classification method of ECG signals based on deep learning realizes the “end-to-end” learning mode, eliminates the manual design process of features, saves manpower, and makes the process of ECG classification simpler and more efficient. Although all the above studies cleverly used the deep neural network to classify ECG signals, the rhythm information between heartbeats has not been fully considered, the interpretability of the network has not been analyzed, and the classification accuracy needs to be improved.

3. Method

The heartbeat classification method of the BiLSTM-Treg algorithm that integrates rhythm information between heartbeats proposed in this paper mainly includes the following steps: firstly, the data are preprocessed to filter out the noise in the ECG signal and segment ECG signal into heartbeats. Secondly, the continuous single heartbeat is combined into heartbeat segments so that the rhythm information between the heartbeats can be retained. Then, the BiLSTM-Treg model was constructed and optimized. Finally, the heartbeats were classified. Section 3.1 is the preprocessing part, Section 3.2 is the representation of the rhythm information part, and Section 3.3 is the model building and optimization part.

3.1. ECG Signal Preprocessing. The preprocessing stage is mainly denoising and segmentation of ECG signals. Generally speaking, the collected ECG signals inevitably contain noise due to the influence of equipment and human body itself [27], which mainly includes baseline drift, power frequency

interference, and EMG interference. It is important to remove as much noise as possible from ECG signals before classifying them. Wavelet transform is a generalization of short-time Fourier transform (STFT) [28], which can perform time-frequency analysis of ECG signals well. Compared with the equally spaced time-frequency localization of STFT, wavelet transform can provide higher frequency resolution at low frequency and higher time resolution at high frequency. In this paper, discrete wavelet transform is used to denoise ECG signals, which can avoid losing important physiological details in ECG signals and better retain the characteristics of ECG signals. Because of the high regularity of the Daubechies wavelet, the reconstructed signal is relatively smooth. And the strength spectrum of the DB6 wavelet [29, 30] is focused on low frequencies. Its moderate filter length and moderate coefficient values, compared with the other wavelets, provide more smoothing and less shift in the ECG fiducials. Therefore, in order to obtain a good classification accuracy, this paper uses the DB6 wavelet in the Daubechies wavelet base to process ECG signals. In terms of implementation, we use python’s open-source wavelet transform tool pywt. The discrete wavelet transform formula [31] is shown in (1) and (2).

$$W_{\Psi}(j, k) = \sum_x f(x) \Psi_{j,k}(x), \quad (1)$$

$$\Psi_{j,k}(x) = a_0^{-j/2} \Psi(a_0^{-j} x - kb_0), \quad (2)$$

where $W_{\Psi}(j, k)$ is the wavelet coefficient, $\Psi_{j,k}(x)$ is the discrete wavelet function at different scales and locations, $f(x)$ is the input ECG signal, $\Psi(k)$ is the wavelet basis function, and j is the order of the scale. The larger j is, the smaller the scale is, which means the higher the frequency is and the closer it is to the details. k is the offset of position. a_0 is the scale parameter and b_0 is the position parameter. Signal comparison before and after pretreatment with discrete wavelet transform is shown in Figure 1 and Figure 2.

Heartbeat segmentation is to divide an ECG record with a complete heartbeat as a unit [32]. A complete heartbeat should contain P wave, QRS compound wave, and T wave [33], as shown in Figure 3(a). In this paper, the peak value of the R wave marked in the MIT-BIH database was used as the reference point for heartbeat segmentation, and 0.25s and 0.4s were extracted before and after the peak of R, as shown in Figure 3(b). We take this 0.65s data as a sample of a single heartbeat. For MIT-BIH ECG data with a sampling rate of 360HZ, we extracted 90 points before R peak and 144 points after R peak. Therefore, the reconstructed sample is 235 points.

3.2. Rhythm Information between Heartbeats. The rhythm information between heartbeats contained in the ECG is an important basis for doctors to diagnose heart diseases. Changes in ECG rhythm can reflect problems in different parts of the heart, which can help medical staff design more rational treatment plans. Common rhythm types are bigeminy, trigeminy, ventricular tachycardia, and atrial tachycardia.

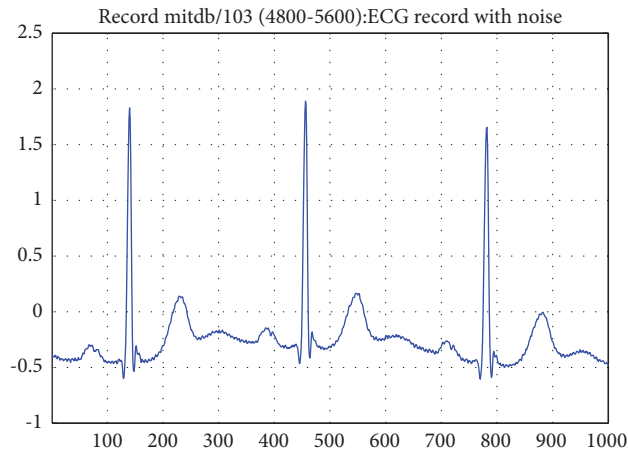


FIGURE 1: The original ECG signal.

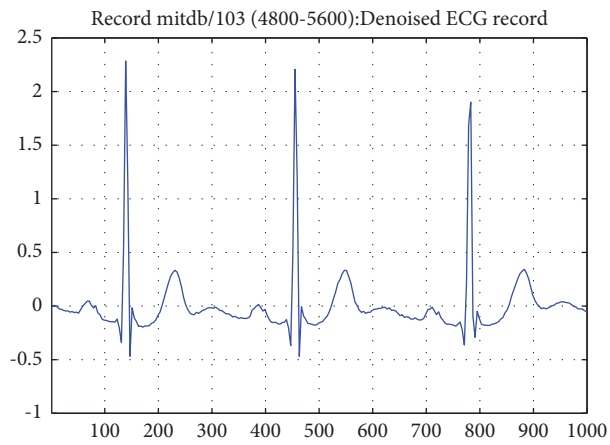


FIGURE 2: The denoised ECG signal.

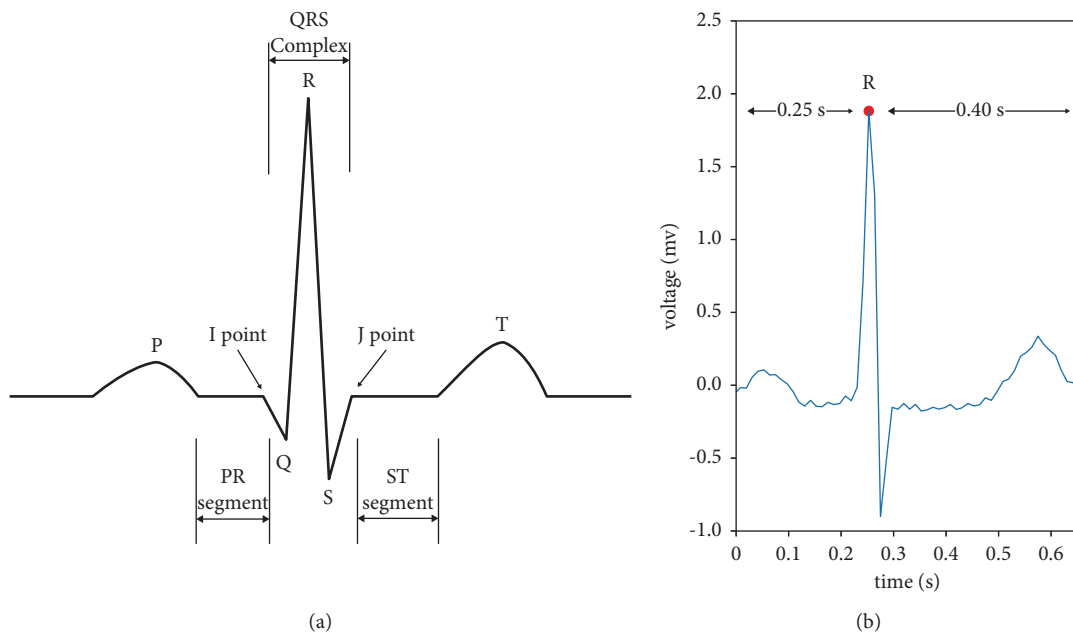


FIGURE 3: Morphology of a single heartbeat and heartbeat segmentation.

Bigeminy. Every normal heartbeat is followed by a premature beat. And the occurrence of three or more groups in a row is called bigeminy. According to the type of premature beat, it can be divided into ventricular bigeminy and atrial bigeminy. For example, the rhythm change of N-V-N-V-N-V is ventricular bigeminy, and the rhythm change of N-S-N-S-N-S is atrial bigeminy.

Trigeminy. A premature beat occurs after every two normal heartbeats. And the occurrence of three or more groups in a row is called trigeminy. According to the type of premature beat, it can be divided into ventricular trigeminy and atrial trigeminy. For example, the rhythm change of N-N-V-N-N-V-N-N-V is ventricular trigeminy, and the rhythm change of N-N-S-N-N-S-N-N-S is atrial trigeminy. The ECG signal with ventricular trigeminy is shown in Figure 4.

Ventricular Tachycardia. Three or more consecutive ventricular premature beats are called ventricular tachycardia, such as the rhythm change V-V-V.

Atrial Tachycardia. Three or more consecutive atrial premature beats are called atrial tachycardia, such as the rhythm change S-S-S.

In addition, the appearance of certain types of heartbeats also reflects changes in ECG rhythm. For example, after a continuous ventricular tachycardia, a ventricular fusion heartbeat is often generated due to electrical signals from the sinus node, followed by ventricular capture. Therefore, ventricular fusion heartbeat and ventricular capture are important characteristics of ventricular tachycardia.

In this paper, this rhythmic information, which is beneficial to the classification of heartbeats, was integrated into the model. Specifically, in processing the dataset, successive single beats were grouped into segments, which preserved information about rhythm between beats. Then, the ECG data is input into the neural network model in the unit of heartbeat segment, which enables the model to make full use of the rhythm information contained in the heartbeat segment when identifying the heartbeat type and improves the classification accuracy. The length of the heartbeat segment is one of the key points of our study.

3.3. BiLSTM-Treg Algorithm. Recurrent neural network (RNN) is a kind of neural network with short-term memory ability, which is very effective in processing data with sequence characteristics. However, in deep neural networks, the gradient is unstable. The gradient close to the input layer is calculated based on the product of the gradients of the subsequent layers [34]. When the neural network has too many hidden layers or the input sequence of the RNN network is too long, it will cause the gradient near the input layer to vanish or blow up, which affects the performance of RNN to some extent. In order to solve this problem, Hochreiter et al. [35] proposed the long short-term memory network (LSTM) in 1997. By adding gating units into RNN, the network can choose whether to retain the historical information so as to solve the problem of gradient

disappearance and gradient explosion caused by long-term dependence of the RNN network.

3.3.1. BiLSTM Neural Network Structure. Compared with RNN, LSTM adds three gating units, which are input gate, forgetting gate, and output gate. In addition, there are two important parts of LSTM, namely, memory unit, and hidden state. The forgetting gate controls whether the information in the memory unit is discarded, the input gate controls whether the information of the current signal and hidden state is added to the memory unit, and the output gate determines the information output in the memory unit. Figure 5 shows the unit structure of the LSTM, where f_t , i_t , and o_t , respectively, represent the forgetting gate at the current moment, the input gate, and the output gate; C_{t-1} and C_t , respectively, represent the state value of the memory unit at the previous moment and the current moment; h_{t-1} and h_t , respectively, represent the hidden state at the previous moment and the current moment. x_t represents the input at the current moment, and \tilde{C}_t is the candidate value of the memory unit at the current moment. σ and \tanh represent the sigmoid activation function and tanh activation function, respectively.

The calculation process of LSTM can be expressed as equations (3–8):

$$i_t = \text{sigmoid}(W_i x_t + U_i h_{t-1} + b_i), \quad (3)$$

$$f_t = \text{sigmoid}(W_f x_t + U_f h_{t-1} + b_f), \quad (4)$$

$$o_t = \text{sigmoid}(W_o x_t + U_o h_{t-1} + b_o), \quad (5)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c), \quad (6)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t, \quad (7)$$

$$y_t = h_t = o_t \otimes \tanh(C_t). \quad (8)$$

Formulas (3)–(6), respectively, represent the calculation formulas for the input gate i_t , forget gate f_t , output gate o_t , and candidate value \tilde{C}_t of the memory unit. They are all determined by the input data x_t at the current moment, the hidden state h_{t-1} at the previous moment, and their corresponding weight matrix, where W_i , W_f , W_o , and W_c are the weight matrix of the current input x_t ; U_i , U_f , U_o , and U_c represent the weight matrix of the hidden state h_{t-1} at the last moment; b_i , b_f , b_o , and b_c are the corresponding bias items, respectively. These weight matrices and bias terms are trained by the way of gradient descent. Formula (7) indicates that the current moment memory unit C_t is adjusted by the current candidate unit \tilde{C}_t and its own state C_{t-1} as well as the input gate and the forgetting gate. Finally, formula (8) indicates that the output at the current moment, that is, the hidden state at the current moment, is determined by the current memory unit C_t and the output gate.

One disadvantage of LSTM is that it cannot encode information upfront and can only use its past context, not its future context. In the classification of heartbeat, if the

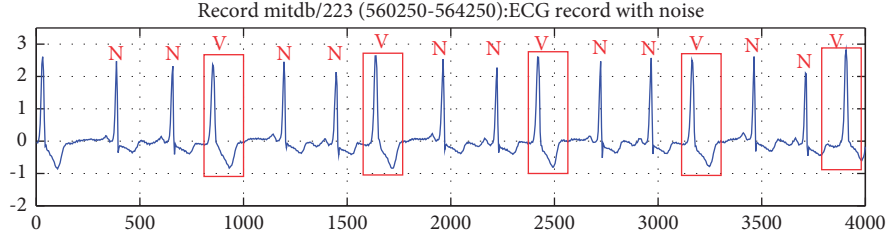


FIGURE 4: The ECG signal with ventricular trigeminy.

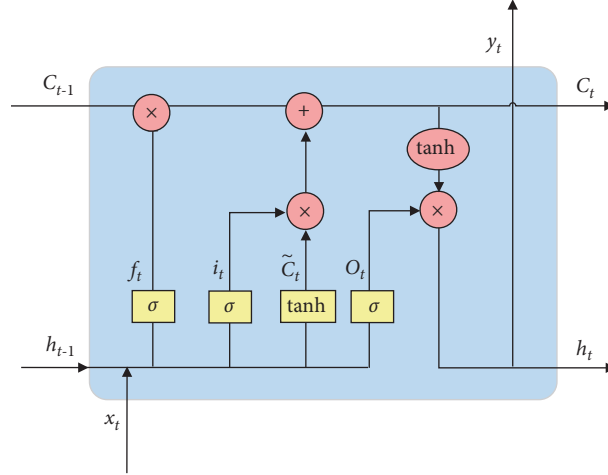


FIGURE 5: The LSTM cell structure.

relevant information of the former and the latter can be obtained at the same time during the classification of the current heartbeat, the rhythm information of the heartbeat will be grasped more accurately, thus improving the classification accuracy of the current heartbeat. And BiLSTM solves this problem well [36]. In each BiLSTM layer, there are two independent LSTM to process sequences in two directions, respectively. The specific formula is shown in (9)–(11). At the time t , the hidden layer state H_t of BiLSTM obtains the heartbeat information \vec{h}_t before the time t through the forward LSTM and the heartbeat information \overleftarrow{h}_t after the time t through the backward LSTM and then carries out the weighted sum of \vec{h}_t and \overleftarrow{h}_t , where \vec{W}_t and \overleftarrow{W}_t are the corresponding weight matrices and b_t is the bias term.

$$\vec{h}_t = \text{LSTM}\left(x_t, \vec{h}_{t-1}\right), \quad (9)$$

$$\overleftarrow{h}_t = \text{LSTM}\left(x_t, \overleftarrow{h}_{t-1}\right), \quad (10)$$

$$H_t = \vec{W}_t \vec{h}_t + \overleftarrow{W}_t \overleftarrow{h}_t + b_t. \quad (11)$$

3.3.2. BiLSTM Network Based on Tree Regularization. In machine learning, there are many strategies designed to reduce model generalization errors, which are collectively referred to as regularization. The form of regularization is

very simple, which is to add an additional term after the objective function to affect the selection of the optimal point of the objective function. The common regularization methods are L1 regularization and L2 regularization. The common regularization methods are L1 regularization and L2 regularization. The objective function is shown in equation (12), where $\lambda\Psi(W)$ is a regular term.

$$\min_W \sum_{n=1}^N \text{loss}(y_n, \hat{y}_n(x_n, W)) + \lambda\Psi(W). \quad (12)$$

Tree regularization is a new regularization method proposed by Wu et al. [37], which can not only effectively improve the generalization ability of the model but also analyze the interpretability of the model. The tree regularization method of deep network model interpretability is a postinterpretable method, that is, the method of applying model analysis after model training to make the model interpretable. This method looks for the decision tree representation of the deep network model and realizes the human understanding of the prediction results of the network model by improving the human simulability of the network model. The implementation method of tree regularization includes the following two stages. First, we train deep neural network while being closely modeled by decision trees. In this way, this decision tree can accurately simulate the prediction process of the network. Secondly, the complexity metric of the decision tree, the average path length (APL), is taken as the penalty term for model

optimization. In this way, the neural network can be encouraged to generate simple decision trees and restricted to generate complex decision trees, which further makes the generated decision trees easier to be simulated by human beings. The decision tree generation formula can be expressed by (13) and (14), where x_n is the sample feature of the training set, $\hat{y}_n(x_n, W)$ is the prediction label of the depth model, W is the weight matrix of the depth model, and \hat{y}_{tn} is the prediction label of the decision tree. The reason why \hat{y}_n is used as the input of the decision tree is to make \hat{y}_{tn} and \hat{y}_n as similar as possible so as to realize the purpose of using the decision tree to simulate the deep network.

$$\text{Tree} = \text{Traintree}(\{x_n, \hat{y}_n(x_n, W)\}), \quad (13)$$

$$\hat{y}_{tn} = \text{Tree.pre dict}(x_n). \quad (14)$$

The calculation formula of tree regularization is shown in (15), where $\text{Path Length}(\text{tree}, x_n)$ is the path length of the n sample and $\Omega(W)$ is the average path length, namely, the penalty term.

$$\Omega(W) = \frac{1}{N} \sum_n \text{Path Length}(\text{tree}, x_n). \quad (15)$$

It can be seen from equation (15) that is not differentiable for network parameter W . Therefore, in order to use the gradient descent strategy in the network optimization process, Wu et al. [37] proposed the surrogate regularization function $\tilde{\Omega}(W)$, which can surrogate the previous APL calculation method, as shown in equations (16) and (17). By training a Multilayer Perceptron (MLP), the mapping relationship between the parameter vector W of the neural network model and APL is established. With W and APL as inputs to MLP, the objective function of MLP is shown in equation (17), where ξ represents the weight matrix of the MLP model, ε represents the regularization intensity, $\{W_j, \Omega(W_j)\}$ represents the known parameter vectors and their corresponding real APL datasets, and J represents the total number of datasets. Therefore, after using the surrogate model, the objective function of the BiLSTM network is shown in equation (18).

$$\tilde{\Omega}(W) \approx \frac{1}{N} \sum_n \text{Path Length}(\text{tree}, x_n), \quad (16)$$

$$\min_{\xi} \sum_{j=1}^J (\Omega(W_j) - \tilde{\Omega}(W_j, \xi))^2 + \varepsilon \|\xi\|_2^2, \quad (17)$$

$$\min_W \sum_{n=1}^N \text{loss}(y_n, \hat{y}_n(x_n, W)) + \lambda \tilde{\Omega}(W). \quad (18)$$

In this paper, tree regularization is used in the BiLSTM model to optimize the model, reduce the generalization error of the model, and improve the classification accuracy. At the same time, the generated simulated decision tree is used to analyze and understand how the BiLSTM model carries out heartbeat classification. The BiLSTM model using tree regularization is shown in Figure 6. Specifically, $x_t = [x_{t1}, x_{t2}, \dots, x_{t235}]$ is used to represent a single

heartbeat sample. The heartbeat segment composed of consecutive single heartbeats is used as the input of the network, and the number of single heartbeats in the heartbeat segment t is the timestep of the network. The model first uses BiLSTM to classify heartbeat. Secondly, the decision tree is used to simulate BiLSTM, and APL is calculated. Then, the MLP model is trained to get the surrogate regularization function $\tilde{\Omega}(W)$, and then $\tilde{\Omega}(W)$ is added to the objective function of the BiLSTM model for the next round of training. Algorithm 1 describes the BiLSTM-Treg model algorithm.

4. Experiment

The processing and analysis of ECG signal is very important to the classification of the heartbeat. The research focus of this paper is on the construction and optimization of the model integrating rhythm information. According to the ANSI/AAMI EC57:2012 classification proposed by the Association for the Advancement of Medical Instruments (AAMI), arrhythmia can be divided into five categories: N (normal or bundle branch block), S (supraventricular ectopic beat), V (ventricular ectopic beat), F (fusion beat), and Q (beat not specified). On the basis of extracting continuous heartbeat segments, this experiment constructs a time-series network that integrates rhythm information between heartbeats and divides heartbeats into the above five types.

4.1. Experimental Environment. The model proposed in this paper is trained and tested on a PC workstation with Xeon(R)Silver-4114CPU, 32 GB memory, and Geforce2080Ti graphics card. The PC workstation runs on Ubuntu 18.04 system. And the algorithm is run under the TensorFlow-GPU V2.2.0 framework.

4.2. Experimental Data. A unified and authoritative standard database is the basis of the automatic analysis of ECG signals. In the research field of ECG signals, the MIT-BIH arrhythmia database is the most widely used database by researchers [38]. The database contains 48 records, each of which is about 30 minutes long, with about 650,000 sampling points and a sampling frequency of 360 Hz. Fifteen categories were labeled in the MIT-BIH arrhythmia database. Table 1 is the corresponding table of the two heartbeat classification methods.

In this paper, we classified 109,454 heartbeats from the MIT-BIH arrhythmia database, including 90,595 N-type heartbeats; 2,781 heartbeats in the S category; 7,235 V-type heartbeats. The number of heartbeats in category F was only 802 and in Q was 8041. In this paper, 90% of the heartbeat data were randomly selected from the dataset as the training set and the remaining 10% as the test set. And the specific distribution of data is shown in Table 2.

4.3. Evaluation Metrics. In order to calculate the performance of the model for heartbeat classification, the classification results were divided into four categories: TP, FP, TN,

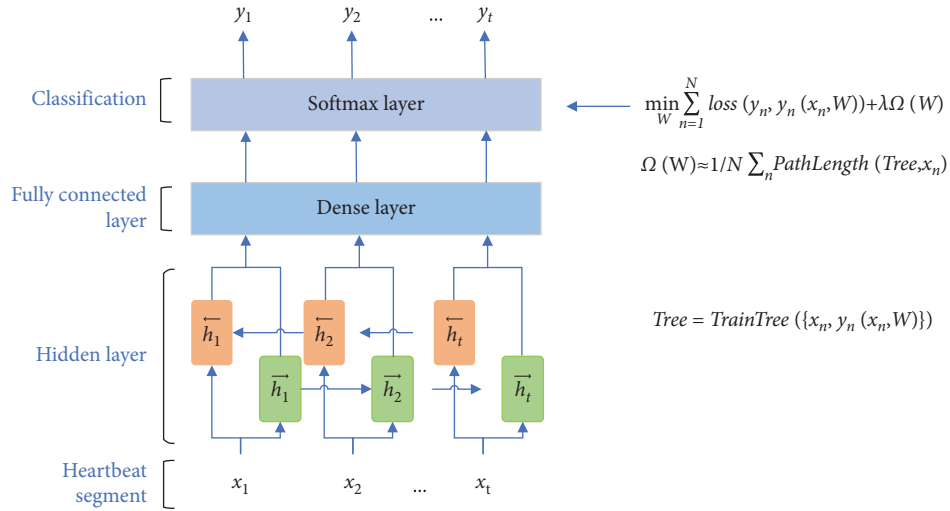


FIGURE 6: The BiLSTM model based on tree regularization.

Input: dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$
Output: classification results and model parameters
Process:

- (1) Build the BiLSTM-Treg model
- (2) **While** training,
- (3) Train the network according to formula (18)
- (4) Get the forecast results $\{\hat{y}_n\}$ for the current round
- (5) Use $\{x_n, \hat{y}_n\}$ to train the decision tree
- (6) Calculate the average path length of the decision tree
- (7) Train MLP model to obtain $\tilde{\Omega}(W)$
- (8) Update the $\tilde{\Omega}(W)$ formula (18)
- (9) **If** training loss stop decreasing, **then**
- (10) Store the model and break
- (11) **End**
- (12) **End**

ALGORITHM 1: The description of the BiLSTM-Treg model algorithm. BiLSTM-Treg model algorithm.

TABLE 1: Correspondence between MIT-BIH arrhythmia database annotations and AAMI heartbeat types.

AAMI heartbeat category	MIT-BIH heartbeat types
N	Normal beat (N); left bundle branch block beat (L); right bundle branch block beat (R); nodal (junctional) escape beat (j); atrial escape beat (e)
S	Aberrated atrial premature beat (a); nodal (junctional) premature beat (J); atrial premature beat (A); premature or ectopic supraventricular beat (S)
V	Premature ventricular contraction (V); ventricular escape beat (E)
F	Fusion of ventricular and normal beat (F)
Q	Paced beat (/); unclassifiable beat (Q); fusion of paced and normal beat (f)

and FN. Take N-type as an example; formulas (19)–(22), respectively, represent the calculation methods of type N true positive heartbeat (TP_N), type N false-positive heartbeat

(FP_N), type N true negative heartbeat (TN_N), and type N false-negative heartbeat (FN_N). Table 3 shows the confusion matrix of the classification results.

TABLE 2: Experimental data statistics.

	Training set	Testing set	Total
N	81,551	9,044	90,595
S	2,501	280	2,781
V	6,519	716	7,235
F	723	79	802
Q	7,242	799	8041

$$TP_N = Nn, \quad (19)$$

$$FP_N = Sn + Vn + Fn + Qn, \quad (20)$$

$$TN_N = Ss + Sv + Sf + Sq + Vs + Vv + Vf + Vq + Fs + Fv + Ff + Fq + Qs + Qv + Qf + Qq, \quad (21)$$

$$FN_N = Ns + Nv + Nf + Nq. \quad (22)$$

In this paper, sensitivity, specificity, positive predictive value, and accuracy are used as indicators of classifier performance. Sensitivity (Se), also known as recall rate, is the proportion of positive samples that are correctly judged to be positive. The higher the sensitivity, the greater the proportion of samples correctly predicted. Specificity (Sp) is the proportion of correctly judged negative samples to actually negative samples. The positive predictive value (+p) refers to the proportion of correctly judged positive samples to all the judged positive samples. Accuracy (Acc) is the ratio of the sum of true positives and true negatives to the total number of samples, reflecting the consistency between test results and actual results. The calculation formula of the above four evaluation metrics is shown in (23–26).

$$Se = \frac{TP}{(TP + FN)}, \quad (23)$$

$$Sp = \frac{TN}{(TN + FP)}, \quad (24)$$

$$+p = \frac{TP}{(TP + FP)}, \quad (25)$$

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)}. \quad (26)$$

5. Results and Analysis

In order to build a time-series network model that is most suitable for the task of heartbeat classification and more accurately distinguish the categories of arrhythmias, we conducted the following five groups of experiments. In this section, we first compare and analyze the performance of RNN, GRU, and LSTM in heartbeat classification (Section 5.1). Secondly, the network is changed to bidirectional, and

the classification results of BiRNN, BiGRU, and BiLSTM are compared (Section 5.2). Thirdly, by comparing the effects of different heartbeat lengths on the classification performance of the BiLSTM model, the optimal heartbeat length was selected (Section 5.3). Then, tree regularization was used to optimize the BiLSTM model. By adding tree regularization, the generalization ability of BiLSTM is improved, and the classification accuracy is improved, compared with the traditional L1 and L2 regularization (Section 5.4). Then, the important features of the simulated decision tree are analyzed and verified by experiments (Section 5.5). Finally, the results are compared with other references (Section 5.6).

5.1. Analysis of Experimental Results of Different Time-Series Networks. In order to select the optimal time-series network model, Experiment 1 selected three network models, namely, RNN, GRU, and LSTM, for heartbeat classification. The experimental results show that the overall classification accuracy of the RNN model and GRU model is 98.98% and 98.97%, respectively. The overall classification accuracy of the LSTM model is 99.09%, which is better than that of the RNN model and GRU model. However, it cannot fully consider the rhythm information by using the one-way recurrent neural network for heartbeat classification. Table 4 shows the classification results and performance of three one-way recurrent neural networks.

5.2. Analysis of Experimental Results of Different Bidirectional Time-Series Networks. The one-way recurrent neural network can only learn the heartbeat information before the current moment when performing heartbeat classification. Therefore, we improve the selected LSTM network to BiLSTM so that the network can consider both the previous heartbeat information and the future heartbeat information. And the BiRNN and BiGRU networks are used for comparison and verification. The experimental results show that the overall classification accuracy of the BiRNN model and BiGRU model is 99.13% and 98.92%, respectively. The overall classification accuracy of the BiLSTM model is 99.18%, which is better than that of the BiRNN model and BiGRU model. Table 5 shows the classification results and performance of the three bidirectional recurrent neural networks.

5.3. Select the Optimal Length of Heartbeat Segment. In order to select the optimal length of the heartbeat segment, a total of 7 experiments were conducted. The length of heartbeat segments selected by us is 1, 5, 10, 15, 20, 25, and 30,

TABLE 3: Classification results of heartbeat statistics.

Reference labels	Predicted labels				
	n	s	v	f	q
N	Nn	Ns	Nv	Nf	Nq
S	Sn	Ss	Sv	Sf	Sq
V	Vn	Vs	Vv	Vf	Vq
F	Fn	Fs	Fv	Ff	Fq
Q	Qn	Qs	Qv	Qf	Qq

TABLE 4: Comparison of the classification results of RNN, GRU, and LSTM network models.

	RNN				GRU				LSTM			
	Se (%)	Sp (%)	+ p (%)	Acc (%)	Se (%)	Sp (%)	+ p (%)	Acc (%)	Se (%)	Sp (%)	+ p (%)	Acc (%)
N	99.61	97.10	99.35	99.15	99.45	97.80	99.51	99.15	99.69	97.40	99.42	99.27
S	90.40	99.80	92.86	99.54	92.72	99.77	92.11	99.58	92.05	99.90	96.19	99.68
V	97.66	99.80	97.26	99.66	97.94	99.76	96.61	99.63	96.98	99.77	96.84	99.59
F	71.26	99.94	91.18	99.72	71.26	99.91	86.11	99.68	75.86	99.91	86.84	99.72
Q	99.43	99.94	99.32	99.90	99.77	99.90	98.88	99.89	99.55	99.97	99.66	99.94

respectively, and the corresponding timestep of the BiLSTM is also 1, 5, 10, 15, 20, 25, and 30, respectively. The experimental results show that the classification accuracy of the network is gradually improved when the length of the heartbeat segment is less than 15. However, when the length of the heartbeat is greater than 15, the classification performance of the network decreases rapidly. The main reason is that the rhythm information of heartbeat, such as bigeminy, trigeminy, atrial tachycardia, and ventricular tachycardia, can be shown within 15 beats. When the heartbeat segment is too long, the heartbeat information considered by the network is too redundant, which will affect the network performance. Table 6 shows the classification results of the BiLSTM network with different lengths of heartbeat segments.

5.4. Analyze the Experimental Results of Different Regularization Methods. In order to improve the generalization ability of BiLSTM and further improve the classification accuracy, we choose tree regularization to constrain the weight of the network and use the traditional L1 and L2 regularization for comparison. Experimental results verify the feasibility and effectiveness of the proposed model, and the overall classification accuracy is 99.32%. The overall classification accuracy of the models using L1 regularization and L2 regularization was 99.26% and 99.23%. Compared with Experiment 2, the overall accuracy of Experiment 4 was improved by 0.14%, and the precision of class S, class V, and class F was all improved, among which the precision of class F was improved more obviously by 5.62%. Through the above analysis, it is concluded that tree regularization can effectively improve the classification accuracy of the network, which is better than the traditional L1 and L2 regularization. Table 7 shows the classification results of BiLSTM models under different regularization methods. Figure 7

shows the confusion matrix of heartbeat classification results based on the BiLSTM-Treg model.

5.5. Analyze the Key Nodes of the Simulated Decision Tree. The tree regularization method used in this paper looks for the decision tree representation of the model in the training process of the network. The generated decision tree simulates the decision process of the BiLSTM network model. We call this decision tree a simulated decision tree (SDT). Since there are many feature points in a single heartbeat, the generated SDT is too large, so we selected the tree generated by the top 10 important feature points of SDT when displaying this decision tree, as shown in Figure 8. The top 10 important feature points are 126, 112, 162, 121, 153, 80, 224, 93, 100, and 120. The positions of these feature points corresponding to the ECG waveform are as follows: sampling points 126, 120, 121, and 153 correspond to ST segment, sampling point 112 to J point, sampling point 224 corresponds to the endpoint of T wave, sampling point 162 corresponds to the beginning point of T wave, sampling point 80 corresponds to the peak value of Q wave, sampling point 93 corresponds to the peak value of R wave, and sampling point 100 corresponds to the peak value of S wave, as shown in Figure 9.

In Figure 8, we have modified the representation of the value field in the decision tree node. We represent the value in the value field as the percentage of the number of heartbeats of N, S, V, F, and Q in the total number of heartbeats of the corresponding category. Taking node 2 as an example, 0.08 in the value field represents that the number of class S heartbeats in this node accounts for 0.08% of the total number of class S heartbeats, which means that this node almost contains no class S heartbeats. Therefore, according to this simulated decision tree, we have the following analysis:

TABLE 5: Comparison of the classification results of BiRNN, BiGRU, and BiLSTM network models.

	BiRNN				BiGRU				BiLSTM			
	Se (%)	Sp (%)	+p (%)	Acc (%)	Se (%)	Sp (%)	+p (%)	Acc (%)	Se (%)	Sp (%)	+p (%)	Acc (%)
<i>N</i>	99.65	97.65	99.47	99.29	99.62	97.05	99.34	99.15	99.64	98.15	99.59	99.37
<i>S</i>	90.40	99.81	93.17	99.55	89.07	99.82	93.40	99.52	92.72	99.83	93.96	99.63
<i>V</i>	98.21	99.83	97.68	99.73	97.12	99.80	97.25	99.63	98.63	99.76	96.64	99.68
<i>F</i>	74.71	99.93	89.04	99.73	67.82	99.90	84.29	99.64	70.11	99.93	88.41	99.69
<i>Q</i>	100.00	99.97	99.66	99.97	99.77	99.91	98.99	99.90	100.00	99.98	99.77	99.98

TABLE 6: Effects of heartbeat segments of different lengths on the classification results of BiLSTM model.

Time step	1	5	10	15	20	25	30
Overall accuracy (%)	99.04	99.02	99.12	99.18	98.55	82.52	84.53

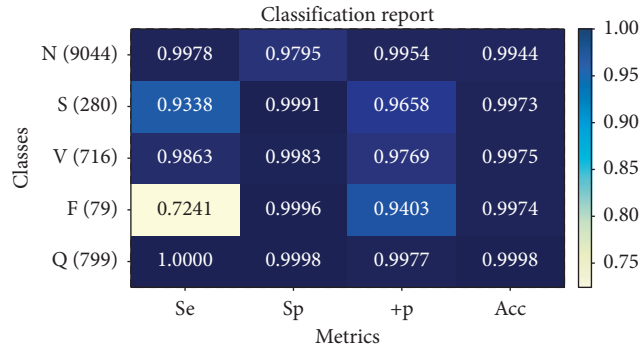


FIGURE 7: Confusion matrix of heartbeat classification results based on the BiLSTM-Treg model.

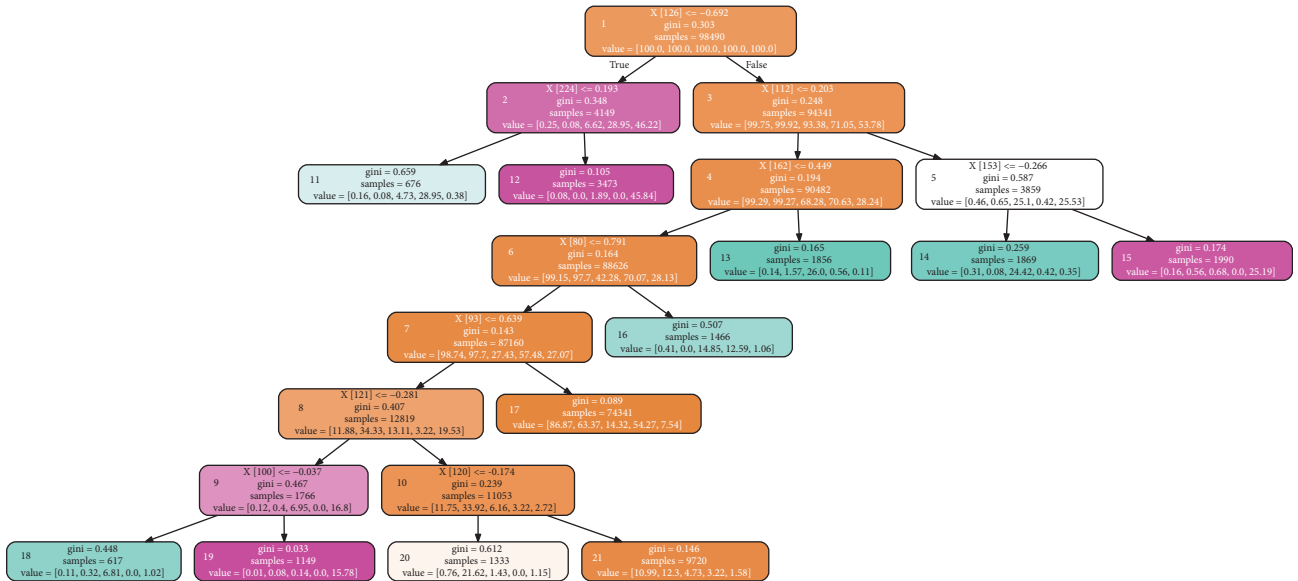


FIGURE 8: The simulated decision tree.

(1) Sampling point 126 is the root node of the simulated decision tree. According to whether the voltage value at this point is less than -0.0692 mV, the sample can be divided into two parts, namely, node 2 and node 3. In node 2, the heartbeat of classes F and Q is relatively large, while the heartbeat of the other three categories is relatively small. Therefore, 28.95% of

class F heartbeats and 46.22% of class Q heartbeats were separated from the total sample according to the sampling point 126. Analysis of the reason: sampling point 126 is the point of ST segment in the ECG waveform. ST segment refers to the segment between the end of the QRS complex and the beginning of the T wave, representing the period

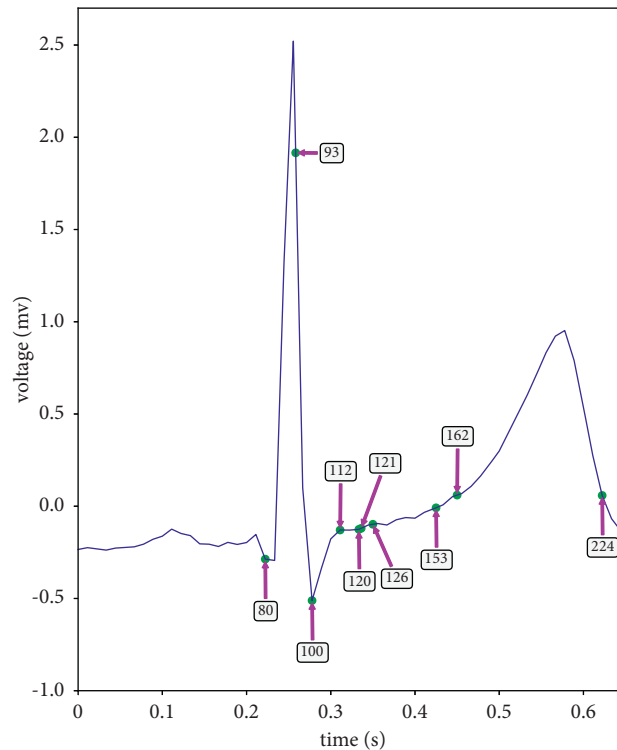


FIGURE 9: The key feature points of the decision tree correspond to the positions in the ECG waveform.

between ventricular depolarization and ventricular repolarization [39]. The normal ST segment is smooth and flush with the baseline.

- (2) It is shown by nodes 11 and 12 that node 2 distinguishes the F heartbeats from the Q heartbeats according to the value of sampling point 224. There is only 0.38% of class Q heartbeat in node 11 and 0% of class S heartbeat in node 12. Analysis of the reason: sampling point 224 is the endpoint of T wave in the ECG waveform. T wave is a wave with a larger amplitude and longer duration after the QRS complex, which shows the process of ventricular repolarization.
- (3) It can be shown from node 5 that node 3 separates 25.1% of class V heartbeats and 25.53% of class Q heartbeats from node 3 according to the value of sampling point 112, and it is shown by node 14 and node 15 that node 5 distinguishes the heartbeats of class S from class Q according to the value of sampling point 153. Analysis of the reason: sampling point 112 is the J point in the ECG waveform, and J point is the junction point between the end of the QRS complex and the beginning of the ST segment.
- (4) As indicated by node 13, node 4 separates 26.00% of class V heartbeats from node 4 according to the value of sampling point 162. As indicated by node 16, node 6 separates 14.85% of class V heartbeats and 12.59% of class F heartbeats from node 6 according to the value of sampling point 80. Sampling point 162 and sampling point 80 are T wave start points and Q wave peak values, respectively.

- (5) It is shown by node 8 that 11.88% of N, 34.33% of S, 13.11% of V, 3.22% of F, and 19.53% of Q heartbeats are separated from the sample of node 7 according to the value of sampling point 93. After passing through nodes 9 and 10, 6.81% of class V is sorted out at node 18, 15.78% of class Q is sorted out at node 19, and 21.62% of class S heartbeat is sorted out at node 20. The reason is analyzed as follows: sampling point 93 and sampling point 100 are, respectively, R wave and S wave in ECG waveform, which together with Q wave corresponding to sampling point 80 constitute QRS complex. The QRS complex is a group of wave groups with complex changes and large amplitude, showing the process of ventricular depolarization [40].

To verify that the BiLSTM-Treg algorithm focuses on and learns from these medically significant feature points, in Experiment 5, we only used these 10 important feature points as the features of a single heartbeat and used the BiLSTM-Treg algorithm to classify the heartbeat. The experimental results are shown in Table 8, and the overall classification accuracy is 98.45%. Compared with Experiment 4, Experiment 5 showed no significant decrease in all other metrics except the sensitivity of class S. The experimental results validate the importance of these medically significant feature points in the model.

5.6. Comparison with Previous Studies. Table 9 compares the classification performance of this method and other literature methods. The experimental data of other pieces of literature also comes from the MIT-BIH arrhythmia database. It can be seen from Table 9 that the method proposed in this paper has

TABLE 7: Classification results of BiLSTM models based on different regularization methods.

	BiLSTM + L1				BiLSTM + L2				BiLSTM + Treg			
	Se (%)	Sp (%)	+p (%)	Acc (%)	Se (%)	Sp (%)	+p (%)	Acc (%)	Se (%)	Sp (%)	+p (%)	Acc (%)
<i>N</i>	99.73	98.00	99.55	99.41	99.73	97.80	99.51	99.38	99.78	97.95	99.54	99.44
<i>S</i>	91.72	99.90	96.18	99.67	91.72	99.89	95.85	99.66	93.38	99.91	96.58	99.73
<i>V</i>	98.63	99.81	97.42	99.73	98.49	99.78	97.02	99.70	98.63	99.83	97.69	99.75
<i>F</i>	74.71	99.92	87.84	99.72	72.41	99.96	94.03	99.74	72.41	99.96	94.03	99.74
<i>Q</i>	100.00	99.98	99.77	99.98	100.00	99.98	99.77	99.98	100.00	99.98	99.77	99.98

TABLE 8: Classification results based on key feature points and BiLSTM-Treg algorithm.

	Se (%)	Sp (%)	+p (%)	Acc (%)
<i>N</i>	99.28	95.45	98.98	98.58
<i>S</i>	77.81	99.61	85.14	99.01
<i>V</i>	97.94	99.74	96.35	99.62
<i>F</i>	73.56	99.93	88.89	99.72
<i>Q</i>	99.89	99.97	99.66	99.96

TABLE 9: Comparison with other studies.

	Reference	Classifier	Performance (%)
Feature engineering	Yang et al., 2021 [9]	Random forest	Acc = 98.1 Se = 75.2 +p = 93.9 Acc = 99.01
	Ji et al., 2021 [15]	Stacking-DWKNN	Se _n = 99.65; Sp _n = 94.94; +P _n = 99.38 Se _s = 89.42; Sp _s = 99.85; +P _s = 94.90 Se _v = 97.21; Sp _v = 99.78; +P _v = 97.07 Se _f = 80.77; Sp _f = 99.94; +P _f = 88.73 Acc = 97.80
	Zhu et al., 2018 [7]	SVM	Se _n = 99.27; +P _n = 98.48 Se _s = 87.47; +P _s = 95.25 Se _v = 94.71; +P _v = 95.22 Se _f = 73.88; +P _f = 86.09 Acc = 94.03
	Pandey et al., 2017 [21]	9-layer CNN	Se _n = 91.54; Sp _n = 96.71; +P _n = 87.43 Se _s = 90.59; Sp _s = 98.63; +P _s = 94.30 Se _v = 94.22; Sp _v = 98.84; +P _v = 95.30 Se _f = 96.06; Sp _f = 98.67; +P _f = 94.76 Se _q = 97.75; Sp _q = 99.69; +P _q = 98.73 Acc = 99.21
Deep learning	Ji et al., 2019 [20]	1D-CNN	Se _n = 98.27; Sp _n = 99.39 Se _v = 97.54; Sp _v = 99.44 Se _f = 98.07 ; Sp _f = 99.50 Acc = 97.29
	Wu et al., 2020 [23]	CNN-BiLSTM	Se _n = 98.57; Sp _n = 93.62; +P _n = 98.81 Se _s = 84.97; Sp _s = 99.13; +P _s = 82.80 Se _v = 94.90; Sp _v = 99.35; +P _v = 94.00 Se _f = 76.89; Sp _f = 99.77; +P _f = 80.45 Acc = 98.58
	Pandey et al., 2021 [25]	BiLSTM	Se _n = 99.54; +P _n = 99.44 Se _s = 92.00; +P _s = 91.02 Se _v = 95.81; +P _v = 96.80 Se _f = 80.55; +P _f = 85.22 Acc = 99.32
	Proposed	BiLSTM-Treg	Se _n = 99.78 ; Sp _n = 97.95; +P _n = 99.54 Se _s = 93.38 ; Sp _s = 99.91 ; +P _s = 96.58 Se _v = 98.63 ; Sp _v = 99.83 ; +P _v = 97.69 Se _f = 72.41; Sp _f = 99.96 ; +P _f = 94.03 Se _q = 100.00 ; Sp _q = 99.98 ; +P _q = 99.77

Bold values represent the best experimental results which correspond to the evaluation criteria for one certain type.

the best classification accuracy, with an overall classification accuracy of 99.32%. The classification methods in literature [23, 25] all use the BiLSTM model. The results show that the proposed method has obvious advantages in all metrics except for the low sensitivity of class F, and the classification accuracy is 2.03% and 0.74% higher than the two methods, respectively. From the perspective of heartbeat type, the sensitivity of class S is significantly improved by the method presented in this paper compared with other methods. Compared with the literature [21], the method presented in this paper makes all metrics of Q heartbeat better, especially the sensitivity of Q heartbeat increased by 2.25%. In this paper, a classification method is proposed to integrate the rhythmic information between heartbeats that doctors are concerned about into the time-series network so that the network can learn this information effectively. Moreover, the bidirectional time-series network model can more conveniently obtain the context information of the heartbeat segment, so the algorithm in this paper can have better classification performance in the heartbeat classification problem.

6. Conclusion

In this paper, an intelligent classification of heartbeat based on the BiLSTM-Treg algorithm is proposed, which integrates rhythm information between heartbeats. This method fully considers the information of heart rhythm, which doctors pay attention to when diagnosing heart disease, and realizes the automatic classification of heartbeats. In this paper, the influence of different lengths of heartbeat segments on the classification results of the model is analyzed to select the best heartbeat segment length. On this basis, the BiLSTM-Treg algorithm was used for heartbeat classification. Experiments were carried out on the MIT-BIH arrhythmia database, and the results showed that the method can effectively distinguish five types of heartbeats, N, S, V, F, and Q, and the overall classification accuracy rate is 99.32%. The significance of this study is to provide patients with more accurate medical care services. The highlight of this study are as follows:

- (1) The heartbeat segment containing rhythm information between heartbeats was selected as the characteristics of the heartbeat sample, and the BiLSTM-Treg algorithm was used to automatically learn the potential rhythm information of individuals
- (2) A tree regularization method is proposed to optimize the BiLSTM-Treg algorithm and improve the accuracy of heartbeat classification
- (3) By analyzing the key nodes of the simulated decision tree, the interpretability of the BiLSTM-Treg algorithm is analyzed
- (4) The experimental results show that the algorithm proposed in this paper can effectively realize the classification of arrhythmia

In the future study, we will collect more class F-type heartbeat data for pretraining of the model so as to obtain more accurate intelligent ECG diagnosis results.

Data Availability

(1) All datasets used to support the findings of this study are included within the paper. (2) All datasets used to support the findings of this study were supplied by the publicly available MIT-BIH database from the Massachusetts Institute of Technology. The URL to access this data is <https://archive.physionet.org/cgi-bin/atm/ATM>. (3) The coding used to support the findings of this study has not been made available because the source code in this paper is part of a national project and is a trade secret, so the source code is not available.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant no. 2017YFB1401200); Key Science and Technology Project of Xinjiang Production and Construction Corps (Grant no. 2018AB017); Key Research, Development, and Dissemination Program of Henan Province (Science and Technology for the People) (Grant no. 182207310002); National Natural Science Foundation of China (Grant no. 61872324); Program for Science & Technology Innovation Talents in Universities of Henan Province (Grant no. 21HASTIT031); Training Plan for Young Backbone Teachers of Colleges and Universities in Henan (Grant no. 2019GGJS018); Key Scientific Research Project of Colleges and Universities in Henan Province (Grant no. 20A520035); and Collaborative Innovation Major Project of Zhengzhou (Grant no. 20XTZX06013).

References

- [1] F. A. Atienza, E. Morgado, L. F. Martinez, A. G. Alberola, and J. L. R. Alvarez, "Detection of life-threatening arrhythmias using feature selection and support vector machines," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 832–840, 2013.
- [2] G. Sannino and G. De Pietro, "A deep learning approach for ECG-based heartbeat classification for arrhythmia detection," *Future Generation Computer Systems*, vol. 86, pp. 446–455, 2018.
- [3] R. E. Mason, I. Likar, R. O. Biern, and R. S. Ross, "Multiple-lead exercise electrocardiography," *Circulation*, vol. 36, no. 4, pp. 517–525, 1967.
- [4] S. Saadatnejad, M. Oveisi, and M. Hashemi, "LSTM-based ECG classification for continuous monitoring on personal wearable devices," *IEEE journal of biomedical and health informatics*, vol. 24, no. 2, pp. 515–523, 2019.
- [5] J.-S. Wang, W.-C. Chiang, Y.-L. Hsu, and Y.-T. C. Yang, "ECG arrhythmia classification using a probabilistic neural network with a feature reduction method," *Neurocomputing*, vol. 116, pp. 38–45, 2013.
- [6] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.

- [7] W. Zhu, X. Chen, Y. Wang, and L. Wang, "Arrhythmia recognition and classification using ECG morphology and segment feature analysis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 1, pp. 131–138, 2018.
- [8] L. Zaorálek, J. Platoš, and V. Snášel, "Patient-adapted and inter-patient ECG classification using neural network and gradient boosting," *Neural Network World*, vol. 28, no. 3, pp. 241–254, 2018.
- [9] P. Yang, D. Wang, W.-B. Zhao, L.-H. Fu, J.-L. Du, and H. Su, "Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification," *Bio-medical Signal Processing and Control*, vol. 63, Article ID 102138, 2021.
- [10] Q. Qin, J. Li, L. Zhang, Y. Yue, and C. Liu, "Combining low-dimensional wavelet features and support vector machine for arrhythmia beat classification," *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.
- [11] Y. Kutlu and D. Kuntalp, "Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients," *Computer Methods and Programs in Biomedicine*, vol. 105, no. 3, pp. 257–267, 2012.
- [12] A. Ebrahimzadeh, M. Ahmadi, and M. Safarnejad, "Classification of ECG signals using hermite functions and MLP neural networks," *Journal of AI and Data Mining*, vol. 4, no. 1, pp. 55–65, 2016.
- [13] A. S. Barhatte, R. Ghongade, and A. S. Thakare, "QRS complex detection and arrhythmia classification using SVM," in *Proceedings of the 2015 Communication, Control and Intelligent Systems (CCIS)*, pp. 239–243, IEEE, Muthura, India, November 2015.
- [14] M. Ayar and S. Sabamoniri, "An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm," *Informatics in Medicine Unlocked*, vol. 13, pp. 167–175, 2018.
- [15] S. Ji, R. Li, S. Shen, B. Li, B. Zhou, and Z. Wang, "Heartbeat classification based on multifeature combination and Stacking-DWKNN algorithm," *Journal of Healthcare Engineering*, vol. 2021, no. 3, 14 pages, Article ID 8811837, 2021.
- [16] M. Wu, Y. Lu, W. Yang, and S. Y. Wong, "A study on arrhythmia via ECG signal classification using the convolutional neural network," *Frontiers in Computational Neuroscience*, vol. 14, no. 1, pp. 1–10, 2020.
- [17] B. Ganguly, A. Ghosal, A. Das, D. Das, D. Chatterjee, and D. Rakshit, "Automated detection and classification of arrhythmia from ECG signals using feature-induced long short-term memory network," *IEEE Sensors Letters*, vol. 4, no. 8, pp. 1–4, 2020.
- [18] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Computers in Biology and Medicine*, vol. 102, pp. 278–287, 2018.
- [19] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM-based auto-encoder model for ECG arrhythmias classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1232–1240, 2019.
- [20] Y. Ji, S. Zhang, and W. Xiao, "Electrocardiogram classification based on faster regions with convolutional neural network," *Sensors*, vol. 19, no. 11, p. 2558, 2019.
- [21] U. R. Acharya, S. L. Oh, Y. Hagiwara et al., "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017.
- [22] M. Ashfaq Khan and Y. Kim, "Cardiac arrhythmia disease classification using LSTM deep learning approach," *Computers, Materials & Continua*, vol. 67, no. 1, pp. 427–443, 2021.
- [23] J. Wu, F. Li, Z. Chen, X. Xiang, and Y. Pu, "Patient-specific ECG classification with integrated long short-term memory and convolutional neural networks," *IEICE - Transactions on Info and Systems*, vol. E103.D, no. 5, pp. 1153–1163, 2020.
- [24] R. Li, X. Zhang, H. Dai, B. Zhou, and Z. Wang, "Interpretability analysis of heartbeat classification based on heartbeat activity's global sequence features and BiLSTM-attention neural network," *IEEE Access*, vol. 7, pp. 109870–109883, 2019.
- [25] S. K. Pandey and R. R. Janghel, "Classification of electrocardiogram signal using an ensemble of deep learning models," *Data Technologies and Applications*, vol. 55, no. 3, pp. 446–460, 2021.
- [26] Ö. Yildirim, "A novel wavelet sequence based on deep bi-directional LSTM network model for ECG signal classification," *Computers in Biology and Medicine*, vol. 96, pp. 189–202, 2018.
- [27] C. Sawant and H. T. Patil, "Wavelet based ECG signal denoising," in *Proceedings of the 2014 First International Conference on Networks & Soft Computing (ICNSC2014)*, pp. 20–24, IEEE, Guntur, India, August 2014.
- [28] Z. A. A. Alyasseri, A. T. Khader, M. A. Al Betar, and M. A. Awadallah, "Hybridizing β -hill climbing with wavelet transform for denoising ECG signals," *Information Sciences*, vol. 429, pp. 229–246, 2017.
- [29] G. Yao, X. Mao, N. Li et al., "Interpretation of electrocardiogram heartbeat by CNN and GRU," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–10, 2021.
- [30] A. Verma, S. Cabrera, A. Mayorga, and H. Nazeran, "A robust algorithm for derivation of heart rate variability spectra from ECG and PPG signals," in *Proceedings of the 2013 29th Southern Biomedical Engineering Conference*, pp. 35–36, IEEE, Miami, FL, USA, May 2013.
- [31] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [32] W. Li and J. Li, "Local deep field for electrocardiogram beat classification," *IEEE Sensors Journal*, vol. 18, no. 4, pp. 1656–1664, 2017.
- [33] A. A. R. Bsoul, S. Y. Ji, K. Ward, and K. Najarian, "Detection of P, QRS, and T components of ECG using wavelet transformation," in *Proceedings of the 2009 ICME International Conference on Complex Medical Engineering*, pp. 1–6, IEEE, Tempe, Arizona, April 2009.
- [34] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber, "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies," *Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies*, Wiley-IEEE Press, Hoboken, NJ, USA, 2001.
- [35] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [36] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [37] M. Wu, M. Hughes, S. Parbhoo, M. Zazzi, V. Roth, and F. Doshi Velez, "Beyond sparsity: tree regularization of deep models for interpretability," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, New Orleans, LA, USA, February 2018.
- [38] Z. F. M. Apandi, R. Ikeura, and S. Hayakawa, "Arrhythmia detection using MIT-BIH dataset: a review," in *Proceedings of*

the 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA), pp. 1–5, IEEE, Kuching, Malaysia, August 2018.

- [39] S. Kaplan Berkaya, A. K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, “A survey on ECG analysis,” *Biomedical Signal Processing and Control*, vol. 43, pp. 216–235, 2018.
- [40] M. Elgendi, “Fast QRS detection with an optimized knowledge-based method: evaluation on 11 standard ECG databases,” *PloS one*, vol. 8, no. 9, Article ID e73557, 2013.

Research Article

The Ensemble Machine Learning-Based Classification of Motor Imagery Tasks in Brain-Computer Interface

Abdulhamit Subasi ^{1,2} and Saeed Mian Qaisar ^{2,3}

¹Institute of Biomedicine, Faculty of Medicine, University of Turku, Kiinanmyllykatu 10, Turku 20520, Finland

²College of Engineering, Effat University, Jeddah 22332, Saudi Arabia

³Communication and Signal Processing Lab, Energy and Technology Research Center, Effat University, Jeddah 22332, Saudi Arabia

Correspondence should be addressed to Saeed Mian Qaisar; sqaisar@effatuniversity.edu.sa

Received 20 August 2021; Revised 30 September 2021; Accepted 25 October 2021; Published 9 November 2021

Academic Editor: G R Sinha

Copyright © 2021 Abdulhamit Subasi and Saeed Mian Qaisar. 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 Brain-Computer Interface (BCI) permits persons with impairments to interact with the real world without using the neuromuscular pathways. BCIs are based on artificial intelligence piloted systems. They collect brain activity patterns linked to the mental process and transform them into commands for actuators. The potential application of BCI systems is in the rehabilitation centres. In this context, a novel method is devised for automated identification of the Motor Imagery (MI) tasks. The contribution is an effective hybridization of the Multiscale Principal Component Analysis (MSPCA), Wavelet Packet Decomposition (WPD), statistical features extraction from subbands, and ensemble learning-based classifiers for categorization of the MI tasks. The intended electroencephalogram (EEG) signals are segmented and denoised. The denoising is achieved with a Daubechies algorithm-based wavelet transform (WT) incorporated in the MSPCA. The WT with the 5th level of decomposition is used. Onward, the Wavelet Packet Decomposition (WPD), with the 4th level of decomposition, is used for subbands formation. The statistical features are selected from each subband, namely, mean absolute value, average power, standard deviation, skewness, and kurtosis. Also, ratios of absolute mean values of adjacent subbands are computed and concatenated with other extracted features. Finally, the ensemble machine learning approach is used for the classification of MI tasks. The usefulness is evaluated by using the BCI competition III, MI dataset IVa. Results revealed that the suggested ensemble learning approach yields the highest classification accuracies of 98.69% and 94.83%, respectively, for the cases of subject-dependent and subject-independent problems.

1. Introduction

A Brain-Computer Interface (BCI) allows individuals to use electroencephalogram (EEG) signals to operate external equipment such as virtual worlds, robots, or spelling machines. The fundamental objective of the BCI is to use brain signals to create the required commands to control peripherals. The most important application is to bypass injured areas of the body or stimulate partly paralyzed organs. BCI devices are viewed as the best solution to mitigate problems for persons with various neuromuscular impairments such as spinal cord damage, amyotrophic lateral sclerosis, cerebral palsy, and stroke [1].

BCI systems may be divided into two categories based on the EEG signals collection methods: noninvasive and invasive. Because of the ease of usage, much current research has focused on noninvasive BCIs. Event-related potentials, steady-state visual-evoked potentials, and slow cortical potentials are the three main noninvasive BCI approaches [2]. In noninvasive approach, different EEG signals can be utilized in BCI. Within the EEG alpha and beta frequency regions of the brain signals, BCI systems typically employ Motor Imagery approaches to produce event-related actions. This form of BCI is mostly utilized for cursor control on computer screens and wheelchair navigation or in virtual environments. Several Motor Imagery (MI) techniques are commonly used, including

tongue movement, left/right hand movement, foot movement, and mental counting [3]. The goal of BCI technology is to assist people with brain diseases including cerebral palsy, amyotrophic lateral sclerosis, and motor neuron disease. EEG is commonly used as a tool for the BCI system [4, 5]. Based on phenomena of event-related synchronization (ERS) and event-related desynchronization (ERD), scientists can interpret and identify MI-related brain signals. The translation of imagination to action involves ERS and ERD. Both ERD and ERS are presented by variations in the EEG signal's oscillatory behaviour and can be investigated by the time-frequency analysis to identify the MI tasks [6]. MI is characterized as a human brain's ability to resynthesize motor experiences with no obvious movement. Such mental images may both appear consciously and be created and controlled deliberately by a subject making MI, which is a flexible and usable method for examining processes of human cognition and motor activity. As various studies have shown, MI uses almost the same neural framework as motor execution, which enables motor activity to be altered by MI training. The MI-based BCI uses variations in the cortical sensorimotor rhythms (SMR), generally ERD related to the different sensorimotor events, including MI [7]. In addition, BCI may serve as a technical bridge for the management of Active and Assisted Living (AAL) systems in the sense of intelligent environments and smart homes. As with any other traditional AAL device interface, the consumer needs to view BCI-enabled control as simple and normal as possible in order to encourage BCI acceptance and effectiveness [8].

Computer-based automated MI signal detection is essential for providing continuous assistance to the intended patients. The preprocessing, feature extraction, dimension reduction, and classification are all parts of the EEG-based automated MI signal detection approaches [9, 10].

Feature extraction and dimension reduction are the most critical aspects of the classification system for EEG-based MI signals since they greatly affect classifier efficiency and computational complexity. If the features retrieved from EEG signals include irrelevant characteristics, the classifier's performance will suffer. The amount of features determines the classifier's processing cost. As a result, extracting the appropriate amount of relevant features from EEG-based MI signals is critical for achieving high classification performance and computational effectiveness for a classifier [9]. In this study, the dataset IVa from the BCI competition is utilized in the experiments [11]. AA, AL, AV, AW, and AY are codes of five healthy participants that contributed to this dataset. Two classes of MI activities, right hand and right foot movement, referred to as class 1 and class 2, respectively, are involved.

1.1. Contribution. The main objective of this work is to extract relevant features from the EEG signals and to design a classifier that can effectively recognize the intended MI tasks.

The major contributions are to propose a novel hybridization of the Multiscale Principal Component Analysis (MSPCA), Wavelet Packet Decomposition (WPD),

subbands statistical features selection, and ensemble learning technique for automated classification of the MI tasks. The functional steps are as follows:

- (i) The Multiscale Principal Component Analysis (MSPCA) is used for denoising.
- (ii) The Wavelet Packet Decomposition is used for producing the subbands.
- (iii) The six different statistical features are extracted from each subband. These are mean absolute value, power, standard deviation, skewness, kurtosis, and ratio of absolute mean values of adjacent subbands.
- (iv) The extracted features are passed to the proposed ensemble learning-based classifiers for automated identification of the MI tasks.

1.2. Organization. The remainder of the paper is organized as follows. Section 2 presents a literature review. In Section 3, materials and methods are introduced, Section 4 discusses the results, and the conclusion is presented in Section 5.

2. Literature Review

The loss of motor function is one of the most concerning effects of injury or disease to the nervous system. The BCI assistive technologies have allowed artificial prostheses, wheelchairs, and computers to be controlled by the electrical activity of the brain in this decade. The major challenges in the BCI systems are precision and processing effectiveness. The current systems have high computational complexity and need advanced and resourceful processing systems to attain a real-time response. Additionally, their classification performance and robustness need to be improved. In this context, several studies have been presented [12, 13].

Zarei et al. [9] used a combination of the Principal Component Analysis (PCA) and the cross-covariance (CCOV) method for features extraction from the EEG signals for the BCI application. The multilayer perceptron neural networks (MLP) and Least Square Support Vector Machine (LS-SVM) are used for classification. The performance of the system is tested by using the BCI competitions dataset IVa. Kayikcioglu and Aydemir [10] extracted features from the EEG signals by using two-dimensional features mining from the 2nd order polynomial coefficients. Then, the functions are categorized using the algorithm k-nearest neighbor (k-NN). They achieved considerable enhancement in speed and accuracy while evolving the performance for the dataset Ia from the 2003 BCI competition. Leamy et al. [12] conducted a comparative experimental research, from a machine learning perspective, for MI-related EEG features in stroke subjects. They try to explore if such features are generalizable to use trained machine learning parameters employing healthy subjects and stroke-affected patients. If BCI is trained with appropriate data, it gives relatively good results to stroke patients; then such a deployment model will make BCI far more realistic in a clinical setting for stroke recovery. On the other hand, if the stroke-affected EEG is significantly different

from healthy EEG or changes over time, it may need more sophisticated architecture from a machine learning perspective for the realistic implementation of BCI in such a setting.

Li et al. [13] proposed a new approach for MI pattern identification. It combines a common spatial pattern algorithm for frequency band selection and features selection, and the classification is carried out with the particle swarm optimized twin Support Vector Machine. They used datasets IIb of BCI competition IV to test the proposed system. For a classification task, Kevric and Subasi [14] employed MSPCA-based denoising of the EEG signals. Comparison among three features extraction techniques, namely, the Empirical Mode Decomposition (EMD), Discrete Wavelet Transform (DWT), and WPD, is conducted. The extracted features sets are classified by using the k-Nearest Neighbor (k-NN) algorithm. The system performance is tested by using the publicly available BCI competition III dataset IVa. Miao et al. [15] have suggested an EEG signals channel selection method. It uses the linear discriminant criteria for automated selection of channels with strong discriminative capabilities. Furthermore, the artificial bee colony algorithm is used for dimension reduction. The performance is tested by using the dataset IVa from the BCI competition III. In [16], Baali et al. have used a signal-dependent orthogonal transformation for features extraction. The classification is carried out by using a tree-based logistic model classifier. In [17], Chaudhary et al. used the flexible analytic wavelet transform (FAWT) for features extraction. The classification is carried out with ensemble learning-based subspace k-Nearest Neighbor (k-NN) classifier. In [18], Rahman et al. have used the Rényi min-entropy-based features extraction approach. The extracted features are used for classifying 4 different BCI categories by using the Random Forest (RF) algorithm. The performance of the proposed method is evaluated by using the BCI competition IV dataset.

Khare and Bajaj [19] employed the extreme learning machine-based classification of the MI tasks. The channels selection is realized by using the multicluster unsupervised learning approach. The signal decomposition is performed by using a flexible variational mode decomposition (F-VMD). Pertinent features from different modes are explored, namely, hjorth, entropy, and quartiles. In [20], the authors have used the flexible analytic wavelet transform (FAWT) for signal decomposition. Time-frequency attributes are calculated from subbands. The PCA, kernel PCA (KPCA), locally linear embedding (LLE), and Laplacian Eigenmaps (LE) are used for feature selection. The Linear Discriminant Analysis (LDA) algorithm is used for the classification. The performance is tested by using the BCI competition III dataset IIIb.

Tiwari et al. [21] proposed a Deep Neural Network (DNN) model for automated identification of the MI tasks by utilizing the EEG signals. The Power Spectral Densities (PSDs) are extracted as features from subbands by applying a bank of Butterworth filters. The performance is tested for the BCI competition III and V dataset MI tasks. Musallam et al. [22] utilized a Convolutional Neural Network (CNN) model that incorporates a number of different methods, including

temporal convolutional networks (TCNs), separable convolution, depthwise convolution, and layer fusion. The intended EEG signals are processed by two successive 1D convolution stages. The first in the time domain and subsequently channelwise and the second based on the image-like representation are used as an input of the main TCN. The performance is tested by using the BCI competition IV, IIa dataset.

3. Materials and Methods

The proposed system's framework is shown in Figure 1. A description and parameterization of different system modules are given in the following section.

3.1. Dataset. The suggested system performance is evaluated by using the well-known BCI competition III, dataset IVa¹ [11]. AA, AL, AV, AW, and AY are codes of five healthy participants that contributed to this dataset. They completed two classes of MI activities involving right hand and right foot movement, referred to as class 1 and class 2, respectively. Subjects are seated in comfortable chairs with armrests. The EEG signals are acquired from 118 electrodes, mounted by following the 10/20 globally accepted standard. Each considered subject performed 140 trials of each category. Being two considered classes of tasks, it resulted in a total of 280 trials per subject. Each trial is carried out for a duration of 3.5 sec. For each category, the data is made up of different-sized training and testing sets.

The training set for subjects AA, AL, AV, AW, and AY has 168, 224, 84, 56, and 28 trials, respectively. The testing set consists of 112, 56, 196, 224, and 252 trials for participants AA, AL, AV, AW, and AY, respectively.

The EEG signals are originally recorded at a rate of 1 kHz. These EEG signals are bandlimited to 50 Hz by using digital filtering and are onward downsampled to the rate of 100 Hz [11]. These downsampled versions of signals are used in this study. The EEG signals from only three channels (C3, Cz, and C4) are selected from a total of 118 available channels. This is because these channels contain the most discriminatory features on Motor Imagery activities involving the hands and feet. For each patient, 280 EEG segments of 3.5 seconds, with 3 selected channels, are prepared [11]. These are from two categories: right hand and foot. In total, 1400 EEG instances were used for the five mentioned subjects. They belong to the two considered classes of the MI tasks.

3.2. Denoising with Multiscale Principal Component Analysis (MSPCA). In multivariate statistical analysis, the PCA is one of the most important models. Let a measurement dataset with m sensors exist, such as $x \in R^m$. Each sensor in the measurement sample contains n separate sampling data, which are combined into a data matrix of size $m \times n$. The process is given by

$$X = [x_1, x_2, x_3, \dots, x_n]. \quad (1)$$

Each row of X represents a sample, and each column represents a measurement variable. The PCA model begins

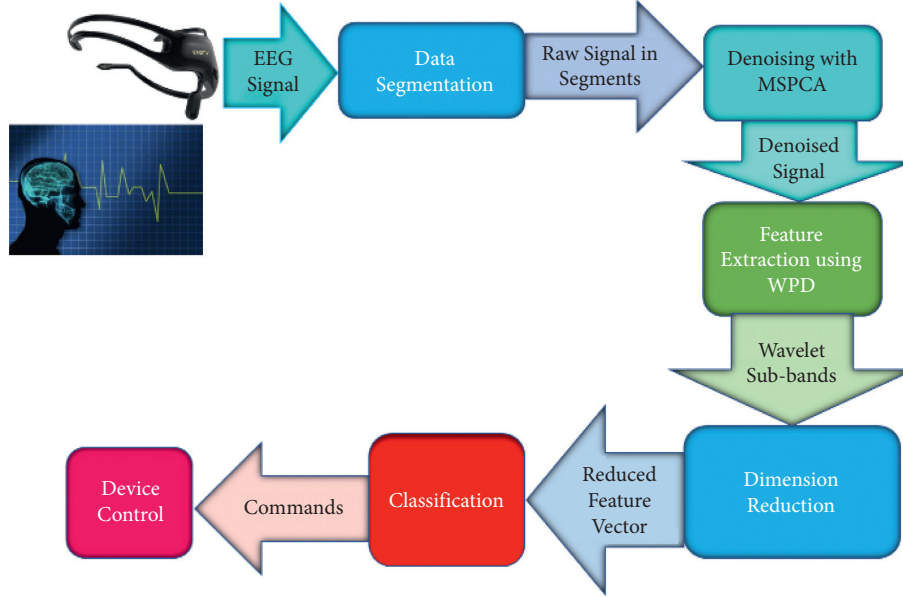


FIGURE 1: The system block diagram.

by standardizing each sample of X by computing the covariance matrix of X . The process is given by

$$\text{cov}(x) \approx \frac{X^T \cdot X}{n-1}. \quad (2)$$

The size of the feature values is ordered from large to small when the feature decomposition of X is done. The process of decomposing X in its principal components is given by equation (3), where $P \in R^{m \times A}$ contains first A feature vectors of $\text{cov}(x)$. $T \in R^{n \times A}$ is a matrix, where each column is known as the principal element variable. A is the count of principal components, and it is equal to the number of columns in T .

$$\begin{cases} X = \hat{X} + Er = T \cdot P^T + Er, \\ T = X \cdot P. \end{cases} \quad (3)$$

Equation (4) can be used to determine the principal component's covariance, where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the first A large eigenvalues of the covariance matrix of X .

$$\Lambda = \frac{X^T \cdot X}{n-1} = \begin{bmatrix} \lambda_1 & \dots & \dots \\ \dots & \lambda_2 & \dots \\ & & \cdot \\ & & \cdot \\ \dots & \dots & \dots & \lambda_n \end{bmatrix}. \quad (4)$$

In this paper, the wavelet transform is combined with the Principal Component Analysis (PCA) to create MSPCA for the incoming signal denoising purpose. The principle of wavelet transform is described in Section 3.3. In this study, the 5th level of decomposition is realized by using the

Daubechies wavelet analysis algorithm [23]. MATLAB is used for implementing the wavelet transform [24].

The ability of standard PCA is enhanced by incorporating the multiscale analysis. Collectively, it results in the multiscale PCA (MSPCA) [25]. In MSPCA, the PCA's capacity to extract covariance between variables is combined with orthonormal wavelets' ability to distinguish deterministic features from stochastic processes and approximately decorrelate the autocorrelation across observations. It identifies linearly related wavelet coefficients at multiple level subbands, obtained with wavelet transform. It allows representing each considered subband with fewer features while removing the autocorrelated coefficients. It results in a simplified representation of the considered subbands at each level of decomposition. The EEG waveforms are decomposed by using the Daubechies wavelet analysis algorithm with the 5th level of decomposition. In the next step, the PCA of detailed coefficients, obtained at each level, is utilized to select the principal components at each scale. Onward, the signal is reconstructed by using the wavelet synthesis. It diminishes the unwanted noise from the incoming signal and generates a simple and noise-free signal version [25, 26]. MATLAB is used for implementing the MSPCA [24].

3.3. Features Extraction with Wavelet Packet Decomposition (WPD). Wavelets are well-known functions and widely used for multiresolution time-frequency analysis. Wavelets can be mathematically described by equation (5) [23], where the dilation parameter is represented by s and the translation parameter is represented by u . The parameters can be generated at the same time with different frequencies.

$$\psi(t) = \frac{1}{\sqrt{S}} \psi\left(\frac{(t-u)}{s}\right). \quad (5)$$

The process of decomposing a signal $x(t)$, by using wavelet transform, can be given by

$$W_x(u, s) = \frac{1}{\sqrt{S}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{(t-u)}{s} \right) dt. \quad (6)$$

A discrete version of the wavelet transform (DWT) is used in this study. The selection of the right number of wavelet decomposition levels, m , is the first key step in the DWT decomposition. The incoming signal $x[n]$ passes concurrently through both the high-pass and low-pass filters, $h[k]$ and $l[k]$. For the m^{th} scale level, the output is represented by two subbands, namely, Detail (D_m) and Approximation (A_m). The process is clear from equations (7) and (8), where H is the order of filters used at different decomposition stages:

$$D_m[k] = \sum_{n=1}^H x[n] \cdot h[2.k - n], \quad (7)$$

$$A_m[k] = \sum_{n=1}^H x[n] \cdot l[2.k - n]. \quad (8)$$

The Wavelet Packet Decomposition is known as the extension of Discrete Wavelet Transform (DWT). The DWT mainly focuses on the low-frequency components, known as approximate coefficients. However, WPD utilizes both approximate and detailed coefficients, high-frequency components [27]. Consequently, when tactfully used, the WPD can result in signal decomposition with superior frequency resolution compared to the DWT [26]. In the studied case, the denoised signal is further analysed by using four levels of WPD. Pertinent statistical features are extracted from multiresolution subbands, obtained with the 4th level of WPD. MATLAB is used for implementing the WPD [24]. The principle of employed WPD with the 4th level of decomposition is shown in Figure 2, where D_m and A_m are, respectively, detailed and approximation coefficients at different decomposition stages and $m \in \{1, 2, 3, 4\}$.

3.4. Dimension Reduction. Since the dimension of the extracted features with WPD is high, the dimension should be reduced. Therefore, in this study, the dimension of extracted features is reduced by using statistical values of the WPD subbands. Using the statistical values of the subbands, the pertinent classifiable features are created from 16 subbands, shown in Figure 2. Five features are extracted from each subband, namely, mean absolute value, average power, standard deviation, skewness, and kurtosis. It results in $16 \times 5 = 80$ features. Additionally, the ratios of absolute mean values of the adjacent subbands are computed, resulting in 15 more features. In this way, in total, 95 features are extracted for each EEG instance, resulting in feature set dimension of 1400×95 for all considered instances.

3.5. Classification Methods. The prepared features set is categorized by using k-Nearest Neighbor (k-NN), C4.5 Decision Tree, REP Tree, Support Vector Machine (SVM),

Random Tree (RT), and RF, which are all well-known robust classification algorithms. Weka is used for evaluating the considered classifiers [28, 29]. To avoid any bias in findings due to the limited volume of the dataset, the 10-fold cross-validation (10-CV) approach is used along with multiple evaluation measures, namely, accuracy, F-measure, and the area under the ROC curve (AUC). Here, ROC stands for receiver operating characteristic [29].

3.5.1. Support Vector Machine (SVM). The SVM searches for hyperplane in an N-dimensional space in the classification of the data points. The SVM can be used for both classification and regression. The system functions by focusing on the decision line. It is a theoretically mature algorithm, only takes tens of instances for training, and is unaffected by the number of dimensions. Furthermore, effective approaches are developed to rapidly train this classifier [30]. In this study, the SVM is used with the cubic polynomial kernel and with a regularization parameter of 100.

3.5.2. K-Nearest Neighbor (k-NN). The k-NN refers to a supervised learning algorithm used in regression and classification problems. The algorithm functions by assuming that every data falling near each other belongs to the same class. It means that the algorithm considers that the classification of information is based on similarities. The technique is highly preferred because of its simplicity [30]. In this study, the k-NN with $k = 1$ is used. Here, k is the number of neighbors, used in the decision.

3.5.3. REP Tree. REP Tree creates a decision or regression tree using information variance reduction and then prunes it using reduced-error pruning. It optimizes speed by only sorting values for numeric attributes once. The minimum number of instances per leaf, maximum tree depth, minimum fraction of training set variance for a split, and the number of folds for pruning are adjustable parameters [31]. In this study, the REP Tree is used with its default configurations, available in Weka [28, 29].

3.5.4. C4.5 Decision Tree. The C4.5 can create classifiers that are redescribed as rulesets. C4.5 starts by growing an initial tree using the divide-and-conquer method. It labels the potential test instances by using two heuristic criteria. The first is the information gain, which tries to minimize the total entropy of subsets. The second is the default gain ratio, which tries to divide the information gain by the information supplied via the test outcomes [30]. In this study, the C4.5 is used with its default configurations, available in Weka [28, 29].

3.5.5. Random Tree (RT) Classifiers. The RT is a supervised learning algorithm that is easy to use and flexible. The algorithm produces excellent results despite lacking hyperparameter tuning. A combination of decision subtrees is trained based on the bagging method. The primary concept

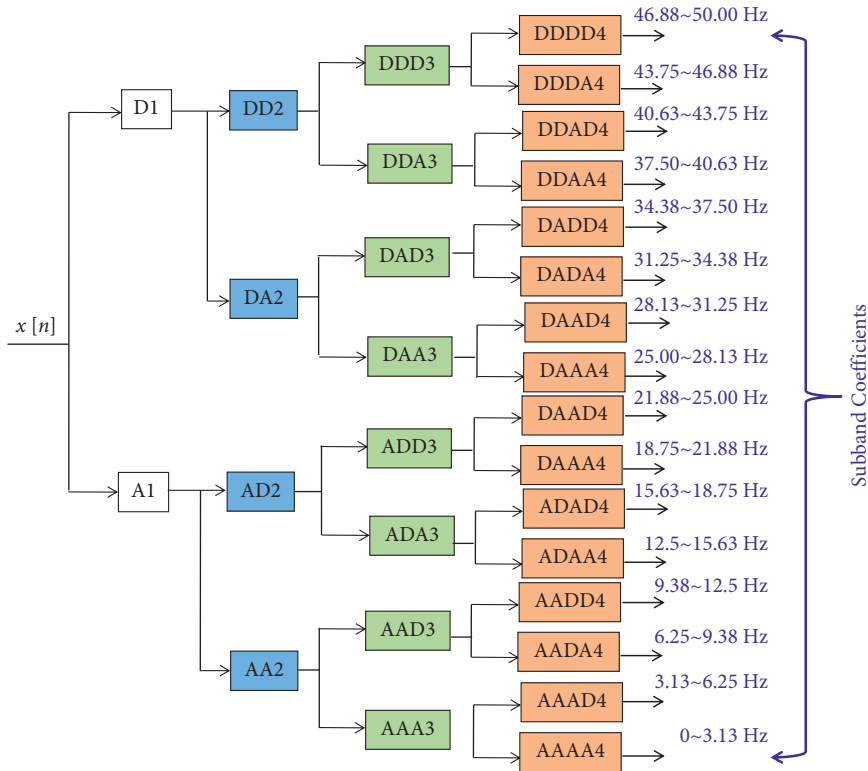


FIGURE 2: The employed WPD scheme.

of the functioning of the Random Tree is that combined learning models will increase the quality of results gained [31]. In this study, the RT is used with its default configurations, available in Weka [28, 29].

3.5.6. Random Forests (RF). The RF refers to a robust machine learning algorithm for various tasks such as classification and regression. The algorithm works by using bagging and randomness when creating each of the trees. It makes an uncorrelated forest of trees where their prediction is more accurate than a single tree [32]. In this study, the RF is used with 100 trees.

3.5.7. Rotation Forest (RoF). The RoF is a feature extraction-based classifier ensemble. We make the training data for a basic classifier by randomly partitioning the feature set into Q subgroups. PCA is applied to each subgroup, and Q is a parameter of the method. To retain the data's variability information, all basic components are kept. As a result, rotating the Q -axis produces additional attributes for a base classifier [33]. All primary components are kept in order to preserve the data's variability information. As a result, new features for a base classifier are formed by rotating the Q -axis [33]. The purpose of the rotation approach is to enhance individual accuracy while also providing variation within the group. Each base classifier's feature extraction contributes to diversity.

3.5.8. The Random Subspace Method (RSM). A well-known ensemble technique is the RSM [34]. The training data is also modified in the RSM. This change, however, is done in the

feature space. The B -dimensional random subspace of the original B -dimensional feature space is thus obtained. As a result, the updated training set has B -dimensional training objects in it. Then, in the final decision rule, classifiers can be built in random subspaces and combined using simple majority voting [35].

3.6. The Ensemble Learning Method. The ensemble learning methods can improve the performance of classification [26]. In this framework, the RoF and the RSM classifiers are employed with single classifiers. Multiple classifiers are used for the considered classification task. Findings of classifiers with various accuracies are combined via an ensemble-based approach [36]. The principle is depicted with the help of Figure 3.

For the case of RoF, by randomly splitting the features set into Q subgroups, we generate training data for a base classifier. After that, the PCA is applied to each subgroup. To maintain the data's variability information, all principal components are taken into consideration. This is how Q -axis rotations are realized to prepare new features for a base classifier. The rotation technique is designed to enhance individual accuracy while simultaneously fostering variation within the ensemble. Each base classifier's diversity is created by feature extraction. In this scenario, accuracy is measured by training each base classifier with the entire dataset [33].

For the case of RSM, the B -dimensional random subspace of the original features set was produced. As a result, the training set comprises B -dimensional training objects. In this approach, we built classifiers in random subspaces and used simple majority voting to aggregate their results [35].

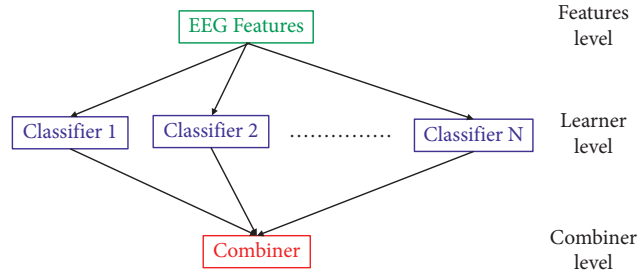


FIGURE 3: The general framework of ensemble classifiers.

3.7. Performance Evaluation Measures. In order to avoid any bias in the classification performance evaluation, multiple evaluation measures, namely, accuracy, F-measure, and AUC, are utilized [29]. The accuracy is defined by equation (9). True positives, true negatives, false positives, and false negatives are represented as tp , tn , fp , and fn , respectively. The F-measure is given by equation (10). The AUC presents the classification performance graphically. It is the area under the curve of the graph, obtained by tracing the True Positive Rate (TPR) with respect to the False Positive Rate (FPR). The TPR and FPR are, respectively, given by equations (11) and (12).

$$\text{accuracy (ACC)} = \frac{tp + tn}{tp + tn + fp + fn} \times 100, \quad (9)$$

$$F - \text{measure} = \frac{tp}{tp + 1/2(fp + fn)}, \quad (10)$$

$$\text{TPR} = \frac{tp}{(tp + fn)}, \quad (11)$$

$$\text{FPR} = \frac{fp}{(fp + tn)}. \quad (12)$$

4. Results

The system performance is tested by using the BCI competition III, dataset IVa [11]. An example of the input EEG signal and its denoised version, obtained with the MSPCA, is shown in Figure 4.

The denoised signal is onward decomposed in 16 subbands by using the 4th level of WPD. An example of obtained subbands is shown in Figure 5.

The overall system performance is studied in terms of classification precision. Findings are outlined in Table 1. These results are also presented graphically. In Figure 6, the accuracy scores, obtained with different classifiers, are shown. Figures 7 and 8, respectively, show the F-Measure and AUC values, obtained with different classifiers.

It is evident from Table 1 that the ensemble of k-NN and RoF attains the superior classification performance in most of the cases, compared to the other studied classifiers. The obtained percentages accuracies obtained for subjects AA, AL, AV, AW, and AY are, respectively, 96.67%, 94.05%, 89.64%, 96.43%, and 90.71%. However, the results are

different for the case of subject AY. For AY, the highest classification accuracy of 98.69% is attained RSM with RF and 98.45% is attained by the RSM with C4.5. The RoF with C4.5 is the third with an accuracy of 97.14%. RSM with RT is the fourth one with an accuracy of 97.02%. RSM with k-NN is the eighth with an accuracy of 92.38%.

While considering the case of each subject, the highest accuracy of 98.69% is achieved by the RSM with RF. However, for all five subjects, the highest classification accuracy of 94.83% is achieved by the RoF with k-NN. It shows that, in general, the used assembly of MSPCA, WPD, and statistical feature selection using RoF with k-NN results in the best classification performance for the studied dataset.

5. Discussion

The results, outlined in the above section, show that, for most of the cases, the proposed framework of MSPCA, WPD, statistical features selection, and RoF with k-NN leads towards a high classification accuracy. However, the best results obtained for the subject AY are obtained for a combination of MSPCA, WPD, and statistical feature selection using RSM with RF. It happens due to the variation in EEG signals magnitudes and response time of subjects while executing an MI task. It has an impact on the shape of EEG signals as well as the performance of the postsegmentation, denoising, feature extraction, and classification algorithms. Therefore, various subjects have varying classification accuracy as a result of this.

The BCI is a well-explored domain, and making a performance comparison with state of the art is a tedious task. It is mainly because of the variety of datasets, pre-processing, features extraction, dimension reduction, and classification techniques used in the previous studies. However, a performance comparison of the suggested framework is made with state-of-the-art solutions using similar datasets. Table 2 provides a review of those studies. It indicates that the suggested method ensures a comparable or superior performance as compared to the previously presented methods. It indicates that the devised denoising, dimension reduction, and ensemble classification approaches have a substantial influence on the overall precision and performance of the system. The self-configurability of ensemble classifiers, as a function of the utilized training dataset, is one of their main advantages. The use of event-driven tools can help in enhancing the computational effectiveness of the suggested method [45–48]. In the future,

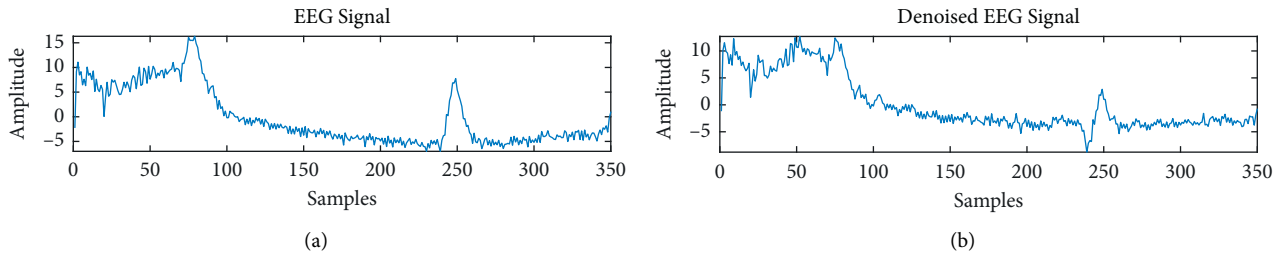


FIGURE 4: (a) The EEG signal and (b) denoised version of EEG signal.

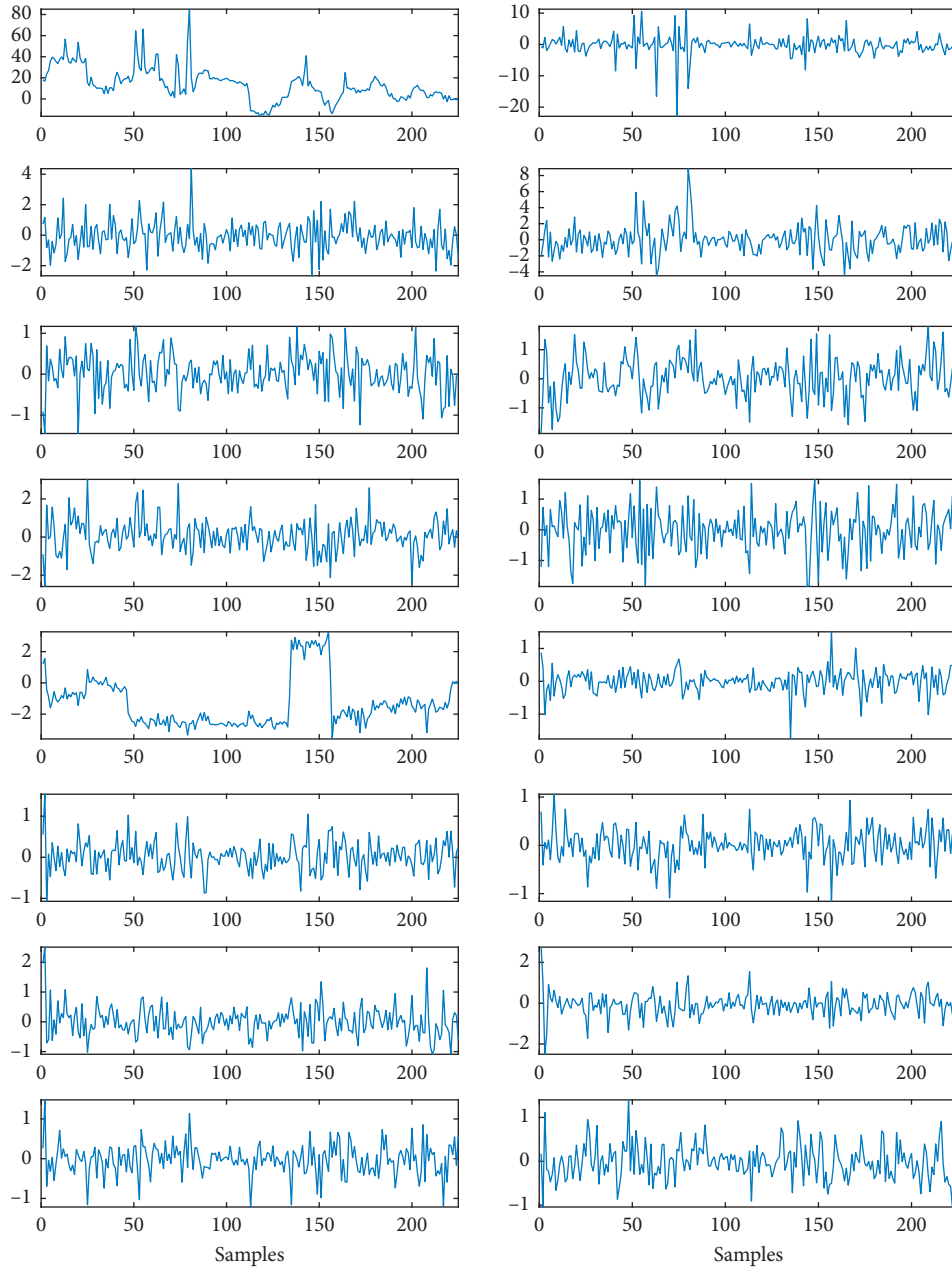


FIGURE 5: The 16 subbands, obtained with WPD.

TABLE 1: Summary of the classification performance measures.

Subj.	Classifier	Accuracy		F-measure		AUC	
		RSM	RoF	RSM	RoF	RSM	RoF
AA	SVM	89.64	90.60	0.896	0.906	0.958	0.929
	k-NN	93.45	96.67	0.935	0.967	0.988	0.985
	RF	82.02	84.76	0.820	0.848	0.900	0.928
	C4.5	67.50	74.76	0.675	0.748	0.764	0.834
	REP tree	69.88	67.50	0.699	0.675	0.765	0.746
	RT	70.83	74.29	0.706	0.742	0.799	0.827
AL	SVM	86.43	84.88	0.864	0.849	0.923	0.879
	k-NN	92.26	94.05	0.923	0.940	0.970	0.968
	RF	75.00	77.26	0.750	0.773	0.830	0.852
	C4.5	71.07	73.69	0.711	0.737	0.786	0.812
	REP tree	68.93	68.69	0.689	0.687	0.751	0.756
	RT	70.95	70.60	0.708	0.703	0.778	0.786
AV	SVM	87.02	84.40	0.870	0.844	0.928	0.886
	k-NN	88.93	89.64	0.889	0.896	0.955	0.939
	RF	73.69	76.90	0.737	0.769	0.818	0.857
	C4.5	61.90	71.31	0.619	0.713	0.677	0.767
	REP tree	62.38	66.19	0.624	0.662	0.671	0.721
	RT	65.00	65.95	0.644	0.654	0.691	0.738
AW	SVM	86.90	85.00	0.869	0.850	0.926	0.876
	k-NN	94.64	96.43	0.946	0.964	0.983	0.976
	RF	77.02	78.81	0.770	0.788	0.844	0.867
	C4.5	67.62	71.19	0.676	0.712	0.727	0.788
	REP tree	63.45	68.33	0.634	0.683	0.683	0.742
	RT	68.69	70.24	0.685	0.700	0.735	0.780
AY	SVM	92.02	91.90	0.920	0.919	0.984	0.942
	k-NN	92.38	90.71	0.924	0.907	0.977	0.952
	RF	98.69	94.29	0.987	0.943	0.999	0.991
	C4.5	98.45	97.14	0.985	0.971	0.999	0.995
	REP tree	95.60	93.10	0.956	0.931	0.988	0.980
	RT	97.02	93.10	0.970	0.931	0.996	0.985
ALL	SVM	89.98	90.12	0.900	0.901	0.962	0.932
	k-NN	93.55	94.83	0.935	0.948	0.984	0.974
	RF	80.36	81.83	0.804	0.818	0.889	0.901
	C4.5	68.57	73.00	0.686	0.730	0.769	0.815
	REP tree	68.60	69.02	0.686	0.690	0.765	0.760
	RT	72.57	72.40	0.724	0.722	0.807	0.809

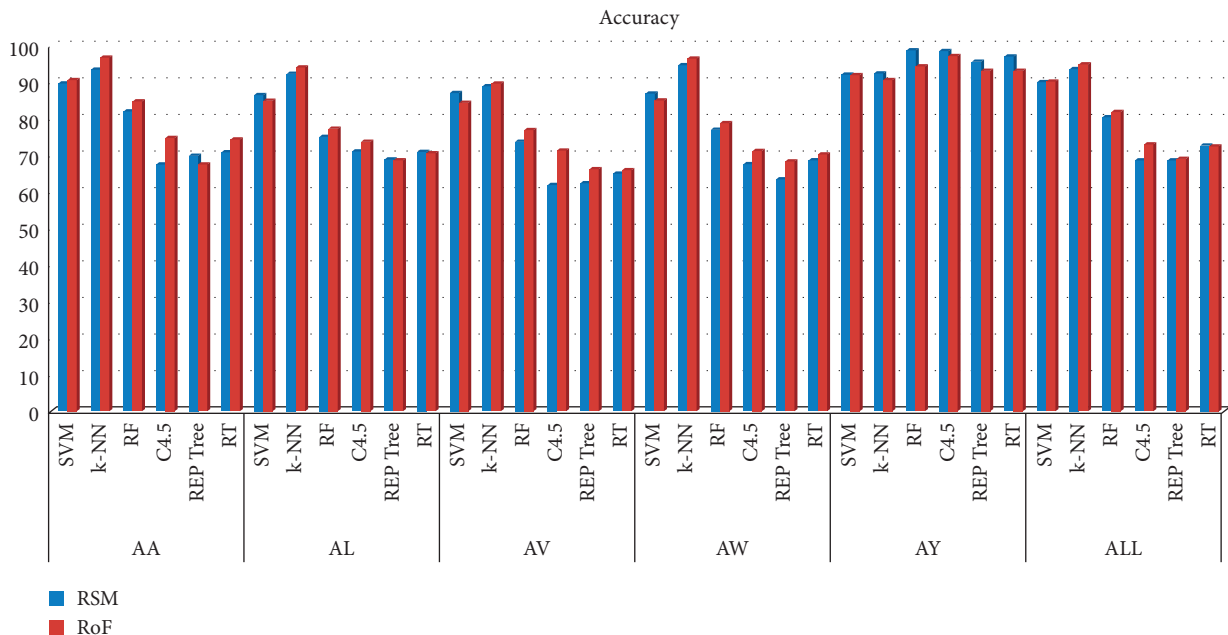


FIGURE 6: Accuracy of different classifiers.

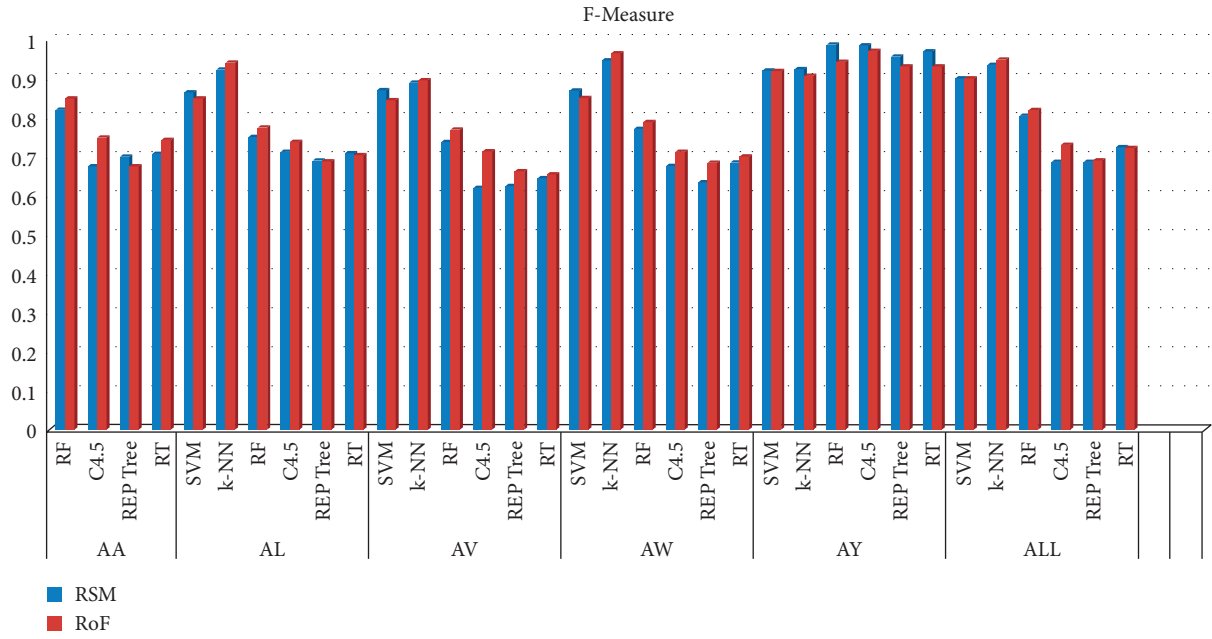


FIGURE 7: F-measure of different classifiers.

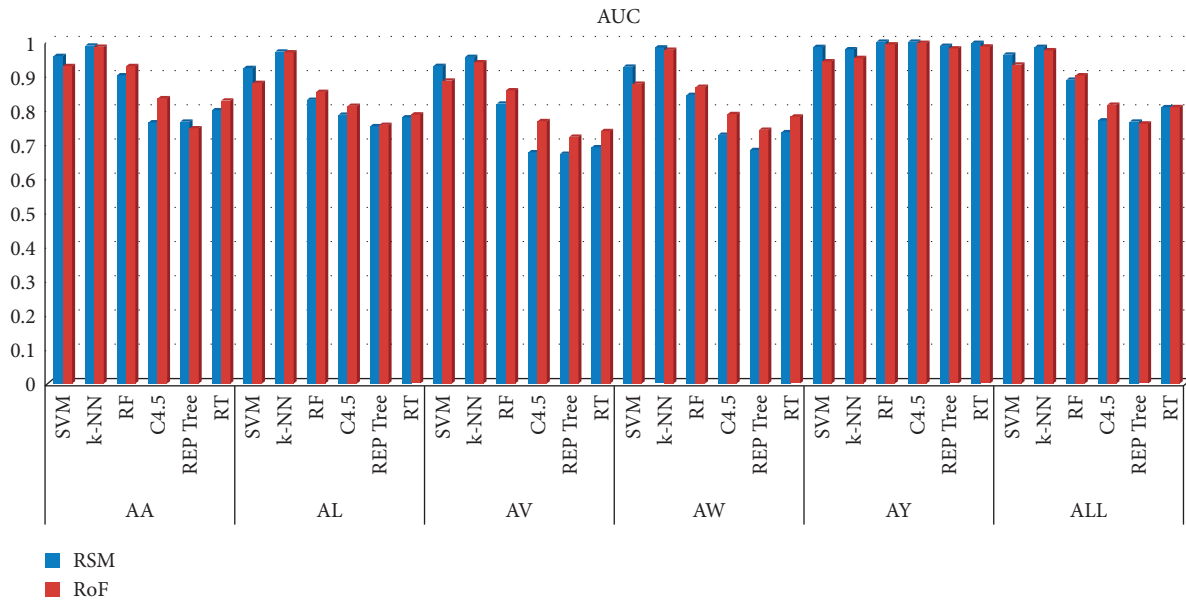


FIGURE 8: AUC of different classifiers.

TABLE 2: Comparison with previous studies.

Study	Feature extraction	Classifier	Classes/subject(s)	Accuracy (%)
[18]	Rényi min-entropy	RF	4/subject independent	80.55
[21]	Subbands PSDs	DNN	2/subject independent	82.48
[37]	Tangent space mapping	SVM	2/1-subject	97.80
[38]	Common spatial pattern	Backpropagation Neural network	2/subject independent	80.73
[39]	Regularized common spatial pattern	SVM	2/subject independent	91.9
[40]	Fisher ratio of time domain parameters	SVM	2/subject independent	89.13
[41]	Common spatial pattern	SVM	2/subject independent	85.01
[42]	Stacked autoencoders (SAE)	CNN	2/subject independent	82.00
[43]	Inverse problem through beamforming	CNN	2/subject independent	90.50
[44]	Granger causality channel selection and common spatial pattern	Linear SVM	2/subject independent	88.46
Proposed	WPD	RF and RSM	2/subject dependent	98.69
	WPD	k-NN and RoF	2/subject independent	94.83

this aspect can be investigated. The developed system has the potential to be integrated into the future generation of Brain-Computer Interface systems. The solution performed well for the intended dataset. Future work is to test its applicability for other potential Motor Imagery datasets. The incorporation of deep learning tools is another axis to explore.

6. Conclusion

In this paper, a novel automated Motor Imagery tasks classification method is proposed. The EEG signals are processed to distinguish between two categories of the brain activities. This approach is an intelligent combination of ensemble learning, Wavelet Packet Decomposition, Multi-scale Principal Component Analysis, and subbands statistical features extraction. Results have shown its effectiveness in classifying the intended Motor Imagery tasks. Using an intelligent ensemble of the Random Subspace classifier with Random Forest, the highest subject-dependent accuracy of 98.69% is realized. The suggested ensemble of the Rotation Forest classifier with k-NN achieved the highest subject-independent accuracy of 94.83%.

Data Availability

The dataset used in this paper is publicly available via the following link: http://www.bbci.de/competition/iii/desc_IVa.html.

Conflicts of Interest

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

Acknowledgments

This project was funded by Effat University with Decision no. UC#7/28 Feb. 2018/10.2-44i.





References

- [1] S. Sanei, *Adaptive Processing of Brain Signals*, John Wiley & Sons, Hoboken, NJ, USA, 2013.
- [2] W. Gao, J.-a. Guan, J. Gao, and D. Zhou, "Multi-ganglion ANN based feature learning with application to P300-BCI signal classification," *Biomedical Signal Processing and Control*, vol. 18, pp. 127–137, 2015.
- [3] C. Guger, S. Daban, E. Sellers et al., "How many people are able to control a P300-based brain-computer interface (BCI)?" *Neuroscience Letters*, vol. 462, no. 1, pp. 94–98, 2009.
- [4] B. Z. Allison, E. W. Wolpaw, and J. R. Wolpaw, "Brain-computer interface systems: progress and prospects," *Expert Review of Medical Devices*, vol. 4, no. 4, pp. 463–474, 2007.
- [5] J. R. Wolpaw, N. Birbaumer, D. J. Mcfarland, G. Pfurtscheller, and T. M. Vaughan, "Brain – computer interfaces for communication and control," vol. 113, pp. 767–791, 2002.
- [6] C. U. Neuper and G. Pfurtscheller, "Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates," *International Journal of Psychophysiology*, vol. 43, no. 1, 2001.
- [7] A. Vasilyev, S. Liburkina, L. Yakovlev, O. Perepelkina, and A. Kaplan, "Assessing motor imagery in brain-computer interface training: psychological and neurophysiological correlates," *Neuropsychologia*, vol. 97, pp. 56–65, 2017.
- [8] N. Mora, I. De Munari, and P. Ciampolini, *A Multi-Modal BCI System for Active and Assisted Living*, Springer, New York, NY, USA, 2016.
- [9] R. Zarei, J. He, S. Siuly, and Y. Zhang, "A PCA aided cross-covariance scheme for discriminative feature extraction from EEG signals," *Computer Methods and Programs in Biomedicine*, vol. 146, pp. 47–57, 2017.
- [10] T. Kayikcioglu and O. Aydemir, "A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1207–1215, 2010.
- [11] B. Blankertz, K.-R. Müller, D. J. Krusienski et al., "The BCI competition III: validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.
- [12] D. J. Leamy, J. Kocijin, K. Domijan et al., "An exploration of EEG features during recovery following stroke—implications for BCI-mediated neurorehabilitation," *J. Neuroengineering Rehabil.*, vol. 11, no. 1, p. 9, 2014.
- [13] D. Li, H. Zhang, M. S. Khan, and F. Mi, "A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition," *Biomedical Signal Processing and Control*, vol. 41, pp. 222–232, 2018.
- [14] 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.
- [15] M. Miao, A. Wang, and F. Liu, "Application of artificial bee colony algorithm in feature optimization for motor imagery EEG classification," *Neural Computing & Applications*, vol. 30, pp. 1–15, 2017.
- [16] H. Baali, A. Khorshidtalab, M. Mesbah, and M. J. Salami, "A transform-based feature extraction approach for motor imagery tasks classification," *IEEE J. Transl. Eng. Health Med.*, vol. 3, pp. 1–8, 2015.
- [17] S. Chaudhary, S. Taran, V. Bajaj, and S. Siuly, "A flexible analytic wavelet transform based approach for motor-imagery tasks classification in BCI applications," *Computer Methods and Programs in Biomedicine*, vol. 187, Article ID 105325, 2020.
- [18] M. A. Rahman, F. Khanam, M. Ahmad, and M. S. Uddin, "Multiclass EEG signal classification utilizing Rényi min-entropy-based feature selection from wavelet packet transformation," *Brain Inform.*, vol. 7, no. 1, pp. 1–11, 2020.
- [19] S. K. Khare and V. Bajaj, "A facile and flexible motor imagery classification using electroencephalogram signals," *Computer Methods and Programs in Biomedicine*, vol. 197, Article ID 105722, 2020.
- [20] Y. You, W. Chen, and T. Zhang, "Motor imagery EEG classification based on flexible analytic wavelet transform," *Biomedical Signal Processing and Control*, vol. 62, Article ID 102069, 2020.
- [21] S. Tiwari, S. Goel, and A. Bhardwaj, "MIDNN—a classification approach for the EEG based motor imagery tasks using deep neural network," *Applied Intelligence*, pp. 1–20, 2021.
- [22] Y. K. Musallam, N. I. Alfassa, G. Muhammad et al., "Electroencephalography-based motor imagery classification using temporal convolutional network fusion," *Biomedical Signal Processing and Control*, vol. 69, Article ID 102826, 2021.

- [23] I. Daubechies, *Different Perspectives on Wavelets*, American Mathematical Soc., Providence, Rhode Island, 2016.
- [24] W. Y. Yang, W. Cao, T.-S. Chung, and J. Morris, *Applied Numerical Methods Using MATLAB*, John Wiley & Sons, Hoboken, NJ, USA, 2020.
- [25] M. T. Sadiq, X. Yu, Z. Yuan, F. Zeming, A. U. Rehman, and I. Ulla, "Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain-computer interfaces," *IEEE Access*, vol. 7, pp. 171431–171451, 2019.
- [26] A. Subasi and S. M. Qaisar, "Surface EMG signal classification using TQWT, Bagging and Boosting for hand movement recognition," *Journal Ambient Intelligence Humanized Computing*, 2020.
- [27] E. Alickovic, J. Kevric, and A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction," *Biomedical Signal Processing and Control*, vol. 39, pp. 94–102, 2018.
- [28] R. R. Bouckaert, E. Frank, M. Hall et al., *WEKA Manual for Version 3-9-1*, Univ. Waikato Hamilt. N. Z., Hamilton, New Zealand, 2016.
- [29] A. Subasi, *Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques: A MATLAB Based Approach*, Academic Press, Cambridge, MA, USA, 2019.
- [30] X. Wu, V. Kumar, J. R. Quinlan et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2008.
- [31] M. Hall, I. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Kaufmann Burlingt, Burlington, MA, USA, 2011.
- [32] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, Amsterdam, Netherlands, 2011.
- [33] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: a new classifier ensemble method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.
- [34] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [35] M. Skurichina and R. P. Duin, "Bagging, boosting and the random subspace method for linear classifiers," *Pattern Analysis & Applications*, vol. 5, no. 2, pp. 121–135, 2002.
- [36] J. Y. Choi, D. H. Kim, K. N. Plataniotis, and Y. M. Ro, "Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography," *Expert Systems with Applications*, vol. 46, pp. 106–121, 2016.
- [37] M. Sartori, S. Fiori, and T. Tanaka, "An experimental study to compare CSP and TSM techniques to extract features during motor imagery tasks," *Signal Process. Mach. Learn. Brain-Mach. Interfaces*, 2018.
- [38] H. Jia, S. Wang, D. Zheng, X. Qu, and S. Fan, "Comparative study of motor imagery classification based on BP-NN and SVM," *Journal of Engineering*, no. 23, pp. 8646–8649, 2019.
- [39] 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.
- [40] Y. Park and W. Chung, "Selective feature generation method based on time domain parameters and correlation coefficients for Filter-Bank-CSP BCI systems," *Sensors*, vol. 19, no. 17, p. 3769, 2019.
- [41] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, "A CSP\AM-BA-SVM approach for motor imagery BCI system," *IEEE Access*, vol. 6, pp. 49192–49208, 2018.
- [42] R. Arabshahi and M. Rouhani, "A Convolutional neural network and stacked autoencoders approach for motor imagery based brain-computer linterface," in *Proceedings of the 2020 International Conference on Computer of Knowledge Engineering (ICCKE)*, pp. 295–300, Mashhad, Iran, October 2020.
- [43] C. Ieracitano, N. Mammone, A. Hussain, and F. C. Morabito, "A novel explainable machine learning approach for EEG-based brain-computer interface systems," *Neural Computing & Applications*, pp. 1–14, 2021.
- [44] H. Varsehi and S. M. P. Firoozabadi, "An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality," *Neural Networks*, vol. 133, pp. 193–206, 2021.
- [45] S. M. Qaisar and A. Subasi, "Cloud-based ECG monitoring using event-driven ECG acquisition and machine learning techniques," *Phys. Eng. Sci. Med*, vol. 43, no. 2, pp. 623–634, 2020.
- [46] S. M. Qaisar and A. Subasi, "Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare," *J. Ambient Intell. Humaniz. Comput*, no. 1–13, 2020.
- [47] S. Mian Qaisar, "Event-driven coulomb counting for effective online approximation of Li-ion battery state of charge," *Energies*, vol. 13, no. 21, p. 5600, 2020.
- [48] S. Mian Qaisar, "Signal-piloted processing and machine learning based efficient power quality disturbances recognition," *PloS One*, vol. 16, no. 5, Article ID e0252104, 2021.

Research Article

Interrater and Intrarater Reliability of Electrical Impedance Myography: A Comparison between Large and Small Handheld Electrode Arrays

Huijing Hu ¹, Wai Leung Ambrose Lo ², Xiaoyun Wang,³ Le Li ¹, and Ping Zhou ⁴

¹Institute of Medical Research, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China

²Department of Rehabilitation Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, Guangdong, China

³Guangdong Work Injury Rehabilitation Center, Guangzhou, Guangdong 510440, China

⁴University of Health and Rehabilitation Sciences, Qingdao, Shandong 266024, China

Correspondence should be addressed to Le Li; lile5@nwpu.edu.cn and Ping Zhou; dr.ping.zhou@outlook.com

Received 13 May 2021; Revised 28 July 2021; Accepted 19 October 2021; Published 2 November 2021

Academic Editor: GR Sinha

Copyright © 2021 Huijing Hu 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 objective of this study was to evaluate the interrater and intrarater reliability of electrical impedance myography (EIM) using handheld sensors of different sizes. Electrical impedance myography of the biceps brachii muscle of twenty healthy individuals was performed by two raters using both large and small sensors. The procedures were also repeated 5 to 8 days after the first recording session. The repeatability of the resistance, reactance, and phase angle at two different current frequencies (50 and 100 kHz) was assessed by the intraclass correlation coefficient (ICC). The ICCs of the large sensor were higher than those of the small sensor for both the intrarater and interrater reliabilities. High-frequency current tended to improve the ICC for the small sensor. These results indicate reasonable repeatability of the handheld electrode arrays for EIM measurements. The findings suggest that electrode array should be selected appropriately according to the size of the tested muscle.

1. Introduction

Electrical impedance myography (EIM) is a noninvasive and bioimpedance-based technique that assesses muscle health by applying very low-amplitude (usually a few milliamperes), high-frequency current through a localized area of tissue. It measures the resulting voltage with sensing electrodes on the skin [1]. There are three most commonly used EIM parameters [2], including resistance (R), reactance (X), and phase angle (θ), calculated as $\theta = \arctan(X/R)$. Electrical impedance myography can be used as a biomarker of neuromuscular diseases given that pathological changes (such as muscle atrophy, muscle fiber denervation/reinnervation, and the development of increasing intramuscular fat and connective tissue, etc.) will collectively influence normal impedance characteristics [3]. It has been reported to

quantify muscle changes in different neuromuscular diseases, such as amyotrophic lateral sclerosis (ALS), spinal muscular atrophy (SMA), and Duchenne muscular dystrophy (DMD) [4–6]. It has also been used to evaluate paretic muscle changes after neurological injuries [7–9].

Linear EIM involves placement of voltage electrodes along a line over the region of interest, and electrical current is injected far from that region. Previous studies adopted the approach of manually placing the two pairs of electrodes over the skin surface of the tested muscle and reported muscle composition alterations with neuromuscular diseases [4, 5]. This method represents the early stage of EIM and has demonstrated good test-retest reproducibility at 50 kHz [10, 11]. A handheld electrode array (HEA) that has been recently developed provides localized voltage and current electrodes for measuring EIM and has demonstrated

very high test-retest reproducibility with multifrequency analysis [12]. Multifrequency EIM is an extension of the linear EIM technique that supplies alternating currents with a range of frequencies rather than just a single frequency. Multifrequency EIM was reported to be more sensitive than single-frequency EIM (50 kHz) in tracking disease progression [13, 14] because multifrequency currents can aid in the extraction of muscle-specific properties (e.g., anisotropy) [15] from the frequency-dependent muscle evaluation.

A gap exists in the field of quantitative measurements of muscle mass, compositional quality, and contractile quality [16], particularly due to the lack of reliable device that can be readily and quickly applied in clinical settings. The EIM assessment procedure involves placing the electrodes on the skin surface. The effects of the skin-electrode interface or the contact sensitivity on bioimpedance outcomes have been a topic of interest [17–19]. Robust EIM measurements can only be obtained via an interface that allows for minor movements without loss of contact; the skin needs to be moist and conductive. Variations in contact levels between raters can result in differences in contact areas, which would reduce the reliability. Geometric factors can also significantly affect measurement reliability, particularly for anisotropic tissues such as skeletal muscles. This is related to the relative direction between the current-injecting electrodes and the underlying muscle fiber orientation. For example, Shiffman found that the muscle impedance was affected by the geometry of the electrode arrangements. The consequences of this entanglement often depend on frequency, making it difficult to extract the properties of the tissue if there are variations in electrode directions among different raters [20]. Particularly, there are different sizes of sensors for measuring EIM, which can affect the relative geometry of the electrode arrangement with respect to the examined muscle. The relative effect of skin and subcutaneous tissues on EIM parameters is also related to sensor size [21]. In addition, the distance between the sensor electrodes can also affect the EIM results. There is gap in the literature in assessing if large or small sensors are sufficiently reliable for clinical application, or if their reliability may differ when measuring a single muscle group by different raters at different time. A recent review by Clark et al. argued that there continues to be a lack of evidence to support the clinical application of EIM in assessing skeletal muscle function [22]. These issues must be addressed to facilitate clinical application of EIM. Therefore, it is critical to evaluate the reliability of EIM measurements and to identify the factors that may affect the EIM outcome via the assessment of between days intrarater and within-day interrater reliability using sensors of different sizes of multifrequency analysis. Motivated by this rationale, the current study assessed the reliability of the two handheld electrode array sensor devices. Intraclass correlation coefficient analysis was conducted to compare the EIM outcomes (R , X , and θ) of the two handheld electrode array sensors, based on the obtained several days apart by the same and different raters. The findings of the study can contribute to our understanding of the reliability of vivo EIM measurement.

2. Materials and Methods

2.1. Subjects. Healthy individuals were recruited from the student and staff population of the host institute via internal announcement. Included subjects had no reported history of neuromuscular disease. No upper limb weakness or functional impairment was present before or during the data collection period. The study was approved by the Committee for the Protection of Human Subjects (CPHS) of the University of Texas Health Science Center at Houston and TIRR Memorial Hermann Hospital (Houston, TX, USA). The Declaration of Helsinki was strictly followed. All of the subjects were provided with participant information sheet and encouraged to ask questions about the study. Informed written consent was obtained from all of the subjects prior to study enrolment.

2.2. Equipment. Impedance measurements were recorded from the biceps brachii muscle of the dominant limb by the HEA system (EIM1103, Skulpt Inc., Boston, MA, USA), which was used in our previous studies [23, 24]. A low-intensity electrical current at the frequencies ranging from 1 kHz to 10 MHz was applied in discrete logarithmic steps. The resulting surface voltages were then measured. Two different-sized sensor arrays (Model 20–00036, Small Sensor, and Model 20–0045, Large Sensor) were used in sequence for the repeated measurements. During each measurement, the sensor array was placed over the center of the biceps brachii muscle belly in a longitudinal direction over the muscle fibers. Each sensor contained a pair of current electrodes and a pair of voltage bar electrodes. For the large sensor, the distance between the pair of current electrodes (3.9 cm long; 0.4 cm wide) was 6.8 cm, and the distance between the pair of voltage bar electrodes (1.3 cm long; 0.4 cm wide) was 1.7 cm. Figure 1(a) shows the configuration of the large sensor. For the small sensor, the distance between the pair of current electrodes (2.6 cm long; 0.2 cm wide) was 3.4 cm, and the distance between the pair of voltage bar electrodes (0.8 cm long; 0.3 cm wide) was 0.8 cm. Figure 1(b) shows the configuration of the small sensor. Parameters recorded from the wide longitudinal configuration measured along the longitudinal direction were analyzed.

2.3. Procedures. The subjects sat on a height-adjustable chair. Their dominant arm rested at 90° flexion and the shoulder at 45° abduction. All data collection took place in the same laboratory. A constant temperature of approximately 22°C was maintained in the laboratory during all data collection sessions. The central air condition system also effectively maintained a stable humidity, although the humidity was not measured. The constant temperature and humidity would ensure similar physical environment for each test session. Each subject was recruited from office work status to participate the study. They were asked not to participate in physical activity on the day before data collection and were given sufficient time to acclimatise to the testing environment upon arrival at the laboratory.

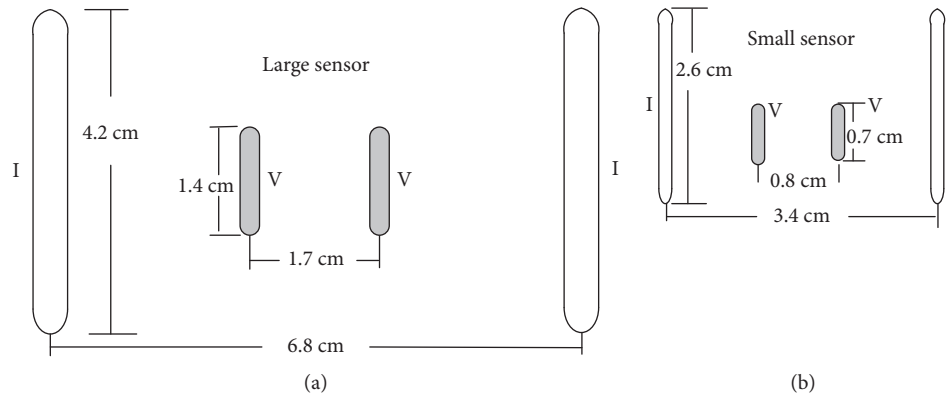


FIGURE 1: Configuration of electrodes and dimensions of large (a) and small (b) sensors.

The skin area of the electrode contact was moistened by sterile saline wipes (Hygea, PDI Inc., Hamilton, NY, USA) prior to performing impedance measurements. The HEA was then placed on the muscle until the green “Begin Test” signal was displayed on the computer screen. This indicated good skin contact and ensured the best measurement. Then, data recording was started. During the recording, the device was kept in place until the results were displayed on the screen. The software also plotted the resistance and reactance across the range of frequencies in real time.

Sensors of two different sizes were used in random sequence by each rater. The position of the two array sensors were marked on the skin to minimize confounding factor that was related to site identification between measurements. For intrarater reliability testing, the first rater conducted repeated measurements on two separate occasions with 5 to 8 days apart. Measurements for interrater reliability were collected on the second visit after the first rater completed the measurements.

The assessors received 4 hours of training from a senior technician on the standard operating procedure of the HEA system. This was followed by two hours of unsupervised practice. The associated software was used to visually inspect the data to ensure consistency over three trials.

2.4. Parameters. The parameters of resistance (R), reactance (X), and phase (θ) recorded from the longitudinal current electrodes were analyzed. All parameters were obtained at the frequencies of 50 kHz and 100 kHz. These two frequencies were chosen because they were within the optimal range to record EIM responses and were the most commonly used frequencies in previous studies [10, 11].

2.5. Data Analysis and Statistics. Statistics analysis was performed using SPSS 23 (IBM Corp., Armonk, NY, USA). The significance level was set at $p < 0.05$. Descriptive analysis was conducted to describe the sample population. The ICC models 2, k and 3, k were adopted to assess the relative interrater and intrarater reliability, respectively. The interpretation of ICC was as follows: ≥ 0.90 = high reliability, 0.80 – 0.89 = good reliability, 0.70 – 0.79 = fair reliability, and ≤ 0.69 = poor reliability [25]. Absolute reliability indices of

standard error of measurements (SEM), smallest real difference (SRD), and Bland–Altman 95% limits of agreement were calculated.

3. Results

Twenty healthy participants were recruited (mean age, 32.9 ± 8.2 ; 12 men and 8 women). A summary of the between-days and within-day longitudinal biceps EIM measurements is shown in Tables 1 and 2. The ICCs of all parameters recorded at 50 kHz and 100 kHz range between 0.90 to 0.98 for the large sensor and between 0.44 and 0.97 for the small sensor. In particular, we note that the ICC value of 50 kHz reactance using the small sensor was 0.440, which indicates poor reliability. The ICCs of the large sensor are higher than the small sensor in all parameters for between-days interrater measurements and within-day measurements. The ICC values recorded by the small sensor at 100 kHz were consistently higher than those recorded at 50 kHz in all parameters, except for reactance, for between-days and within-day measurements. Similar ICC values were observed at the measurements recorded at 50 kHz and 100 kHz by the large sensor. These results indicated that the reliability of the small sensor is prone to be affected by sampling frequency. Tables 3 and 4 present the results of ICC analyses of between-days and within-day measurements recorded by each sensor. Figures 2 and 3 present the ICC plots of longitudinal bicep measurements recorded by the large sensor and small sensor taken by two different raters and by the same rater on different days, respectively.

The SEM and SRD recorded by the large sensor were smaller than those recorded by the small sensor for all measurements, indicating a higher variation around the “true” score and that a larger change is required to be deemed real change. Bland–Altman 95% LOA indicates larger error range between measurements recorded by the small sensor than the large sensor. Tables 3 and 4 present the results of absolute reliability indices for the between-days and within-day measurements recorded by the two assessors using both sensors.

TABLE 1: Between-day longitudinal biceps EIM measurements.

		Large sensor			Small sensor		
		1st measurement	2nd measurement	d	1st measurement	2nd measurement	d
50 kHz	Phase	20.93(3.52)	21.38 (3.25)	-0.45	12.23 (4.31)	11.75 (3.58)	0.48
	Reactance	9.83 (1.08)	10.13 (1.27)	-0.31	13.31 (3.10)	12.61 (2.72)	0.70
	Resistance	26.59 (6.91)	26.70 (7.16)	-0.11	64.54 (16.60)	63.18 (16.40)	1.36
100 kHz	Phase	22.13 (3.54)	22.56 (3.51)	-0.43	15.17 (3.82)	14.64 (3.27)	0.53
	Reactance	8.46 (1.11)	8.63 (1.31)	-0.17	14.97 (3.16)	14.19 (3.08)	0.77
	Resistance	21.63 (6.38)	21.56 (6.52)	0.07	57.11(15.54)	56.27 (15.53)	0.85

Note: mean (standard deviation); d : mean difference between first and second measurements.

TABLE 2: Within-day longitudinal biceps EIM measurements.

		Large sensor			Small sensor		
		1st measurement	2nd measurement	d	1st measurement	2nd measurement	d
50 kHz	Phase	21.54 (3.34)	21.38 (3.25)	0.15	12.11 (4.07)	11.75 (3.58)	0.36
	Reactance	9.85 (1.16)	10.13 (1.27)	-0.32	12.73 (3.51)	12.61 (2.72)	0.12
	Resistance	25.32 (5.37)	26.70 (7.16)	-1.38	61.67 (7.16)	63.18 (16.39)	-1.38
100 kHz	Phase	22.73 (3.80)	22.56 (3.51)	0.18	14.98 (3.56)	14.64 (3.27)	0.34
	Reactance	8.36 (1.01)	8.63 (1.31)	-0.27	14.14 (3.11)	14.19 (3.08)	-0.05
	Resistance	20.49 (5.04)	21.56 (3.52)	-1.06	54.59 (14.19)	56.27 (15.53)	-1.68

Note: mean (standard deviation); d : mean difference between first and second measurements.

TABLE 3: Reliability indices for within-day interrater reliability of both sensors.

Within-day interrater		Large sensor				Small sensor			
		ICC	SEM	SRD	95% LOA	ICC	SEM	SRD	95% LOA
50 kHz	Phase	0.97 (0.92–0.99)	0.19	1.22	-6.38–4.17	0.92 (0.81–0.97)	0.59	2.13	-10.44–6.20
	Reactance	0.90 (0.73–0.96)	0.21	1.27	-4.62–3.29	0.44 (-0.41–0.78)	2.62	4.49	-15.97–8.96
	Resistance	0.99 (0.98–1.00)	0.12	0.96	-7.23–4.59	0.95 (0.88–0.98)	1.53	3.42	-30.09–16.03
100 kHz	Phase	0.98 (0.94–0.99)	0.13	0.99	-5.80–3.88	0.92 (0.81–0.97)	0.51	1.98	-9.41–5.69
	Reactance	0.95 (0.88–0.98)	0.11	0.94	-4.00–2.98	0.67 (0.19–0.87)	1.76	3.68	-14.31–8.13
	Resistance	0.99 (0.98–1.00)	0.11	0.92	-6.34–4.15	0.97 (0.93–0.99)	0.82	2.51	-22.12–12.04

TABLE 4: Reliability indices for between-days intrarater reliability of both sensors.

Between-days intrarater		Large sensor				Small sensor			
		ICC	SEM	SRD	95% LOA	ICC	SEM	SRD	95% LOA
50 kHz	Phase	0.98 (0.95–0.99)	0.13	0.99	-5.66–3.81	0.84 (0.59–0.94)	1.87	3.79	-13.59–7.78
	Reactance	0.90 (0.73–0.96)	0.22	1.30	-4.73–3.35	0.63 (0.02–0.85)	3.30	5.04	-15.28–8.62
	Resistance	0.95 (0.85–0.98)	0.54	2.03	-11.73–6.85	0.95 (0.88–0.98)	2.38	4.27	-29.09–15.52
100 kHz	Phase	0.98 (0.96–0.99)	0.13	1.00	-5.85–3.90	0.91 (0.76–0.96)	0.99	2.76	-10.10–6.30
	Reactance	0.90 (0.74–0.96)	0.22	1.31	-4.76–3.36	0.84 (0.60–0.94)	1.47	3.36	-11.24–6.60
	Resistance	0.96 (0.88–0.98)	0.43	1.81	-10.49–6.22	0.96 (0.89–0.98)	1.98	3.90	-25.98–13.97

4. Discussion

This study aimed to assess the between-day intrarater reliability and within-day interrater reliability of two difference-sized array sensors to record muscle impedance. The results of the present study indicated that the large sensor was consistently more reliable than the small sensor under all tested conditions. The high ICCs are consistent with previous reliability evaluations of localized EIM applications on healthy tibialis anterior muscles [12], as well as in DMD [6, 26]. Rutkove et al. reported high test-retest reliability of linear EIM performed on biceps, quadriceps, and tibialis

anterior muscles (ICC = 0.970, 0.971, and 0.938, respectively), 250 days apart at 50 kHz [10].

One of the possible factors that contributes to the lower reliability of the small sensor is the effect of subcutaneous fat over a smaller distance between electrodes. Jafarpour et al. applied a finite-element model to mimic human upper arm muscles and found that the electrode distance dramatically affected the EIM outputs [27]. All of the tissues beneath the voltage electrodes can in theory contribute to the impedance measurement, including subcutaneous fat and bone. The muscle inherent resistivity is lower than both subcutaneous fat and bones [28]. Therefore, the electrical current tends to

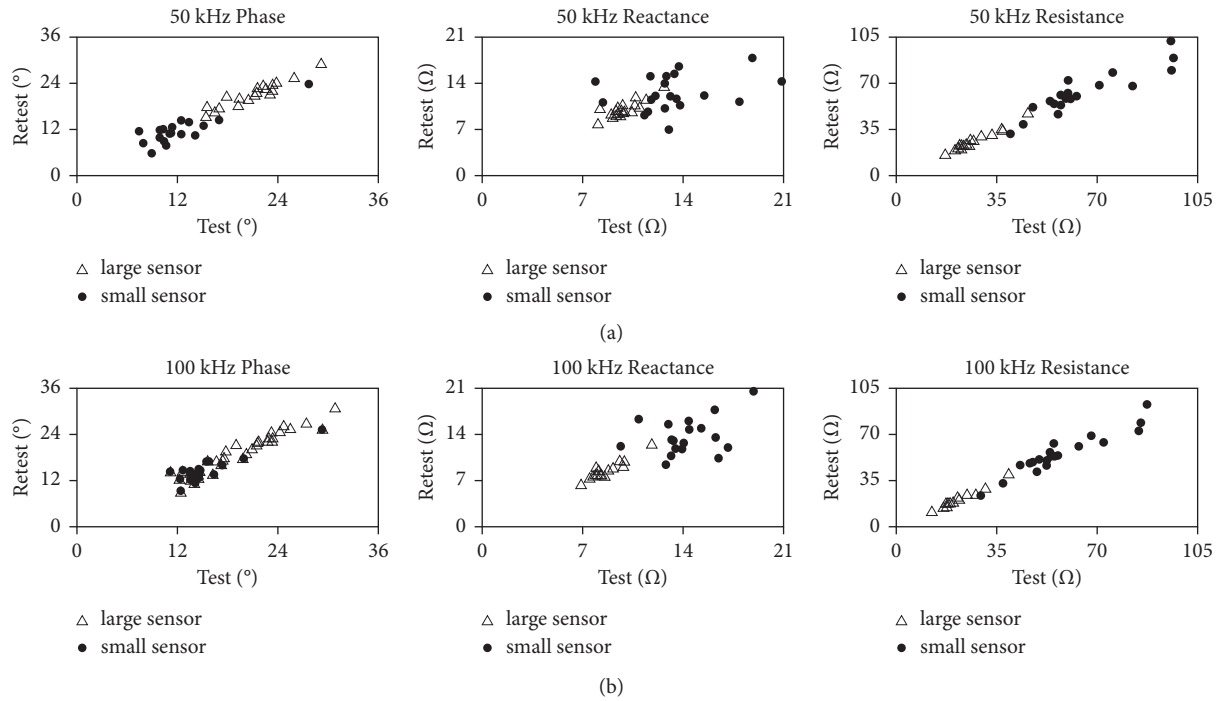


FIGURE 2: ICC plots of longitudinal bicep measurements in healthy subjects using large sensors (open triangles) and small sensors (filled circles) taken by two different (inter-) raters. (a) Plots of 50 kHz data. (b) Plots of 100 kHz data.

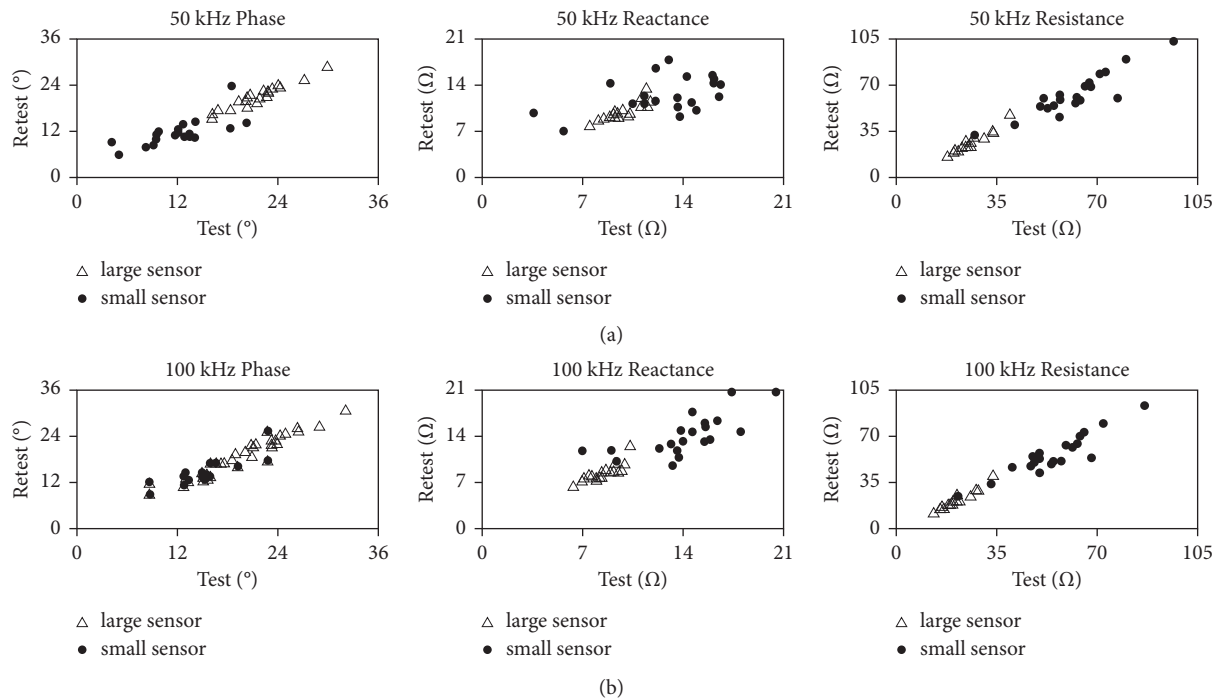


FIGURE 3: ICC plots of longitudinal bicep measurements in healthy subjects using large sensors (open triangles) and small sensors (filled circles) taken at baseline and 5–8 days later by the same (intra-) rater. (a) Plots of 50 kHz data. (b) Plots of 100 kHz data.

pass through the muscle. However, Sung et al. suggested that when currents passed through tissues that were underneath the voltage-measuring electrodes, a considerable proportion of the currents went through the fat at the points closest to

where the voltage was recorded [29]. A significant correlation was found between the subcutaneous fat layer (SFL) thickness and the resistance and phase for the medial gastrocnemius muscle in healthy subjects. Jafarpour et al. also

reported that the thickness of the SFL of the quadriceps muscle did contribute to resistance and reactance [27]. Therefore, a shorter electrode distance (small sensor) is more likely to be affected by the SFL than a larger electrode distance (large sensor) [20]. In addition, bioimpedance analysis is influenced not only by the degree of fat mass but also by the electrode-skin interface [19]. The electrode area and distance effects have been evaluated in linear EIM. Rutkove et al. found that a fraction of the effect is related to the electrochemical properties of the electrode-skin-fat-layer complex, whose contribution is inversely proportional to the electrode area [11]. Therefore, the EIM output of the small sensor is more prone to be affected by the electrode-skin-fat-layer than the larger sensor. This in turn may affect the reliability of the small sensor. Findings of this study indicated that EIM measurements should be recorded by electrodes that are properly sized and spaced. Measurement recorded by a relatively small sensor should be considered with caution.

This study observed smaller SEM and SRD recorded by the small sensor at 100 kHz than at 50 kHz, suggesting that the reliability of the instrument is affected by recording frequency. Several studies indicated that multifrequency measures may be more sensitive to disease status and progression over time [13, 14]. By changing the frequency of the current injected, it can shift the relative weights of the resistive (fluid) and reactive (membranes) contributions to the total impedance (i.e., the cell membrane acts like a capacitor in an electrical circuit, such that a very high frequency makes nearly no reactance contribution to the impedance) [30]. At low frequencies, the contact impedance between the sensor and skin is higher, which results in less-reliable data. Hewson et al. studied the impedance of electrode-skin interface using multifrequency (1–16,384 Hz) and found large individual differences in the level of impedance at low frequencies [31]. The findings of this study suggest that higher frequencies (i.e., higher than 50 kHz) should be considered when using the small sensor to record impedance properties.

Reactance has the lowest value of reliability indices among the three EIM parameters recorded by either sensor. Possible contributing factors to the low reactance reliability might be the angle at which the current was applied and the angle measured relative to the muscle fiber direction, and the contact pressure on the skin. There might be small variations in the electrode positions and the angles relative to the biceps brachii each time the electrodes were placed on the skin surface by the rater. The variation in electrode position and the angle relative to biceps brachii may affect the EIM parameters. Tarulli et al. reported that reactance was most sensitive to the angle at which current was applied and the angle measured relative to the muscle fiber direction in bovine skeletal muscles [32]. The sensitivity may be related to the anisotropy of muscle tissue because anisotropy represents the inherent muscle fiber geometry within the muscle, and the electrical currents flow more readily along muscle fibers than passing across them [33]. Therefore, any alteration of the electrode angle affects the direction of injecting currents to the muscle, and in turn, it contributes to

the variation of reactance measurements. In addition, reactance is associated with the tendency for oscillating charges to accumulate against the capacitors in the muscle cell membranes [2]. Resistor-capacitor models of the electrode-skin-fat layer interface revealed that the interface resulted in a drift in the EIM results known as the “voltage divider effect,” which depends on the input characteristics of the impedance-measuring instrument, the area, and the conditions of the electrode-skin interface [11]. It can be reduced by increasing the input impedance of the instrument and by using active, low-capacitance probes. Because the measured reactance itself is modeled as a “capacitor,” it is prone to poor contacts between electrode and interface, which in turn can cause the low reliability of reactance.

The limitations of the study should be acknowledged. Although the findings of the current study with healthy population provides useful information about the reliability of small and large EIM sensors, a clinical population was not included in the study while it is important to further extend the reliability investigation to different clinical populations where the EIM can find most important applications. The current study only examined the biceps brachii muscle, although different types of muscles should be further examined. In addition, the same EIM experiment performed by one of the two raters (i.e., the same data set) was used for both the between-days intrarater reliability and within-day interrater reliability investigations. The protocol can be further improved by having both raters take measurements on both days. According to the literature, the variation of muscle size can affect resistance and reactance measurement [29]. For example, the decrease in muscle size might increase measured R and X . However, the phase angle (θ) would be affected to a less extent because its calculation would potentially cancel out simple volumetric effects of the measured muscle [29, 30]. The current study investigated the performance of the large and small sensors over the biceps brachii muscle. It remains a future study to further examine how the EIM parameters change with different muscles.

5. Conclusions

EIM measurements of the biceps brachii muscle recorded by the large sensor demonstrated high between-days intrarater and within-day interrater reliability for all three parameters. The small sensor demonstrated poor to high reliability, and repeated measurements were not as consistent as the large sensor. Reactance recorded by the small sensor was the least reliable parameter. The reliability of small sensor was prone to be affected by measuring frequency. Findings of this study suggest that electrode size and interelectrode distance should be selected appropriately in accordance with the size of the tested muscle to achieve the desired reproducibility of EIM measurements.

Data Availability

The data sets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Huijing Hu and Wai Leung Ambrose Lo contributed equally to this work.

Acknowledgments

This work was supported in part by the Guangzhou Science and Technology Program (201704030039 and 201907010034), the National Natural Science Foundation of China (81702227, 31771016, 32071316, and 81971124), and the Shandong Provincial Natural Science Foundation (ZR2020KF012).

References

- [1] S. B. Rutkove, "Electrical impedance myography: background, current state, and future directions," *Muscle & Nerve*, vol. 40, no. 6, pp. 936–946, 2009.
- [2] S. B. Rutkove, R. Aaron, and C. A. Shiffman, "Localized bioimpedance analysis in the evaluation of neuromuscular disease," *Muscle & Nerve*, vol. 25, no. 3, pp. 390–397, 2002.
- [3] J. Li, T. R. Geisbush, G. D. Rosen, J. Lachey, A. Mulivor, and S. B. Rutkove, "Electrical impedance myography for the in vivo and ex vivo assessment of muscular dystrophy (mdx) mouse muscle," *Muscle & Nerve*, vol. 49, no. 6, pp. 829–835, 2014.
- [4] S. B. Rutkove, J. M. Shefner, M. Gregas et al., "Characterizing spinal muscular atrophy with electrical impedance myography," *Muscle & Nerve*, vol. 42, no. 6, pp. 915–921, 2010.
- [5] A. W. Tarulli, L. P. Garmirian, P. M. Fogerson, and S. B. Rutkove, "Localized muscle impedance abnormalities in amyotrophic lateral sclerosis," *Journal of Clinical Neuromuscular Disease*, vol. 10, no. 3, pp. 90–96, 2009.
- [6] C. M. Zaidman, L. L. Wang, A. M. Connolly et al., "Electrical impedance myography in Duchenne muscular dystrophy and healthy controls: a multicenter study of reliability and validity," *Muscle & Nerve*, vol. 52, no. 4, pp. 592–597, 2015.
- [7] X. Li, L. Li, H. Shin, S. Li, and P. Zhou, "Electrical impedance myography for evaluating paretic muscle changes after stroke," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2113–2121, 2017.
- [8] L. Li, H. Shin, A. Stampas, X. Li, and P. Zhou, "Electrical impedance myography changes after incomplete cervical spinal cord injury: an examination of hand muscles," *Clinical Neurophysiology*, vol. 128, no. 11, pp. 2242–2247, 2017.
- [9] L. Li, A. Stampas, H. Shin, X. Li, and P. Zhou, "Alterations in localized electrical impedance myography of biceps brachii muscles paralyzed by spinal cord injury," *Frontiers in Neurology*, vol. 8, p. 253, 2017.
- [10] S. B. Rutkove, K. S. Lee, C. A. Shiffman, and R. Aaron, "Test-retest reproducibility of 50kHz linear-electrical impedance myography," *Clinical Neurophysiology*, vol. 117, no. 6, pp. 1244–1248, 2006.
- [11] S. B. Rutkove, R. A. Partida, G. J. Esper, R. Aaron, and C. A. Shiffman, "Electrode position and size in electrical impedance myography," *Clinical Neurophysiology*, vol. 116, no. 2, pp. 290–299, 2005.
- [12] P. Narayanaswami, A. J. Spieker, P. Mongiovi, J. C. Keel, S. C. Muzin, and S. B. Rutkove, "Utilizing a handheld electrode array for localized muscle impedance measurements," *Muscle & Nerve*, vol. 46, no. 2, pp. 257–263, 2012.
- [13] S. B. Rutkove, M. C. Gregas, and B. T. Darras, "Electrical impedance myography in spinal muscular atrophy: a longitudinal study," *Muscle & Nerve*, vol. 45, no. 5, pp. 642–647, 2012.
- [14] S. Schwartz, T. R. Geisbush, A. Mijailovic, A. Pasternak, B. T. Darras, and S. B. Rutkove, "Optimizing electrical impedance myography measurements by using a multifrequency ratio: a study in Duchenne muscular dystrophy," *Clinical Neurophysiology*, vol. 126, no. 1, pp. 202–208, 2015.
- [15] A. B. Chin, L. P. Garmirian, R. Nie, and S. B. Rutkove, "Optimizing measurement of the electrical anisotropy of muscle," *Muscle & Nerve*, vol. 37, no. 5, pp. 560–565, 2008.
- [16] F. Buckinx, F. Landi, M. Cesari et al., "Pitfalls in the measurement of muscle mass: a need for a reference standard," *Journal of Cachexia, Sarcopenia and Muscle*, vol. 9, no. 2, pp. 269–278, 2018.
- [17] F. L. H. Gielen, W. Wallinga-de Jonge, and K. L. Boon, "Electrical conductivity of skeletal muscle tissue: experimental results from different muscles in vivo," *Medical, & Biological Engineering & Computing*, vol. 22, no. 6, pp. 569–577, 1984.
- [18] Y. Wang, P. H. Schimpf, D. R. Haynor, and Y. Kim, "Geometric effects on resistivity measurements with four-electrode probes in isotropic and anisotropic tissues," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 7, pp. 877–884, 1998.
- [19] F. Zhu, E. F. Leonard, and N. W. Levin, "Body composition modeling in the calf using an equivalent circuit model of multi-frequency bioimpedance analysis," *Physiological Measurement*, vol. 26, no. 2, pp. S133–S143, 2005.
- [20] C. A. Shiffman, "Adverse effects of near current-electrode placement in non-invasive bio-impedance measurements," *Physiological Measurement*, vol. 34, no. 11, pp. 1513–1529, 2013.
- [21] S. B. Rutkove, A. Pacheck, and B. Sanchez, "Sensitivity distribution simulations of surface electrode configurations for electrical impedance myography," *Muscle & Nerve*, vol. 56, no. 5, pp. 887–895, 2017.
- [22] B. C. Clark, S. Rutkove, E. C. Lupton, C. J. Padilla, and W. D. Arnold, "Potential utility of electrical impedance myography in evaluating age-related skeletal muscle function deficits," *Frontiers in Physiology*, vol. 12, Article ID 666964, 2021.
- [23] J. Li, W. L. Staats, A. Spieker, M. Sung, and S. B. Rutkove, "A technique for performing electrical impedance myography in the mouse hind limb: data in normal and ALS SOD1 G93A animals," *PLoS One*, vol. 7, no. 9, Article ID e45004, 2012.
- [24] L. Li, X. Li, H. Hu, H. Shin, and P. Zhou, "The effect of subcutaneous fat on electrical impedance myography: electrode configuration and multi-frequency analyses," *PLoS One*, vol. 11, no. 5, Article ID e0156154, 2016.
- [25] C. R. Meyers and T. E. Blesh, *Measurement in Physical Education*, Taylor & Francis, New York, USA, 1962.
- [26] T. R. Geisbush, N. Visyak, L. Madabusi, S. B. Rutkove, and B. T. Darras, "Inter-session reliability of electrical impedance myography in children in a clinical trial setting," *Clinical Neurophysiology*, vol. 126, no. 9, pp. 1790–1796, 2015.
- [27] M. Jafarpoor, L. Jia, J. K. White, and S. B. Rutkove, "Optimizing electrode configuration for electrical impedance measurements of muscle via the finite element method," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1446–1452, 2013.

- [28] T. J. C. Faes, H. A. V. D. Meij, J. C. D. Munck, and R. M. Heethaar, "The electric resistivity of human tissues (100 Hz-10 MHz): a meta-analysis of review studies," *Physiological Measurement*, vol. 20, no. 4, pp. R1-R10, 1999.
- [29] M. Sung, A. J. Spieker, P. Narayanaswami, and S. B. Rutkove, "The effect of subcutaneous fat on electrical impedance myography when using a handheld electrode array: the case for measuring reactance," *Clinical Neurophysiology*, vol. 124, no. 2, pp. 400-404, 2013.
- [30] C. A. Shiffman, R. Aaron, V. Amoss, J. Therrien, and K. Coomler, "Resistivity and phase in localized BIA," *Physics in Medicine and Biology*, vol. 44, no. 10, pp. 2409-2429, 1999.
- [31] D. J. Hewson, J.-Y. Hogrel, Y. Langeron, and J. Duchène, "Evolution in impedance at the electrode-skin interface of two types of surface EMG electrodes during long-term recordings," *Journal of Electromyography and Kinesiology*, vol. 13, no. 3, pp. 273-279, 2003.
- [32] A. W. Tarulli, A. B. Chin, R. A. Partida, and S. B. Rutkove, "Electrical impedance in bovine skeletal muscle as a model for the study of neuromuscular disease," *Physiological Measurement*, vol. 27, no. 12, pp. 1269-1279, 2006.
- [33] L. P. Garmirian, A. B. Chin, and S. B. Rutkove, "Discriminating neurogenic from myopathic disease via measurement of muscle anisotropy," *Muscle & Nerve*, vol. 39, no. 1, pp. 16-24, 2009.

Research Article

Persistent Homology-Based Topological Analysis on the Gestalt Patterns during Human Brain Cognition Process

Zaisheng Liu ¹, Fei Ni ¹, Rongpeng Li ¹, Honggang Zhang ¹, Chang Liu ²,
Jiefang Zhang ² and Songyun Xie ³

¹College of Information Science and Electronic Engineering, Zhejiang University, Zheda Road 38, Hangzhou 310027, China

²Communication University of Zhejiang, Hangzhou 310018, China

³Northwestern Polytechnical University, Xi'an 710129, China

Correspondence should be addressed to Rongpeng Li; lirongpeng@zju.edu.cn

Received 19 May 2021; Revised 26 August 2021; Accepted 17 September 2021; Published 1 November 2021

Academic Editor: Siti Anom Ahmad

Copyright © 2021 Zaisheng 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.

The neuropsychological characteristics inside the brain are still not sufficiently understood in previous Gestalt psychological analyses. In particular, the extraction and analysis of human brain consciousness information itself have not received enough attention for the time being. In this paper, we aim to investigate the features of EEG signals from different conscious thoughts. Specifically, we try to extract the physiologically meaningful features of the brain responding to different contours and shapes in images in Gestalt cognitive tests by combining persistent homology analysis with electroencephalogram (EEG). The experimental results show that more brain regions in the frontal lobe are involved when the subject perceives the random and disordered combination of images compared to the ordered Gestalt images. Meanwhile, the persistence entropy of EEG data evoked by random sequence diagram (RSD) is significantly different from that evoked by the ordered Gestalt (GST) images in several frequency bands, which indicate that the human cognition of the shape and contour of images can be separated to some extent through topological analysis. This implies the feasibility to digitize the neural signals while preserving the whole and local features of the original signals, which are further verified by our extensive experiments. In general, this paper evaluates and quantifies cognitively related neural correlates by persistent homology features of EEG signals, which provides an approach to realizing the digitization of neural signals. Preliminary verification of the analyzability of human consciousness signals provides reliable research ideas and directions for the realization of feature extraction and analysis of human brain consciousness cognition.

1. Introduction

In recent years, with the development of neural networks, researchers are committed to explaining the intrinsic nature of human consciousness generation and artificial intelligence (AI). One of the research directions is to explore the laws of human brain cognition and consciousness generation process to promote the development of machine learning technology. In communication technology, the realization of brain-to-brain communication (B2BC) under the support of future 6G technology also urgently needs a method to realize the digitization of human brain nerve signals to support the development of its research. The most typical analysis method of electroencephalogram (EEG) signals is based on the brain signals' characteristics by

filtering, artifacts removing, event-related potentials (ERP) analysis, and brain domain heat map with respect to the original time-domain signals. The complex and dynamic multichannel time-domain signal is not an ideal carrier for information transmission. Currently, various digital analysis methods based on the EEG signals are constantly being proposed and improved [1–4], such as single-trial analysis and other diverse methods, which take into account the significant differences in EEG signals between different subjects. Furthermore, [3] attempts to extract the digital features that may be more relevant and simpler regarding the signal, which coincides with the first step of B2BC: the digitization process of neural signals. Combining the relevance enlightenment of B2BC and AI, the analysis of the human brain's cognitive process is of forward-looking value.

Gestalt psychology theory is the pioneering foundation of modern cognitive learning theory, which was established in the early twentieth century by psychologists Westheimer W. Kohler and K. Koffka based on similitude study [5]. They believe that thinking is a holistic and meaningful perception rather than a simple collection of connected representations and argue that learning lies in forming a gestalt, which aims to change one gestalt into another. This cognition process is fundamentally different from the current image recognition model of deep learning empowered by artificial neural networks (ANNs).

At present, with the rapid development of deep learning, newer and stronger algorithm models emerge endlessly, and their computing power and learning ability for specified tasks become increasingly powerful. On top of this basis, some researchers initiate ambitious new goals and turn to focus on making AI more “intelligent,” that is, achieving brain-like intelligence. They expect machine learning to achieve what the human brain can do and solve problems or recognize things like the human brain. It is well known that conventional neural networks are designed by the inspiration of the fundamental principle of signal transmission in a single nerve cell. Therefore, the linkage nature of the whole biological neural network is the direction we need to explore as the next frontier. It can inspire us to build brain-like intelligence and transfer from the traditional machine learning process to a more advanced consciousness level. By exploring the act of brain cognition, it is potentially possible to probe into the generation of consciousness [6], not limited to what kind of consciousness the brain produces.

Researchers have made attempts to explore neural networks and the human brain’s biological patterns [7–15]. To gain insight into the brain’s response to external stimuli, scientists have developed functional network analysis methodology because they assert that brain functionality is determined by the internal interaction between different neurons as well as different brain regions [11]. They set out to analyze the neural signals displayed by the brain as a whole. The spatial and connective relationship of neurons within the brain structure is a complex connection model, which has been used to analyze the human brain’s activity with topological tools for a long time. The initial focus of this kind of research is on the somatosensory sensations (e.g., hot and cold sensations and pain sensations) that are easy to recognize in brain signals [12]. Afterward, abnormal EEG responses (epilepsy seizures) [13], steady-state visually evoked potential (SSVEP), movement intention detection [14], and emotion classification [15] were analyzed. We compared the results obtained by various EEG analysis methods in various fields. Among them, the detection accuracy of schizophrenia reached a balanced accuracy of 89.59%; the detection of moving images reached a recognition rate of 64.9%–79.5%; and the classification of emotional EEG signals based on gender reached 90.4% (SVM) and 92% (KNN) [16].

In particular, inspired by the recent research on Gestalt recognition for which Baker et al. [17] and Been Kim et al. [18] provided totally conflicting conclusions, the discussion on the differences between artificial neural networks and the human brain cognition process has motivated us to follow research on the cognitive process at the level of consciousness.

Nevertheless, it is well realized that the human brain’s cognition on the overall outline of geometric patterns faces local and global problems in the previous Gestalt experiments.

The mathematic tools of algebraic topology are uniquely equipped to provide quantitative information about both the local and global properties of an arbitrary graph [8]. Accordingly, topological data analysis (TDA) is capable of providing a series of new topological and geometric methods to analyze the brain’s neural networks covering EEG signals, among which persistent homology is one of the key approaches. [19–24]. Persistent homology analysis provides efficient algorithms for calculating the Betti number of each complex graph in the network families under consideration and encodes the evolution of the nested complex homology groups at different networking scales. Consequently, it helps understand the EEG data better and keeps analytical stability concerning perturbations or noise in the EEG signals.

In our study, 20 participants were considered in traditional visual stimulation experimental methods and collecting EEG signals at the same time. The neurophysiological evaluation of the contour in the Gestalt experiment was investigated by exploiting Euler characteristics and persistent homology features of the EEG signal. On that basis, how the regions of the brain are involved in contour recognition was interpreted by selecting Vietoris–Rips filtration.

The main contributions of this work are as follows:

- (1) The topology calculation method adopted in this experiment provides effective separability of the EEG signals of contour cognitive behavior and realizes the digital feature extraction of EEG signals
- (2) We provide reference significance and a reference method for B2BC and other work that needs to realize the digital feature extraction of EEG signals
- (3) We demonstrated the feasibility of using persistent homology modeling to analyze EEG signals

2. Experiments and Methods for Assessing Brain Cognition’s Gestalt Patterns

The general framework for the neurophysiological assessment process and method of Gestalt contour cognition is illustrated in Figure 1, which is based on topological data analysis enabled by persistent homology. During the brain cognition process, subjects first watch random sequence diagram (RSD) pictures repeatedly at fixed intervals and then watch Gestalt (GST) images in the same manner. Meanwhile, the EEG data are collected by a special cap with sensing electrodes synchronously (Step I). Then two methods are used for calculating the correlation coefficient: one is to calculate the phase correlation coefficient (0–1) of the EEG signals between the sensors through the algorithm based on Hilbert transform to construct the correlation matrix and the other is to calculate the standardized Euclidean distance between the sensors to construct a distance matrix (Step II) [13] for obtaining the topological Vietoris–Rips simplex (Step III). Finally, persistent homology methodology is applied to analyze the brain’s neurophysiological features stimulated by various pictures across different qualities (Steps IV and V).

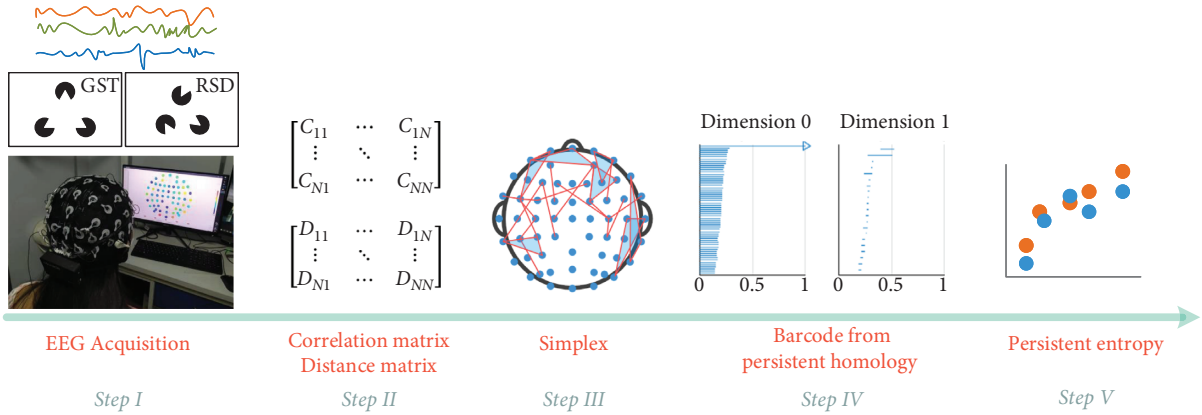


FIGURE 1: The neurophysiological process and method of Gestalt contour cognition with topological data analysis.

2.1. Stimuli. As shown in Figure 2, we have selected two representative types of Gestalt pictures with the existence and nonexistence of a specific outline of a standard triangle. One is the picture of GST that people can easily recognize the outline of the triangle, and the others are the pictures of RSD. The size and quality of the two types of pictures are the same, and both are $1,440 \times 1,080$ resolution. To explore the characteristics of changing consciousness in a subject's cognition process, we repeat the RSD 30 times and then repeated the GST 10 times to increase the samples' amount and eliminate potential experimental errors.

2.2. Procedure. After a general introduction to the experiment and the EEG cap preparation, the subjects start the test with EEG recording, which is described in Figure 3. The EEG recordings of two cognition periods correspond to two continuous stages in the whole process. The first stage is to collect the EEG signals when the subject does not have a clear cognition of RSD, and the second stage is to collect the signals when the subject recognizes the outline of the triangle from the GST. When the subjects start to identify the triangle's intrinsic outline from the GST, each trial began with a fixed time slot that lasts 1 second, and then the RSD or GST image appear for 10 seconds. After that, a rest time slot appears to remind the subjects that they could take a break for 1 second. One by one, the subsequent trials start to run.

2.3. Subjects and Equipment. The EEG data are measured from 20 healthy volunteers (9 males and 11 females, in the age group 19–27) with normal (or corrected to normal) vision. Volunteers are mainly sophomores and juniors. The main age group is 22 years old, with an average age of 22.4 and a standard deviation of 1.71. The experiment equipment is a standard Neuracle 64 System, which includes a 64-channel adult-sized head cap with the sensor array, EEG recorder with EEG acquisition software, and amplifier (NSW364). The sampling rate is 1,000 Hz for the EEG signals, and the filtering window is changed with frequency from 0.3 to 100 Hz.

2.4. Topological Data Analysis for EEG Data. Topological data analysis for the EEG data has been summarized in Figure 1, and the following provides the corresponding details.

2.4.1. EEG Signals Acquisition and Preprocessing. The EEG data are collected by the EEG cap and downsampled to 250 Hz. Filtered EEG signals of different wavebands are obtained by a set of filters, including δ band (1–3 Hz), θ band (4–7 Hz), α band (8–13 Hz), β band (14–30 Hz), and the whole band (1–45 Hz).

The specific operations are as follows: during the entire acquisition process, we mark the EEG signals corresponding to different events to facilitate subsequent trial segmentation. Since the entire acquisition process is continuous, considering the activity frequency of the human brain under normal conditions, we first perform (1–45 Hz) filtering on the entire time-domain signal and then try segment to extract the target data we need, based on this perform subsequent data adjustments such as baseline calibration and downsampling.

The filtered signals from each electrode of the EEG cap correspond to a set of measuring points G . As explained before, two data analysis methods are used for characterizing the brain cognition process: one is to calculate the correlation matrix through the real-time phase relationship and the other is to define the distance matrix for each point through the signal-level correlation.

2.4.2. Correlation Matrix Computing. The calculation steps are as follows:

- (1) After the key feature extraction and preprocessing of the EEG signals, we get the signal of each trial period as follows:

$$F_{\text{EEG}} = \begin{bmatrix} f_{11} & \cdots & f_{1N} \\ \vdots & \ddots & \vdots \\ f_{M1} & \cdots & f_{MN} \end{bmatrix}, \quad (1)$$



FIGURE 2: Two types of Gestalt pictures used for the brain cognition experiments: GST and RSD.

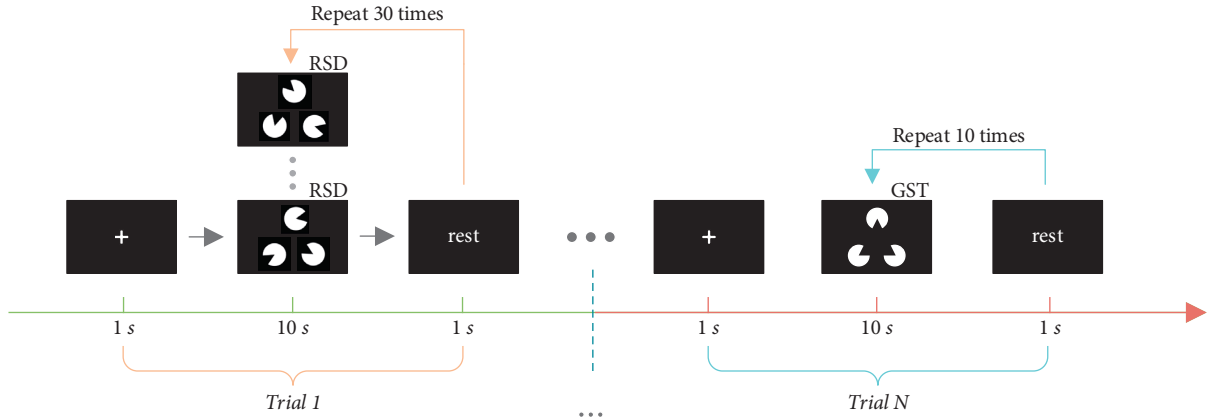


FIGURE 3: The procedure of Gestalt pattern cognition.

where N is the total data length, which is equal to the sampling rate multiplied by the measurement time, and M is the total number of electrodes that collect the EEG signals. Each row in the matrix F_{EEG} represents the signals collected by one electrode.

- (2) Hilbert transform [25] is performed on each signal in F_{EEG} , that is, each row, to obtain a new matrix $H(F_{\text{EEG}})$.
- (3) $H(F_{\text{EEG}})$ obtained by step 2 is used to calculate the instantaneous phase of each electrode:

$$\phi = \arctan\left(\frac{H(F_{\text{EEG}})}{F_{\text{EEG}}}\right). \quad (2)$$

- (4) The value of the corresponding element of the incidence matrix is calculated by equation (3). The absolute value is taken and then combined to obtain the incidence matrix equation (4):

$$C_{pq} = \begin{cases} \frac{1}{N} \left| \sum_{n=1}^N \exp(j\{\phi^p(n) - \phi^q(n)\}) \right|, & p \neq q, \\ 0, & p = q, \end{cases} \quad (3)$$

$$C_{M \times M} = \begin{bmatrix} C_{11} & \cdots & C_{1M} \\ \vdots & \ddots & \vdots \\ C_{M1} & \cdots & C_{MM} \end{bmatrix}, \quad (4)$$

where j is an imaginary unit and $\phi^p(n)$ and $\phi^q(n)$ represent the n -th instantaneous phase in the electrode p and q , respectively.

2.4.3. Distance Matrix Computing. The filtered signal from each electrode in the EEG cap constitutes a set of sampling points G , and the distance between electrodes with different channels is calculated by [13]

$$d(r, t) = \sqrt{\sum_{k=1}^N \left(\frac{r|_k}{s_k} - \frac{t|_k}{s_k} \right)^2}, \quad (5)$$

where $r|_k$ and $t|_k$ stand for the y -component of different electrodes in (x_k, y_k) and s_k is the sample standard deviation calculated from all y -components at position k in channel r .

2.4.4. Simplicial Complexes Construction. Simplicial complexes are constructed by Vietoris–Rips filtration according to either the correlation matrix or the distance matrix obtained in Step II, which is illustrated in Figure 4.

2.4.5. Euler Characteristics. Before calculating the persistent entropy, we supplement Euler entropy to do a preliminary analysis of the separability of the topological properties of the data, which also provides a basis for the persistent homology separability. The topological structures of original EEG signals are constructed by Vietoris–Rips filtration: one uses the phase-locked value (PLV) of the EEG signals data as

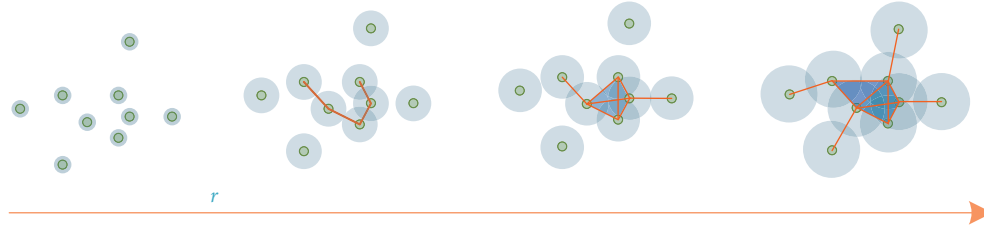


FIGURE 4: Vietoris-Rips filtration in persistent homology-based topological data analysis.

the normalized correlation coefficient (C-matrix) between electrodes and the other uses the level correlation distance as the normalized correlation coefficient (D -matrix). Euler entropy can be calculated according to the Betti numbers. In the process of Vietoris-Rips filtration, the Betti numbers change all the time, so we can restore an Euler entropy curve. The Euler entropies of brain networks for different values of ϵ are calculated, and it is a remarkable fact that Euler entropy has a negative peak with the change of ϵ . Since the ϵ value corresponding to the negative peak of Euler entropy is different between the clear and unclear situation, we further calculate the ϵ value when the negative peak of Euler entropy appears, which is taken as a phase transition point in this work. In the topological modeling of human brain structure, the phase transition point is shown by Euler entropy often represents a critical point change in brain activity [8].

2.4.6. Persistent Homology Analysis. Persistent homology is an algebraic topology methodology that counts the number of n -dimensional holes in a topological space, that is, Betti number. The Betti number of a generic topological space S is composed of β_0 , β_1 , and β_2 in this paper. β_0 is the number of connected components in S ; β_1 is the number of holes in S ; and β_2 is the number of voids in S . During the filtration, the time when a k -dimensional hole appears in the simplicial complex is recorded as T_{start} , while T_{end} is the time when the k -dimensional hole disappears. Accordingly, the k -dimensional Betti interval is defined by $[T_{\text{start}}, T_{\text{end}}]$, and the corresponding persistence barcode is its graphical representation of it [8, 26, 27]. On the other hand, persistent entropy (PE) provides a new entropy measure to extract the feature of topological space by persistence barcode. In this paper, $B = \{(x_i, y_i) | i \in I\}$ is set to the persistent barcode group associated with the filtration of topological space S , where i is a set of indexes (Figure 5). Accordingly, the persistent entropy H of the simplicial complex filtration is calculated by the following equation:

$$H = - \sum_{i \in I} p_i \log(p_i), \quad (6)$$

where $p_i = y_i - x_i / L$ and $L = \sum_{i \in I} (y_i - x_i)$. Moreover, H can be rescaled, and \hat{H} is treated as the persistent homology feature of the EEG data in this paper and expressed as follows:

$$\hat{H} = \frac{H}{\log \ell_{\max}}, \quad (7)$$

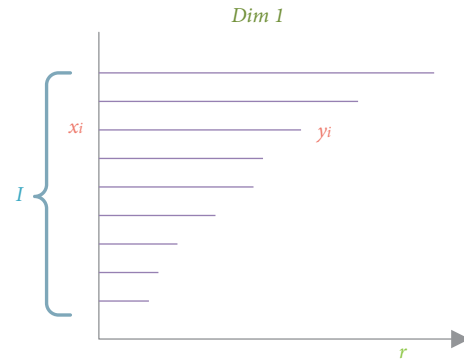


FIGURE 5: Persistent barcode group.

where ℓ_{\max} is the maximum interval in the considered persistent barcode group.

Topological patterns of the EEG data evoked by the RSD/GST pictures are constructed by Vietoris-Rips filtration, as shown in Figure 6. To examine the relationship between different brain regions and the perception of image shape and contour, the EEG mapping results are supplemented and drawn in Figure 6. When the subjects perceive irregularly distributed images, more brain regions are involved, with nonprominent features, but when they perceive ordered images, there will be clear reaction areas with more prominent features. Therefore, we hypothesize that vague cognitive goals make the task more difficult and lead to more mental activities. In addition, due to the intense brain activity observed in the frontal lobe, the frontal lobe's function needs further investigation. The frontal lobe is the physiological basis of the most complex mental activities. It is responsible for planning, regulating, and controlling human mental activities, which plays a vital role in human's advanced and purposeful behaviors. Figure 6 indicates the correlation between human perception of shapes and higher-level cognitive processes covering Gestalt patterns. These results verify our method's effectiveness in describing the intrinsic correlation between the EEG signals and shape cognition, and our method is closer to the actual biological response process.

In this regard, our preliminary Euler entropy analysis diagram is shown in Figure 7. The two types of calculated Euler entropy show the difference in their phase transition points. The red line represents the RSD in the state of unclear recognition, and the blue line represents GST that can recognize the outline of a triangle. It can be seen that the phase transition point of the GST sample will appear earlier

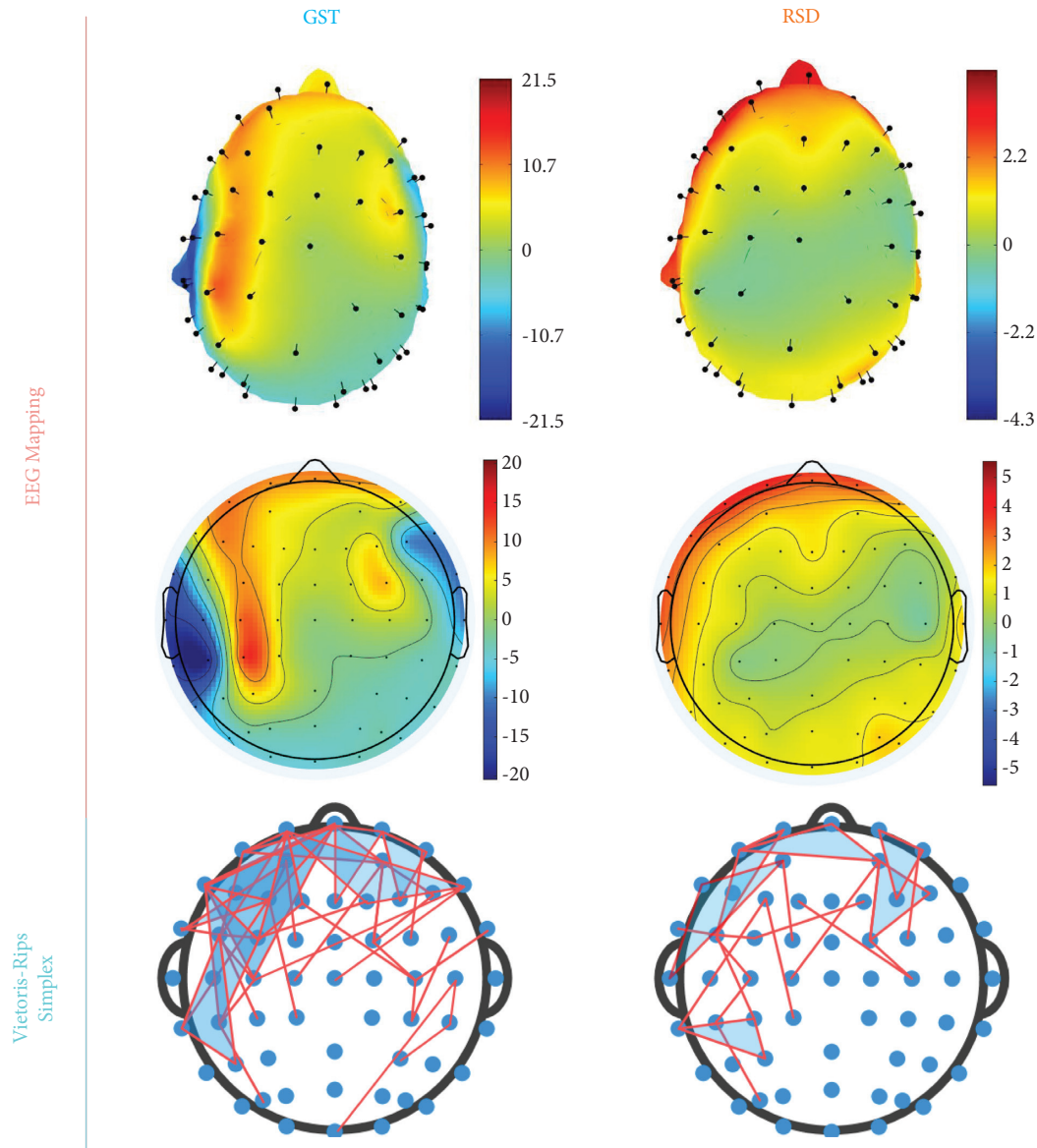


FIGURE 6: EEG mapping and the corresponding Vietoris–Rips simplex. Red indicates high levels of neuronal activation in EEG mapping.

than the RSD. The data result of the overall sample is shown in Figure 8.

3. Result and Discussions

It can be observed from Figure 8 that the phase transition point of GST trials of most samples is before the RSD, which intuitively shows the separability of the overall topological characteristics of the whole brain signal, so further continuous coherence analysis can be carried out on this basis.

To reduce the computation time of analyzing topological features, the change of persistent entropy of the subjects at different time latencies after the picture appearance is investigated first. There is a difference between the RSD and GST trials. Without being informed of the experiment's purpose, the subjects observed the RSD pictures first, which are disorderly and random. Accordingly, the overall EEG levels are shown in each trial. They were all at a certain level

in a relatively balanced manner, while regarding the GST pictures, it was intuitively reflected within 2 s, and the EEG level was stable afterward. Therefore, the EEG signals of two seconds after displaying the image are selected for the persistent homology analysis in this paper.

Table 1 shows the range, mean value, and maximum value of the distinguishing degree of the two types of cognitive behaviors investigated by adopting persistent entropy as the overall experiment's discrimination standard. The average distinguishing rates of each frequency band classified by C - and D -matrix were all greater than 70%, and the optimal distinguishing rates reached 90% and 85% for C - and D -matrix, respectively.

Figure 9 shows the respective performance and comparison of the persistent entropy obtained by the two different matrix calculation methods. The blue line represents the participants' response to GST, and the red line represents their response to RSD. Based on the statistical classification

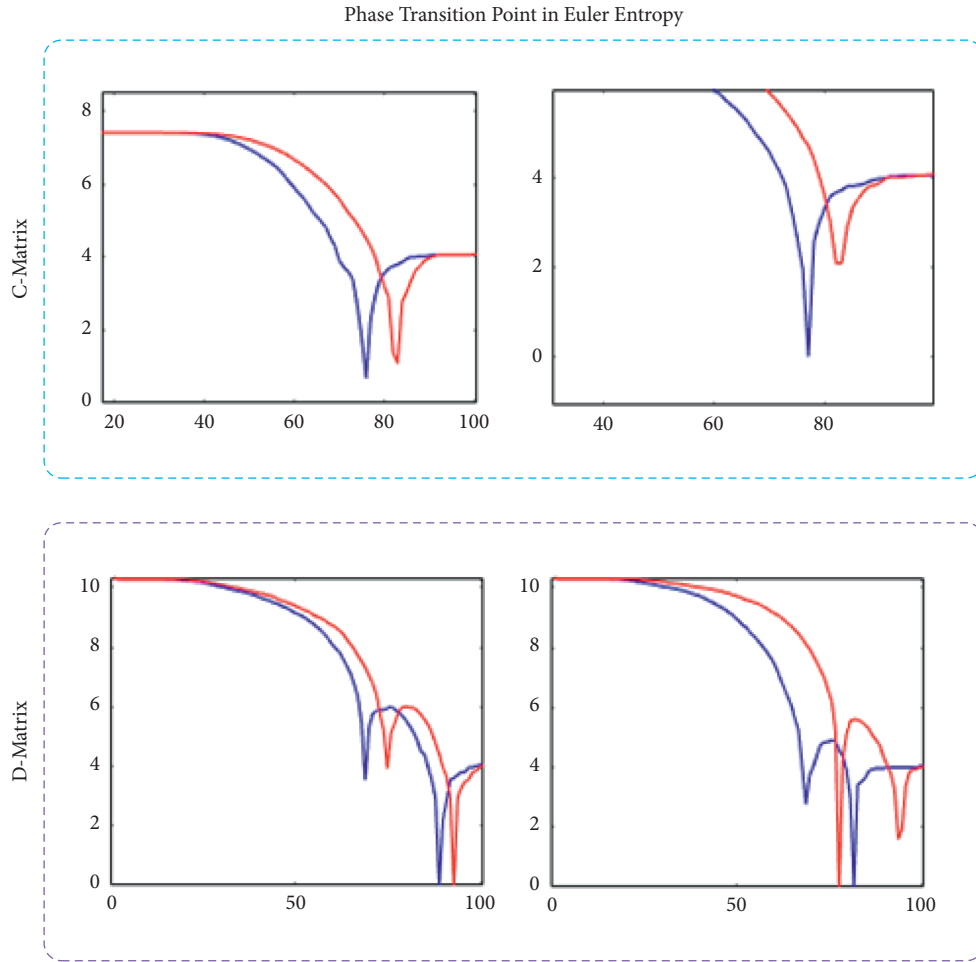


FIGURE 7: Phase transition points in Euler entropy curve.

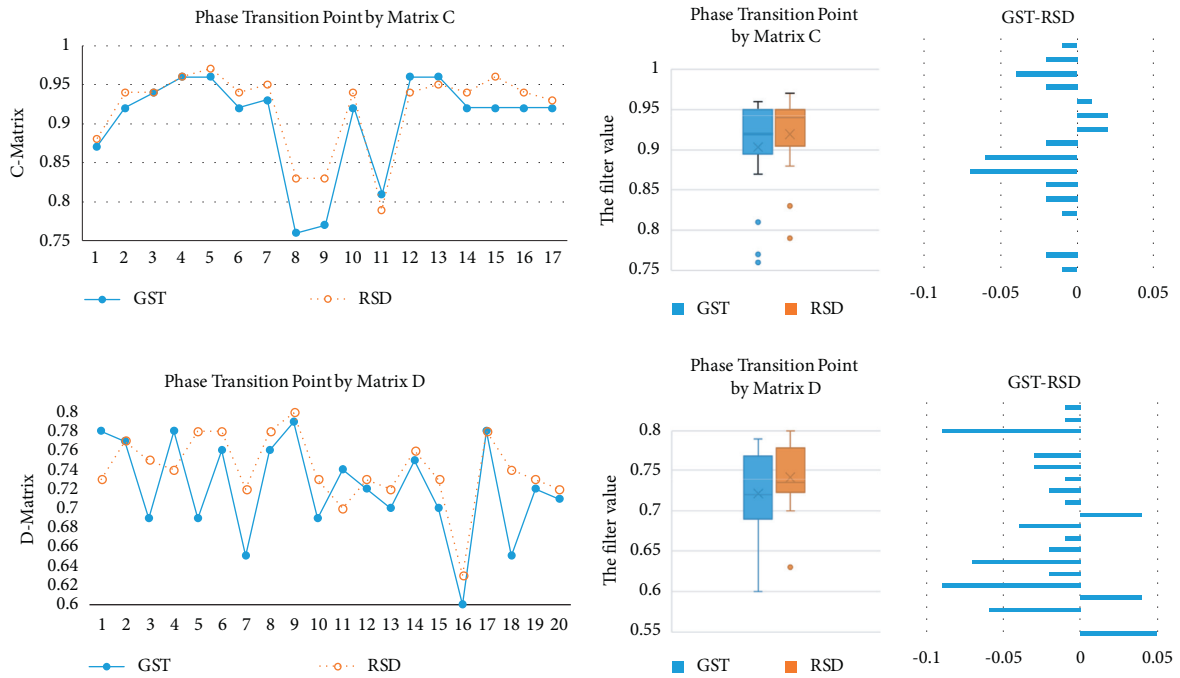


FIGURE 8: Comparison of phase transition points by C- and D-matrix.

TABLE 1: The table of distinguishing accuracy.

	Distinguishing rate range (%)	Average (%)	Max (%)
C-matrix	65–90	73.33	90
D-matrix	55–85	71.67	85

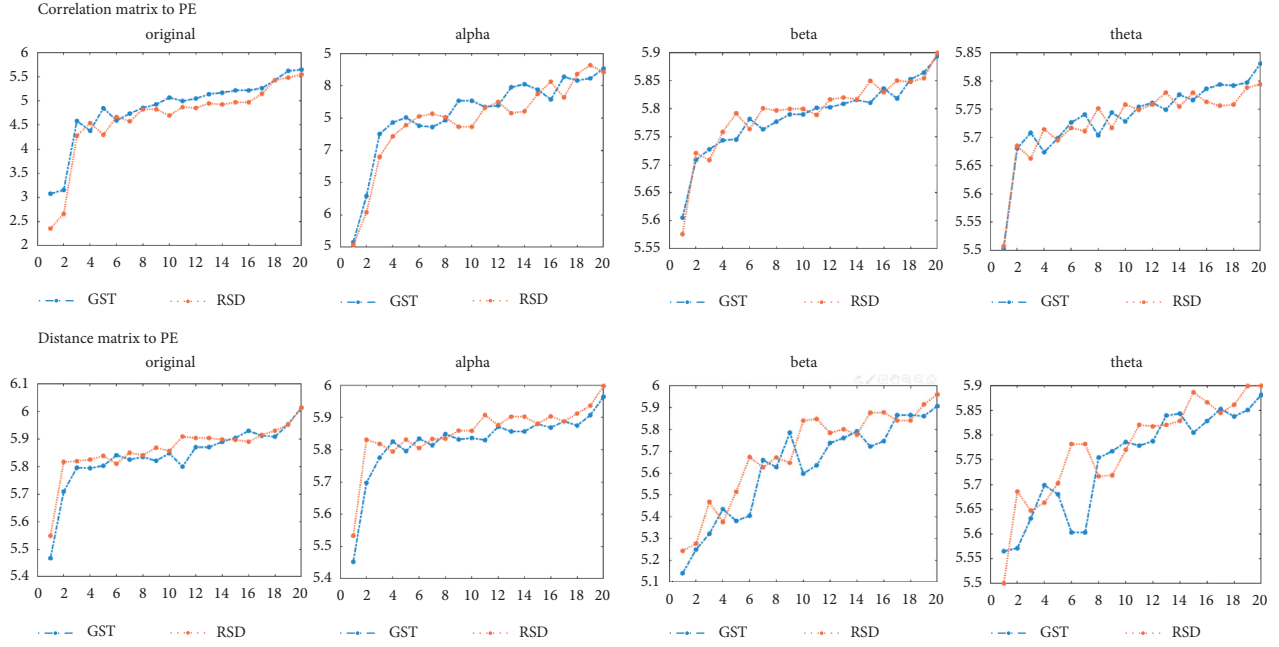


FIGURE 9: Comparison of persistent entropy calculated by C- and D-matrix (1).

of the 20 subjects and the frequency comparison, it is clear that both methods clearly describe the separation feature in terms of persistent entropy between the two topological patterns in the two types of brain cognition situations. We further draw a comparison chart of the GST and RSD values calculated by the correlation matrix as shown in Figure 10. According to Figures 9 and 10, almost all the persistent entropy values for the topological structure of the EEG signals induced by GST (blue) are higher than that induced by RSD (red); the opposite is the case when the persistent entropy is calculated by the distance matrix. From the measurement and analysis results, it is evident that the bands with significant differences are the α and θ bands, which are in line with the corresponding trend of the overall original signal. However, in the β band, the properties of the two types of methods are similar, and the results of the comparison of RSD and GST are similar as depicted in Figure 11.

As a summary, we have proposed a neurophysiological approach for cognitive assessment of the shape and contour of the Gestalt images via EEG. When the subjects perceive RSD images, compared with the GST image, more brain regions are involved in the cognition process. It can be understood that the human brain is in a state of randomness in this case. TDA is used to extract the physiological features of EEG signals induced by the shape contours. The results verify that the EEG data induced by the GST image are in the beta band, and the persistent entropy values obtained by the

two calculation methods are lower than that of the RSD image. The persistent entropy values in the α and θ bands and the overall 1–45 Hz band consistently show that $PE_{GST} > PE_{RSD}$ with the correlation matrix calculation method and $PE_{GST} < PE_{RSD}$ with the distance matrix calculation method.

Compared with the conventional neurophysiological methods based on evoked potentials (requiring a specific experimental paradigm), our approach provides a generalizable method that can extract the overall information from the whole brain signal, not just the characteristic performance. Our approach focuses not only on the brain response to external stimuli but also on the algorithms designed to normalize and extract numeric features that can be reliably classified and represent different cognitive perceptions. The algebraic topology is used to explain the coordination relationship between various neural regions in the human brain. This work can serve as an inspiration for the analytical approach to the collaborative work of complex neural networks. The dimensionality of the complex neural network model is reduced to one-dimensional persistent entropy to measure its characteristics.

Since this paper focuses on a specific case of Gestalt contour cognition, future research may extend to more analysis of different Gestalt contour cognitions and even color or content cognition and progressively try to leverage TDA to explain the cognitive process. Unlike the previous

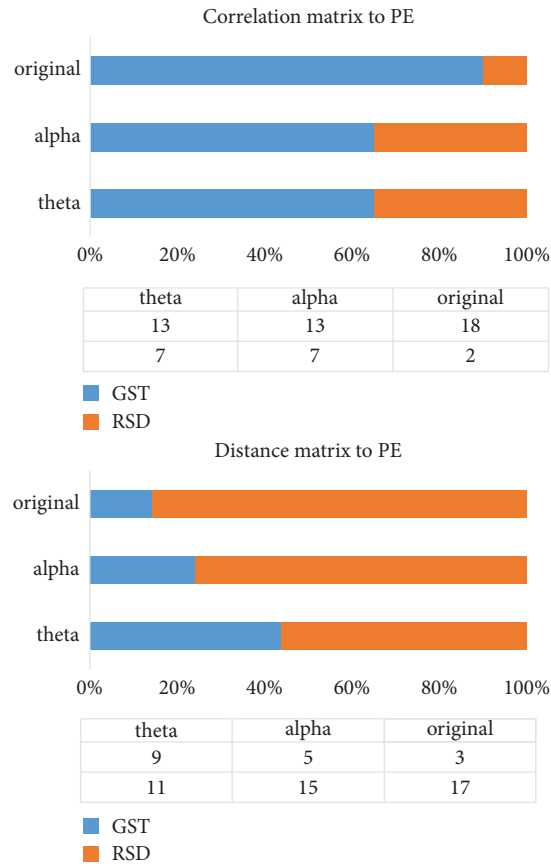


FIGURE 10: Comparison of persistent entropy calculated by C- and D-matrix (2).

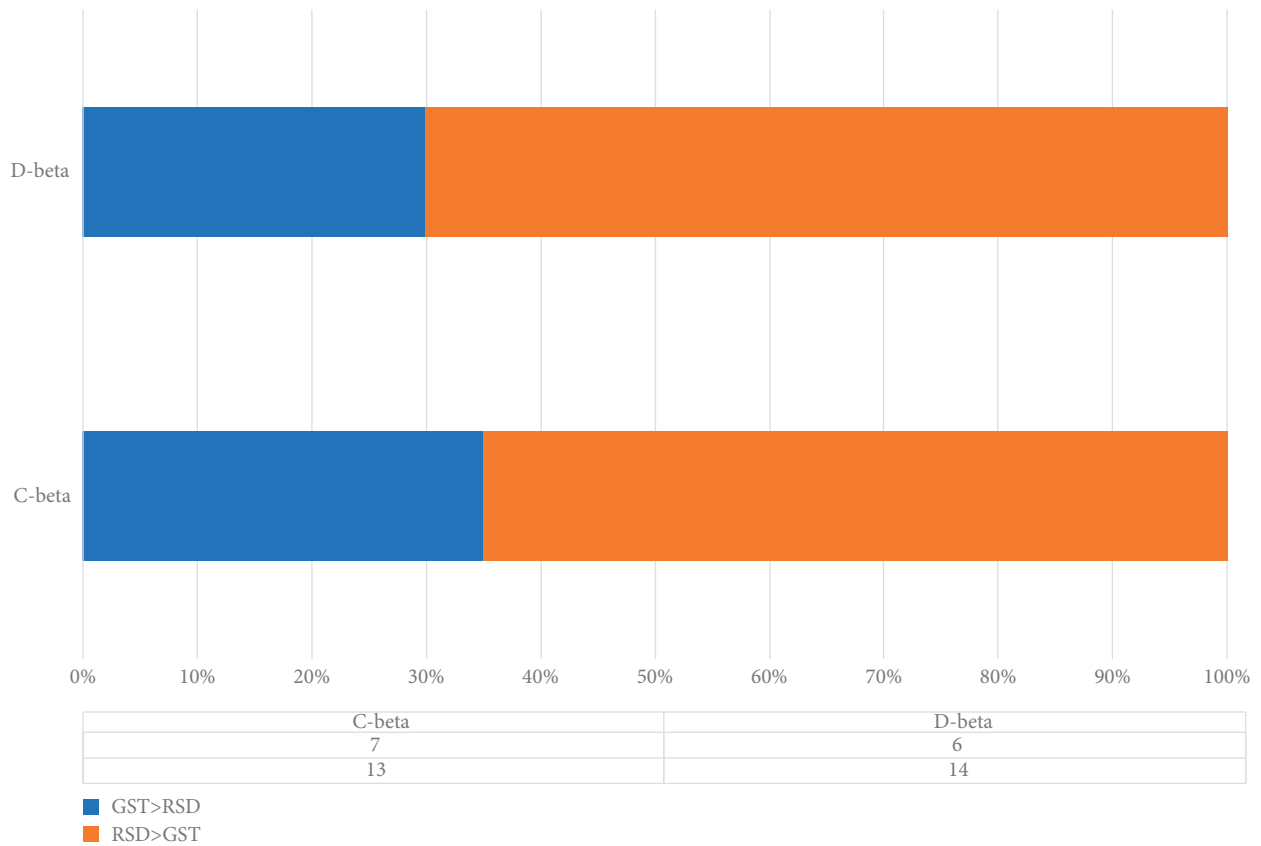


FIGURE 11: Beta bands.

EEG signal analysis experiments, this work implies the feasibility to interpret the human brain consciousness patterns in a divisible manner, and it is a preliminary exploration from feeling to consciousness. Furthermore, the deepening grasp of the brain neural network's linkage behavior in the brain response process from external stimuli to digital features may inspire us to build a new artificial neural network structure, which requires further research and experimentation.

4. Conclusions and Future Work

In this paper, we have proposed an approach physiological evaluation of contour cognition from EEG by using persistent homology of brain network and extracted its separable digital feature, persistent entropy (PE). Our approach has acquired cognitively related neural information via integrating the EEG collection with the traditional Gestalt psychology test procedure and obtained physiologically meaningful features of brain responses to different shape outlines by topological data analysis (TDA). The validation experiment results show that when subjects perceive chaotically distributed images, more brain regions are involved, but the level values are more average, and when they perceive ordered images, there will be clear reaction areas with more prominent features. The PE calculated by using two different EEG correlation feature extraction matrices is all separable. In the α , θ , and (1–45 Hz) bands, the overall performance is consistent, and the two types of calculations in the β band have reached a unified result of the calculated value and the classification situation. The above results can intuitively show that in some specific B2BC interaction scenarios, the transmission of a specific human brain nerve signal into a characteristic signal (PE) can be achieved.

The neurophysiological assessment process of Gestalt contour cognition is a preliminary study to explore human consciousness formation. The experimental results show that the conventional EEG signal can be digitized and converted into the matrix relationship between the electrode points, and then the Vietoris–Rips complex is constructed to use the topological calculation to express the characteristics. It encouragingly shows good separation, which provides a possibility for the development of B2BC.

At present, there still exist some limitations to be addressed, specifically in the following two aspects:

- (1) One is that noninvasive EEG acquisition equipment cannot completely restore the spatial location generated by electrical signals, which means that the accuracy of our topological reconstruction construction cannot restore the original signals generated by consciousness
- (2) Second, the use of algebraic topology is still in the preliminary stage, and more experiments are required to verify the robustness of the method.

The outlook for future work can be expanded from two dimensions of breadth and depth. The breadth is that there are many forms of conscious thinking because the project is an exploratory experiment, and we use contour recognition

as the starting point. Subsequent work can be developed to the consciousness analysis of more advanced cognitive behaviors, such as the calculation of simple mathematical problems, the judgment of the right and wrong of simple logic, and so on. The depth requires us to further enhance and strengthen this method on the basis of existing research. We can try to refine the research based on gender differences, brain region selection, more detailed trial segmentation, and frequency band selection to verify the robustness and reliability of the method proposed in this paper. The realization and gradual advancement of these tasks will lay a solid foundation for our future realization of brain-computer interconnection and brain-to-brain interconnection technology, and this is also an effective means to simulate and realize human intelligence. The analysis of human consciousness and thinking activities in this work also expands the breadth and depth of EEG analysis. The research in this area is still in the preliminary stage for the time being, and we provide enlightening significance for reference.

Data Availability

The data in this study are collected by our own experiment and are available from the first author (liuzs@zju.edu.cn) upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1804800, in part by the National Natural Science Foundation of China under Grants 61731002 and 62071425, in part by the Zhejiang Key Research and Development Plan under Grants 2019C01002 and 2019C03131, in part by Huawei Cooperation Project, in part by project sponsored by Zhejiang Lab under Grant 2019LC0AB01, in part by project sponsored by Ministry of Industry and Information Technology under Grant 2019-00891-2-1, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY20F010016.

References

- [1] S. Siuly, S. K. Khare, V. Bajaj, H. Wang, and Y. Zhang, "A computerized method for automatic detection of schizophrenia using EEG signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 11, pp. 2390–2400, 2020.
- [2] Y. Wang, H. Ombao, and M. K. Chung, "Topological data analysis of single-trial electroencephalographic signals," *Annals of Applied Statistics*, vol. 12, no. 3, pp. 1506–1534, 2018.
- [3] S. Scholler, S. Bosse, M. S. Treder et al., "Toward a direct measure of video quality perception using EEG," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2619–2629, 2012.
- [4] C. Liu, X. Ma, J. Wang et al., "Neurophysiological assessment of image quality from EEG using persistent homology of brain network," in *Proceeding of the 2021 IEEE International*

- Conference on Multimedia and Expo*, pp. 1–6, ICME, Shenzhen, China, July 2021.
- [5] K. Koffka, “Perception: an introduction to the gestalt-t,” *Psychological Bulletin*, vol. 19, no. 10, pp. 531–585, 1922.
 - [6] D. George, M. L’azaro-Gredilla, and J. S. Guntupalli, “From captcha to commonsense: how brain can teach us about artificial intelligence,” *Frontiers in Computational Neuroscience*, vol. 14, p. 97, 2020.
 - [7] A. Lavin, J. S. Guntupalli, M. L’azaro-Gredilla, W. Lehrach, and D. George, “Explaining visual cortex phenomena using recursive cortical network,” 2018, <https://www.biorxiv.org/content/early/2018/07/30/380048>.
 - [8] F. A. N. Santos, E. P. Raposo, M. D. Coutinho-Filho, M. Copelli, C. J. Stam, and L. Douw, “Topological phase transitions in functional brain networks,” *Physical Review*, vol. 100, no. 3, Article ID 032414, 2019.
 - [9] A. Subasi and E. Erçelebi, “Classification of EEG signals using neural network and logistic regression,” *Computer Methods and Programs in Biomedicine*, vol. 78, no. 2, pp. 87–99, 2005.
 - [10] V. Srinivasan, C. Eswaran, and N. Sriraam, “Approximate entropy-based epileptic EEG detection using artificial neural networks,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 3, pp. 288–295, 2007.
 - [11] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of Brain Network Analysis*, Academic Press, Cambridge, MA, USA, 2016.
 - [12] A. C. N. Chen, P. Rappelsberger, and O. Filz, “Topology of EEG coherence changes may reflect differential neural network activation in cold and pain perception,” *Brain Topography*, vol. 11, no. 2, pp. 125–132, 1998.
 - [13] M. Piangerelli, M. Rucco, L. Tesei, and E. Merelli, “Topological classifier for detecting the emergence of epileptic seizures,” *BMC Research Notes*, vol. 11, no. 1, pp. 1–7, 2018.
 - [14] D. Zhang, L. Yao, K. Chen, S. Wang, P. D. Haghighi, and C. Sullivan, “A graph-based hierarchical attention model for movement intention detection from EEG signals,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 11, pp. 2247–2253, 2019.
 - [15] R. Gonzalez-Diaz, E. Paluzo-Hidalgo, and J. F. Quesada, “Towards emotion recognition: a persistent entropy application,” in *Proceedings of the International Workshop on Computational Topology in Image Context*, pp. 96–109, Springer, Cham, 2019.
 - [16] N. K. Al-Qazzaz, M. K. Sabir, and K. Grammer, “Correlation indices of electroencephalogram-based relative powers during human emotion processing,” in *Proceedings of the 2019 9th International Conference on Biomedical Engineering and Technology*, pp. 64–70, Tokyo Japan, March 2019.
 - [17] N. Baker, G. Erlikhman, P. J. Kellman, and H. Lu, “Deep convolutional networks do not perceive illusory contours,” *Cognitive Science*, 2018.
 - [18] B. Kim, E. Reif, M. Wattenberg, S. Bengio, and M. C. Mozer, “Neural networks trained on natural scenes exhibit gestalt closure,” *Computational Brain & Behavior*, vol. 4, no. 3, pp. 251–263, 2020.
 - [19] A. Zomorodian and G. Carlsson, “Computing persistent homology,” *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.
 - [20] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, American Mathematical Society, Ann Arbor, MI, USA, 2010.
 - [21] R. Hartshorne, *Algebraic Geometry*, Springer Science & Business Media, Berlin, Germany, 2013.
 - [22] M. Petti, J. Toppi, F. Babiloni, F. Cincotti, D. Mattia, and L. Astolfi, “Eeg resting-state brain topological reorganization as a function of age,” *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 6243694, 10 pages, 2016.
 - [23] M. Rucco, R. Gonzalez-Diaz, M.-J. Jimenez et al., “A new topological entropy-based approach for measuring similarities among piecewise linear functions,” *Signal Processing*, vol. 134, pp. 130–138, 2017.
 - [24] A. Khalid, B. S. Kim, M. K. Chung, J. C. Ye, and D. Jeon, “Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology,” *NeuroImage*, vol. 101, pp. 351–363, 2014.
 - [25] M. Johansson, “The Hilbert transform,” Mathematics Master’s thesis, V”axj”o University, V”axj”o, Sweden, 1999.
 - [26] H. Adams, T. Emerson, M. Kirby et al., “Persistence images: a stable vector representation of persistent homology,” *Journal of Machine Learning Research*, vol. 18, pp. 1–35, 2017.
 - [27] Y. Wang, H. Ombao, and M. K. Chung, “Statistical persistent homology of brain signals,” in *Proceeding of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Brighton, UK, May 2019.

Research Article

Investigating Feature Ranking Methods for Sub-Band and Relative Power Features in Motor Imagery Task Classification

Samrudhi Mohdiwale ¹, Mridu Sahu ¹, G. R. Sinha ² and Humaira Nisar ³

¹National Institute of Technology Raipur, Raipur, India

²Myanmar Institute of Information Technology, Mandalay, Myanmar

³Universiti Tunku Abdul Rahman, Kampar, Malaysia

Correspondence should be addressed to G. R. Sinha; gr_sinha@miit.edu.mm

Received 28 May 2021; Revised 31 August 2021; Accepted 11 September 2021; Published 27 September 2021

Academic Editor: Daniel H.K. Chow

Copyright © 2021 Samrudhi Mohdiwale 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.

Interpreting the brain commands is now easier using brain-computer interface (BCI) technologies. Motor imagery (MI) signal detection is one of the BCI applications, where the movements of the hand and feet can be recognized via brain commands that can be further used to handle emergency situations. Design of BCI techniques encountered challenges of BCI illiteracy, poor signal to noise ratio, intersubject variability, complexity, and performance. The automated models designed for emergency should have lesser complexity and higher performance. To deal with the challenges related to the complexity performance tradeoff, the frequency features of brain signal are utilized in this study. Feature matrix is created from the power of brain frequencies, and newly proposed relative power features are used. Analysis of the relative power of alpha sub-band to beta, gamma, and theta sub-band has been done. These proposed relative features are evaluated with the help of different classifiers. For motor imagery classification, the proposed approach resulted in a maximum accuracy of 93.51% compared to other existing approaches. To check the significance of newly added features, feature ranking approaches, namely, mutual information, chi-square, and correlation, are used. The ranking of features shows that the relative power features are significant for MI task classification. The chi-square provides the best tradeoff between accuracy and feature space. We found that the addition of relative power features improves the overall performance. The proposed models could also provide quick response having reduced complexity.

1. Introduction and Background

Brain activates the sensory motor rhythm for virtual motor movements such as hand or feet; however, the actual motor movement of the body parts is not essential. The activation properties of brain correlate the activities with the motor movements, which help in different emergency situations to provide quick response in the system [1, 2]. Amyotrophic lateral sclerosis (ALS) is one of the serious diseases of brain, where the patient loses their control over the body and only the mind is active. The brain-computer interface (BCI) technology designed for motor imagery task can assist the patients in communication [3]. BCI technology enables the translation of brain commands of motor actions to read the brain signal and thus is considered as effective method for

providing faster and accurate response [4, 5]. The accuracy of BCI is very sensitive to internal and external noises, intersubject variabilities, and nonstationarity and nonlinearity of brain signal. Also, when BCI models are considered, the performance is limited by the algorithm simplicity.

The structure of brain consists of four lobes, which are frontal, parietal, occipital, and temporal lobe. Each lobe is responsible for a specific task. Frontal lobe is active when emotion, problem solving, speech, and movement related tasks are performed; and the parietal lobe is responsible for sensation, taste, speech, reading, and so forth. Occipital lobe is responsible for vision, visual stimuli, and interpretation. The temporal lobe is related to hearing, comprehension of language, and information retrieval. The movement related actions are performed by the motor cortex which is located

in back part of the frontal lobe almost in the center of brain. Thus, the signal received from the motor cortex from the frontal and parietal lobes is helpful in understanding and classifying motor imagery actions [6].

Numerous studies are reported in the literature to recognize and interpret the brain signals. The motor imagery commands are decoded using Fourier transform, short time Fourier transform, common spatial pattern and its variants, local mean decomposition, wavelet packet decomposition, and power spectral density [5, 7–11]. Wavelets are used for various applications such as denoising, feature extraction, and frequency sub-band categorization [12, 13]. Band power is also one of the widely used features. Brodu et al. [14] presented a comparative study on band power extraction for motor imagery (MI) task in which periodogram, autoregression, Butterworth filter, spectrogram, and Morlet wavelets are evaluated and wavelets are recommended for obtaining satisfactory results. Wang et al. [15] evaluated time frequency representation synthesis for spatial filters. Qin and He [16] developed wavelet and event-related desynchronization-based method for MI task classification, in which further comparison is based on the weighted energy difference of electrode pairs for MI task. Kim et al. [17] proposed an optimal channel based feature extraction via difference weighted power spectral density for MI task classification. The single session and session-to-session accuracies were evaluated for reliability of the proposed model. Tidare et al. [18] studied a single limb hand open and close movement by power measures. Linear regression and convolution neural network (CNN) are used to evaluate the performance and CNN was found to be outperforming. However, it still suffers from less accuracy, nearly 60% for the used dataset. This shows the chances of improvement via deep learning with higher variety of datasets. Mu-beta rhythms were proved efficient; however, to enhance the training feedback, SSVEP based hybrid BCI is proposed in [19]. Further studies also reported combination of EEG from motor and somatosensory cortex together for improving the performance of BCI [20]. Discriminative feature learning, sliding window common spatial patterns are some recent approaches used for MI task classification [21, 22]. EEG-Net with Temporary Constrained Sparse Group Lasso also proves its efficiency in MI task classification [23]. Akbulut et al. [24] proposed alpha and beta frequency power for MI task classification as the frequency represented as most responsible frequency of motor tasks. The performance evaluation is based on nearest neighbor, SVM, logistic regression, naïve Bayes, and decision tree classifiers. From these studies, decision tree and random forest are reported as the most widely used classifiers for MI task classification and, thus, in the current study, we have taken them into consideration. Deep learning is a widely accepted area nowadays, but the computational cost, complexity of model, and lack of sufficient data still create implementation challenges.

1.1. Motivation and Objectives. Motor imagery task classification is one of the open challenging tasks for which various methods are proposed. Common spatial pattern (CSP) and its

variants show the improved accuracy for MI task classification, but these are sensitive to noise; and the spectral and temporal characteristics of signals are neglected. Moreover, the variants are susceptible to channel specific data. To mitigate these limitations, frequency-based features are gaining popularity nowadays. Keeping in mind the popularity of frequency specific features, the current study aims to

- (1) Improve the performance of motor imagery task classification using frequency-based features
- (2) Analyze the relative power of frequency for classification of motor imagery task
- (3) Maintain the tradeoff between accuracy and feature space for minimal complexity of the model

The rest of the paper is organized as follows. Section 2 presents the methodology for the process adopted to improve the performance of motor imagery task classification. Section 3 provides the details of the results obtained from different classifiers. Section 4 provides the detailed discussion on the results and Section 5 concludes the paper with future aspects in the area.

2. Methodology

In this section, the details of the method incorporated to classify the motor imagery signal are described. First, the dataset used in the current work is described. Second, the features and feature extraction methods from brain signal used for MI task are discussed. Furthermore, feature ranking algorithms are used to evaluate the performance of newly added features. Figure 1 represents the overall flow of processing involved in MI task classification.

In Figure 1, the processing of work accomplishes on EEG based MI task dataset. The description of the dataset is provided in Section 2.1. EEG signals are preprocessed using Butterworth filter and variety of features extracted from the wavelets of different power bands which are alpha, beta, gamma, and theta. To get the most significant feature and evaluate the suitability of feature ranking method, three different feature ranking methods, namely, mutual information, chi-square, and correlation, are performed. Classification on all feature sets and ranked feature set has been done using listed classification techniques. Results are evaluated based on accuracy, precision, recall, and F1-score. The results are also compared with existing techniques on the same dataset. Further details on each block of Figure 1 are presented further in the section.

2.1. Dataset Description. The current study utilizes an open source dataset available for MI task classification, which has been accessed from BNCI Horizon 2020 website. In this dataset, cue grazed recordings of 10 subjects in a single session of 8 runs, each of 20 trials, are available. The subjects were asked to perform hand and feet movements as per the cue. Participants had the task of performing sustained (5 seconds) kinesthetic motor imagery (MI) of the right hand and of the feet each as instructed by the cue. Feedback was presented in form of a white colored bar graph. The length of

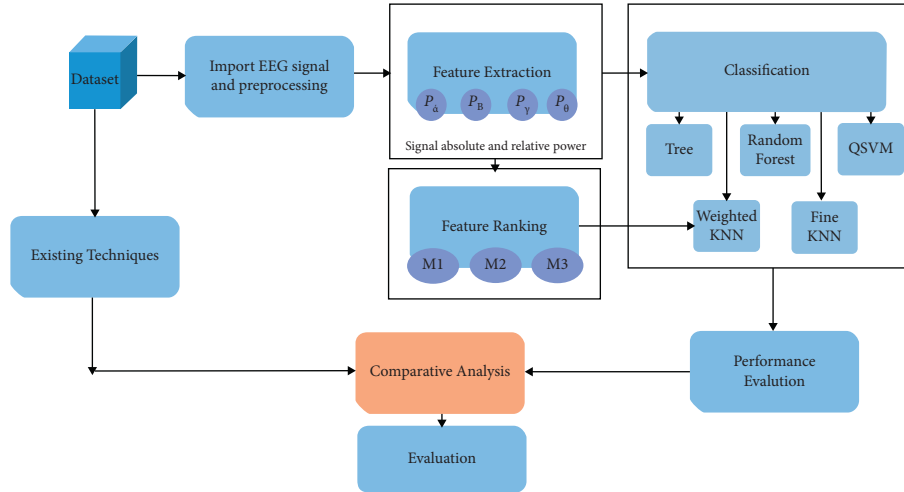


FIGURE 1: Flowchart of the proposed method.

the bar graph reflected the amount of correct classifications over the last second. EEG was measured with a biosignal amplifier and active Ag/AgCl electrodes at a sampling rate of 512 Hz. The electrodes placement was designed for obtaining three Laplacian derivations. To record the EEG signals, wet electrodes were fixed at central lobe, that is, C3, C4, and Cz [25, 26]. In this experiment, we have used 5-fold cross-validation throughout the work. Although there are different datasets available for MI task classification, this dataset has been chosen, since it provides a large number of data samples along with the different subjects available for recording with reduced intersubject variability issue.

2.2. Signal Preprocessing. The signals obtained from the source electrode contain artifacts such as undesired frequencies which are removed by preprocessing the data using Butterworth band-pass filter of 5th order having a passband frequency of 0.5 Hz and stopband frequency of 100 Hz. The filter transfer function is given in (1) [27]. Maximally flat response and uniform passband property make the filter more suitable for preprocessing of brain signals.

2.3. Feature Extraction. Feature extraction is an important step to reduce the dimensionality of a signal and simultaneously extract the important information. EEG recordings consist of large oscillations of different frequencies and various studies on brain oscillatory frequencies reflect the event-related synchronization and desynchronization in alpha, beta, and gamma rhythms [28]. Power spectral density (PSD) of these frequencies helps to analyze the impact of signal while performing MI task. Since the results from different studies suggest that only average power of these features is not sufficient to discriminate hand and feet movement, relative power and variance of the PSD are proposed in the current study for analysis.

To obtain the oscillations of different frequencies from EEG signal, discrete wavelet transform (DWT) of 5th level decomposition and “dB4” wavelet are used. The technique localizes the changes in frequency of signal over time and

thus both time information and frequency information are taken into consideration unlike CSP features with reduced computational complexity. In wavelet transform, the signal is downsampled by 2 and up to 5 levels, and, using downsampling, we get 5 detailed coefficients and 1 approximate coefficient. The 5th level DWT has been chosen because its decomposition provides the range of frequencies distributed similar to brain oscillatory frequencies. The “dB4” is chosen based on the effective results from different studies [17, 18, 29]. The frequency ranges for the coefficients are as follows: D1: 50–100 Hz (called noise and rejected), D2: 25–50 Hz (gamma), D3: 12.5–25 Hz (beta), D4: 6.25–12.5 Hz (alpha), D5: 3.125–6.25 Hz (theta), and A5: 0–3.125 Hz (delta, none of our interest).

In the current work, average power and relative and varied powers from each frequency band of interest are used as a new feature combination which can effectively distinguish the MI task. In Figure 2, it can be seen that the average as well as variance of each oscillatory frequency can play a major role. While the concept of relative power is considered here because the event-related synchronization and desynchronization between alpha, beta, and gamma rhythm are used for the MI task to happen, this will add extra efficient features for MI task classification. Average power of each frequency and variance of power distribution are obtained by using signal reconstruction corresponding to the wavelet coefficients of alpha, beta, and gamma. Pwelch function of MATLAB is used for the calculation of power spectral density, and the feature matrix is as follows: $[A_{11}, A_{12}, A_{11}, A_{13}, A_{14}, A_{15}, A_{16}, A_{17}, A_{18}, A_{19}, A_{110}, A_{111}, A_{111}]$, where $A_{11} - A_{14}$ denote average powers of alpha, beta, gamma, and theta band, respectively, $A_{15} - A_{17}$ denote relative powers of alpha to beta, gamma, and theta band, respectively, and $A_{18} - A_{111}$ denote varied powers of alpha, beta, gamma, and theta band, respectively. The algorithm used in the current study is presented as Algorithm 1.

The obtained feature matrix has dimension of 1×11 for each trial from the dataset. The overall feature matrix of size 1260×11 is obtained. To evaluate the significance of

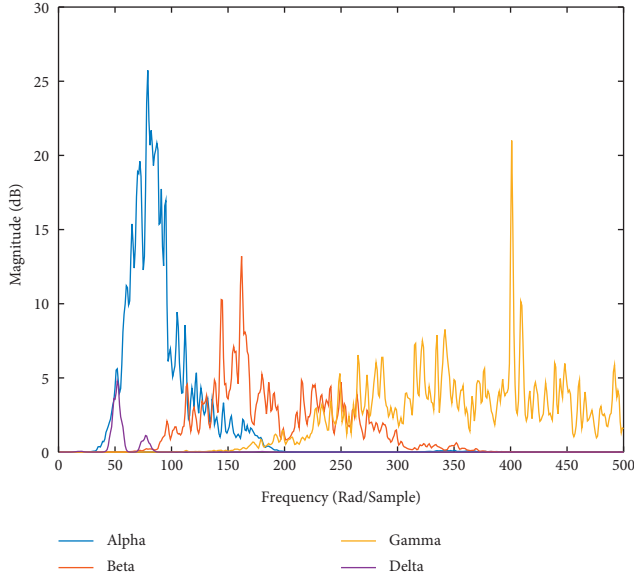


FIGURE 2: Power spectral density of different brain oscillatory frequencies.

features, feature ranking method is used and comparative analysis is carried out. In the next section, feature ranking method is described in detail.

2.4. Feature Ranking. Feature ranking methods help in evaluating the importance of the proposed features which are used in the work. These methods will provide rank of features based on different methods. Mutual information, chi-square (χ^2), and correlation are the most widely used feature ranking methods. In this study, these three techniques are used to find the most as well as least important features. The details of the method are discussed below.

2.4.1. Mutual Information. Mutual information provides the measure of dependency between features and can be obtained by

$$I(A, B) = \sum_A \sum_B p(A, B) \log \frac{p(A, B)}{p(A)p(B)}, \quad (1)$$

where A and B are two different features, respectively; $p(A, B)$ is the joint probability; $p(A)$ and $p(B)$ are individual probabilities and $I(A, B)$ denotes the mutual information between two features. The higher the mutual information is, the higher the dependency is; hence, features having higher mutual information will be ranked higher than others [30, 31].

2.4.2. Chi-Square Method. Chi-square method of feature selection ranks the features based on the dependency of features on the respective class. We are interested in features which are highly dependent on the class and this method gives the higher rank to the feature which is more dependent on class than others. These feature frequencies are calculated from each sample. Null hypothesis for the test is formulated

as the features are highly dependent with an alternative hypothesis as the features are independent. The value of χ^2 is calculated by using the following formula [30, 32]:

$$\chi^2 = \frac{(F_O - F_E)^2}{F_E}, \quad (2)$$

where F_O is the observed frequency of dependent features and F_E is the expected frequency of the dependent features. Alpha (confidence interval) is chosen as 0.05.

2.4.3. Correlation Method. Correlation measure provides a method to identify highly correlated features of the data. Higher correlation signifies the lesser generalizability of the model. Hence, these features need to be removed to reduce the dimension and to improve the generalizability of the selected classifier. The correlation is calculated using the following formula [30, 33]:

$$r = \frac{k(\sum AB) - (\sum A)(\sum B)}{\sqrt{[(\sum A^2) - (\sum A)^2][k((\sum B^2) - \sum B)]}} \quad (3)$$

where A and B are two features from the set of features and r denotes the correlation coefficient between features. The correlation between more than two features can be visualized in heat map. In this map, highly correlated features are darker, while the features having lesser correlation show less intensity of the color. From the heat map, one can rank the highly positively or negatively correlated features [30, 33].

2.5. Classification and Comparative Analysis. The feature matrix obtained using the method is fed into different classifiers to evaluate the performance of features. Five well-known classifiers, namely, decision tree (DT), fine k-nearest neighbor (KNN), weighted KNN (WKNN), quadratic support vector machine (QSVM), and random forest, are used for classification. The results obtained from classification are shown and discussed in Section 3. The results obtained after classification are analyzed based on the classification accuracy of classifier, precision, recall, and F1-score [34]. The comparative analysis of the current study with different existing approaches is also presented in the next section.

3. Results and Discussion

The experiment was performed on a system configuration of Intel Core i5 processor and 8 GB RAM. Open source dataset of motor imagery task classification is evaluated with the proposed approach and compared with different existing approaches. The results obtained from the current study are presented in this section.

3.1. Results of Different Classifiers without Feature Ranking. In this study, five classifiers are used for classification of MI tasks and the results obtained are shown in Table 1. Five of the most widely used classifiers are chosen for the study, which are decision tree, fine KNN, weighted KNN, quadratic SVM, and random forest.

```

Input: EEG Signal for each trail, Class Label No. of Subject =  $n$ . No. of trails =  $T$ . No. of Segment =  $m$ 
Output: Accuracy, precision, recall, F-Score
for subject  $i = 1 : n$ 
  for segment  $j = 1 : T$ 
    for trail  $k = 1 : T$ 
      Data = Import data from each trail,
      Pro_data = Preprocessing using Butterworth filter of band [0.5–100 HZ] for Data ( $k$ )
      Features = Wavelet (Processed data with dB 4,5 level)
      calculate avg power, Variance of power, Relative power to alpha;
      Feature_set = Feature, Class ( $k$ );
    end
    Updated_Feature = Update feature set after each trail by concatenation;
  end
  Updated_Feature_Final = Update feature set after each segment by concatenation;
end
  Confusion Matrix = Function_Classifier (Updated_Feature_Final, Cross validation ( $k$  times));
Accuracy, Precision, Recall, F-Score = [Obtain from confusion matrix];
  Calculate Ranks of features from Mutual Information, chi-square, Correlation and Repeat the steps to calculate the matrix and Accuracy

```

ALGORITHM 1: Algorithm for the proposed approach.

TABLE 1: Classification accuracy of different classifiers.

Subjects	Classification accuracy (in %)				
	Tree	Fine KNN	Weighted KNN	Quadratic SVM	Random forest
S01	89.1	92.9	94.3	65.3	89.4
S02	87.3	91.9	92.8	70.5	91.1
S03	89.5	92.9	94.2	71	91.34
S04	88.2	90.6	93.6	67.3	90.625
S05	85.5	91.8	93.1	66.1	86.77
S06	87.7	90.2	93.1	58.4	90.86
S07	88.7	89.1	91.6	69	87.01
S08	86.7	91.3	92.9	64.4	89.66
S09	92	92.7	95.2	73.7	92.78
S10	87.7	92.9	94.3	67.5	91.34
AVG	88.24	91.63	93.51	67.32	90.0885

Bold letters show the maximum classification accuracy of the classifier.

Decision tree takes the decision based on experience by splitting the data and it works like a human brain. KNN is based on nearness criteria; that is, the lower the distance, the higher the chances of data to lie in that class. SVM creates a hyperplane to classify the objects. Random forest consists of multiple small decision trees on a random subset of data, and each tree will act as an expert to take the decision of split [34].

From the results, it is clear that, except SVM, all classifiers provide competitive accuracy. The highest accuracy obtained from the weighted KNN classifier is 93.51%.

Precision, recall, and F1-score measures are also important criteria while evaluating the performance of classifiers. The higher the precision is, the lower the number of false positive errors committed by the classifier is. Classifiers with large recall have very few positive examples misclassified. F-measure represents a harmonic mean between recall and precision. A high value of F-measure ensures that both precision and recall are reasonably high [34]. The precision, recall, and F1-scores of different classifiers are shown in Figure 3.

In Figure 3, it can be seen that, for values of precision, recall, and F1-score for classifiers ranging from 0.75 to 0.95, in decision tree, fine KNN, and random forest, we get outlier values; however, weighted KNN and QSVM does not have outliers. In weighted KNN, maximum values are above the 1st quartile range and have significant score of more than 0.85 which can be termed as good precision, recall, and F1-scores for MI task classification. Fine KNN also shows better values but has less accuracy than weighted KNN. Hence weighted KNN is used further in the study.

Feature ranking methods are evaluated for analysis in order to understand the role of power features. The detailed results on the ranking method are provided in the subsequent section.

3.2. Results of Different Classifiers with Feature Ranking Methods. Analysis of features is essential and crucial step to obtain the dominant features and reduce the dimensionality. As feature ranking can provide the relevant set of features, it

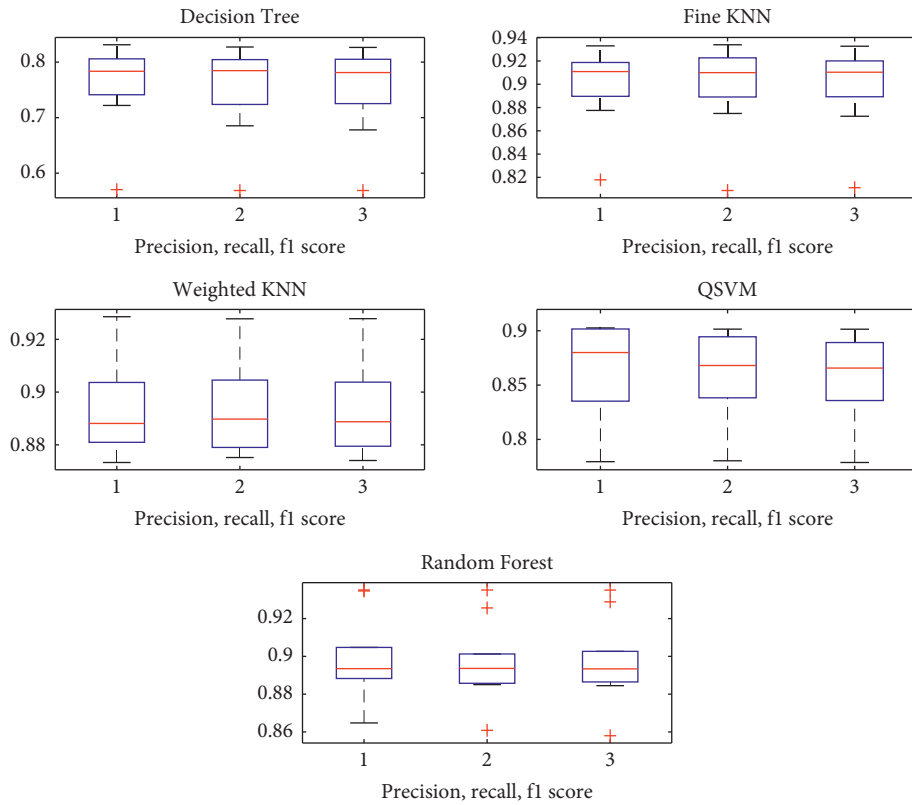


FIGURE 3: Precision, recall, and F1-scores for different classifiers.

also helps to reduce the overfitting and improve the generalizability of the classification model. This section discusses the results of different methods of feature ranking in detail.

3.2.1. Results of Feature Ranking with Mutual Information. Mutual information is a very prominent method of feature ranking. The detailed method was discussed in Section 2.4.1. To obtain the results of feature ranking, the method is evaluated 10 times and the ranking obtained after each iteration is stored. Mutual information shows the dependency of features with respect to class. The higher the mutual information, the higher the significance of the feature.

Top 6 features are selected out of 11 which are obtained from 10 iterations for all ten subjects. Figure 4 represents the ranking of features obtained with the mutual information method. From Figure 4, it is clear that features {5, 7, 8, 10, 11} are the most valuable features, out of which 5 and 7 are relative power features and the others are varied power features. It can be seen that mutual information method of feature selection selects the variety of features containing alpha, theta, and gamma band.

3.2.2. Results of Feature Ranking with Chi-Square Method. Chi-square test is performed for the ranking of features. The higher dependence of features on the class is preferred for ranking. Chi-square test rankings for 11 features are shown in Figure 5.

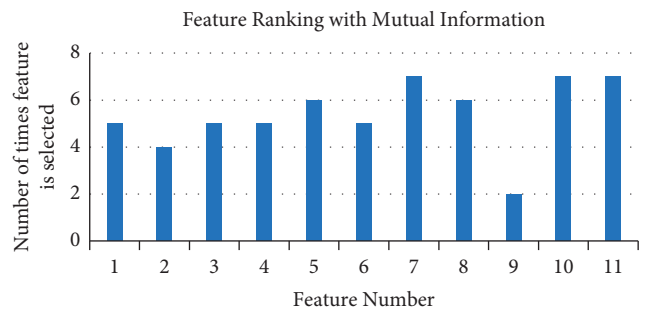


FIGURE 4: Feature ranking with mutual information.

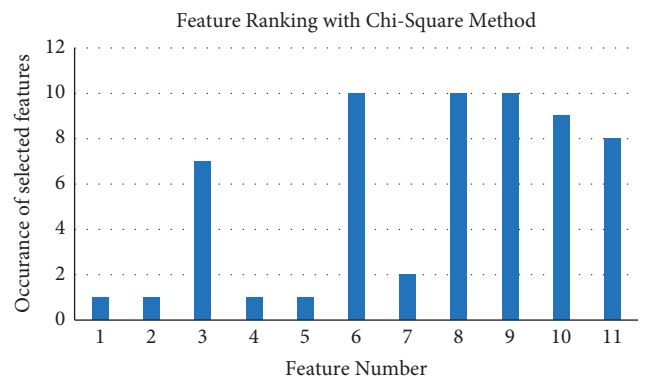


FIGURE 5: Feature rankings with chi-square method.

Chi-square method is another popular approach of feature selection. This method focuses on two values named F-Score and P values. The higher F-Score and the lower P values are the most significant features. In the current work, the same procedure is followed and top five features out of eleven are selected by the rank provided by chi-square method. The feature selection is performed for 10 iterations and the results are presented in Figure 5. Top five features are selected based on occurrence of features in multiple iterations. Feature set {6, 8, 9, 10, 11} is selected as most significant feature for further analysis.

3.2.3. Results of Feature Ranking with Correlation Measure. The correlation method is another measure for feature ranking, which provides information of mutually correlated features. The heat map for the features of the ten subjects is shown in Figure 6.

In the heat map, darker color shows higher correlation, while lighter color has less correlation among the features. The correlated pairs obtained after 10 iterations of the correlation feature selection module are shown in 7.

Figure 7 shows the correlated pairs of feature. The higher the correlation, the lower the importance of feature; hence, feature {3, 7, 8, 9, 10, 11} can be discarded and accuracy can be calculated among the rest of the features for further analysis. From this, it can also be seen that there is no correlation among the features of relative power, that is, feature {5, 6}. This represents the significance of relative power features with respect to the correlation among them.

3.2.4. Overall Analysis of Feature Ranking Method on Accuracy. In this section, the overall analysis of the impact of relative and varied power on accuracy for MI task classification is presented. The accuracy has been calculated by considering the best features obtained from feature ranking method. The results are shown in Table 2.

When all features are taken into consideration, the average accuracy of 93.51% has been obtained, which is a significant result. We analyze the importance of relative power features by calculating the accuracy without considering them. The results show that, for six out of ten subjects, the accuracy is reduced and, for the rest of the subjects, there is a slight change in the accuracy, nearly ± 0.2 to 0.6%, which is very small. Further, in the analysis, the best features from different ranking algorithms are selected based on the rank and the rest of the features are dropped out. The results for dropout of those features do not have more impact on the classification accuracy. In the subsequent section, a comparative analysis of different existing models and feature ranking methods is presented.

3.3. Comparative Analysis. To compare the proposed method with existing models, a comparative analysis is presented in this section. The results for comparison are based on the research papers using the same datasets [25, 26]. The proposed approach is compared with the most widely used feature, that is, CSP for MI task classification.

The comparative results are shown in Figure 8. From the figure, it is clear that the average and varied power features with RF classifier outperform the CSP feature for most of the subjects. The figure also represents that the performance of CSP method is varied with the subjects, but our proposed approach does not have this issue. This ensures that the performance of the proposed approach mitigates the challenge related to intersubject variability.

Apart from the results presented in this dataset, some other studies reported performance for a few subjects. Sahu et al. reported the average performance of 60.5% for 3 subjects using principal component analysis [35]. In another study, M. Sahu and S. Shukla used a feature selection approach for improved classification of MI signals. In this study, an average accuracy of 58.25% for 4 subjects is reported [36]. Kumar and Sahu proposed PSO based analysis for MI task classification and this resulted in the highest accuracy achieved which is 68.75% [37]. To analyze the performance of feature ranking approach, we further evaluate the average loss of features and accuracy for different feature ranking approach.

Figure 9 represents the loss or reduction in accuracy for different methods. If we select top 5 features from the feature set and calculate the accuracy, the variation in accuracy with existing accuracy of particular subject is termed as loss in accuracy. The higher the loss, the lower the significance of method. Correlation method shows the highest loss in accuracy with respect to accuracy of all the features. Mutual information method shows significant loss in accuracy, while chi-square method provides minimum loss in accuracy. So, chi-square method can be used to maintain the tradeoff if higher complexity is the issue.

Further, in this study, to know the effect of dropout features with respect to accuracy, Figure 10 highlights the change in accuracy.

From Figure 10, it is clear that the highest accuracy is obtained by considering all eleven features. In mutual information, top six features are selected and accuracy of 93% is achieved, whereas, in chi-square method, top five features are selected and accuracy of 93.5% is obtained. In correlation method, top 5 features are selected but accuracy drastically reduces to 91.5%. Hence, it can be concluded that, in mutual information method and chi-square method, the tradeoff between feature space and accuracy is maintained.

4. Discussion

In the current study, the impact of relative power and variance of power is taken into consideration for classification of MI tasks. Signals obtained from brain are taken using electrodes placed on the motor cortex; hence, the study is based on frequency and power of signal from the motor cortex, that is, C3, C4, and Cz and nearby electrodes. To analyze the impact, different strategies are adopted. Ten subjects are considered for the evaluation of the proposed method for MI task classification. The results show that the proposed approach outperforms CSP based approach for classification; the reason behind this is the nature of

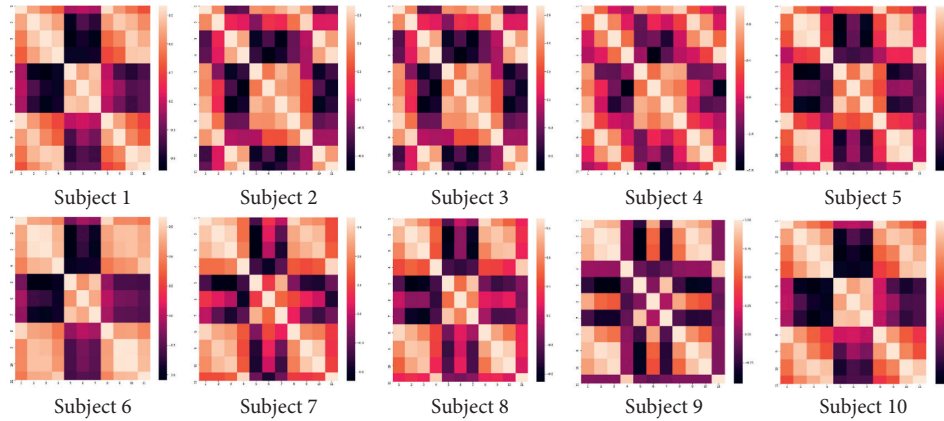


FIGURE 6: Heat map of all ten subjects for correlated and uncorrelated features.

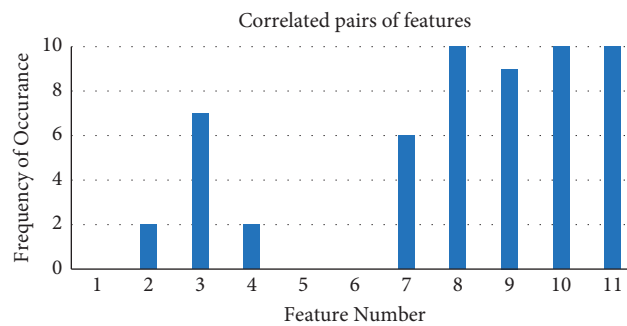


FIGURE 7: Correlated pairs of features.

TABLE 2: Analysis of accuracy (in %) for impact of relative and varied power features.

Subjects	Accuracy with all features	Accuracy without relative power features	Accuracy with best features from mutual information method	Accuracy with best features from chi-square method	Accuracy with dropping out correlated features
S01	94.3	94.9	93.8	94.2	92.4
S02	92.8	93	93.7	93.1	92.5
S03	94.2	93.3	93.2	94.2	90.4
S04	93.6	91.9	92.1	93.6	91.7
S05	93.1	92.5	92.6	93.1	91.8
S06	93.1	92.7	92.3	93.1	90.1
S07	91.6	91.4	91.7	91.6	89.7
S08	92.9	92.6	92.2	93.01	90.6
S09	95.2	95.4	95.4	95.2	91.7
S10	94.3	94.9	93.8	94.3	92.1

Bold letters show the maximum classification accuracy of the classifier.

frequency-based features which allows holding significant characteristics for each task efficiently. Overall average accuracy obtained in the study is 93.51%, which is consistent for all subjects, not like other approaches having higher accuracy for a few subjects. As an analysis of feature ranking method, the mutual information-based method delivers an average classification accuracy of 93.05% for the feature set {5, 7, 8, 10, 11}. In this method, the relative power features and varied power features show their significance with the slightest loss in accuracy. In the chi-square method, the analysis provides accuracy of 93.4% on the feature set {6, 8, 9,

10, 11}. In this method, the dominance of varied power feature is indicated, but it also shows the contribution of relative power feature of alpha to gamma ratio (6th feature). In correlation-based method, again the feature {5, 6} shows the least correlation. Hence, these proposed features are significant not only in the aspect of accuracy enhancement but also in different scenarios considered having mutual information and correlation of features as major concerns. The analysis also shows that the combination of features exhibits the tradeoff between accuracy and complexity of feature space when complexity is a major concern.

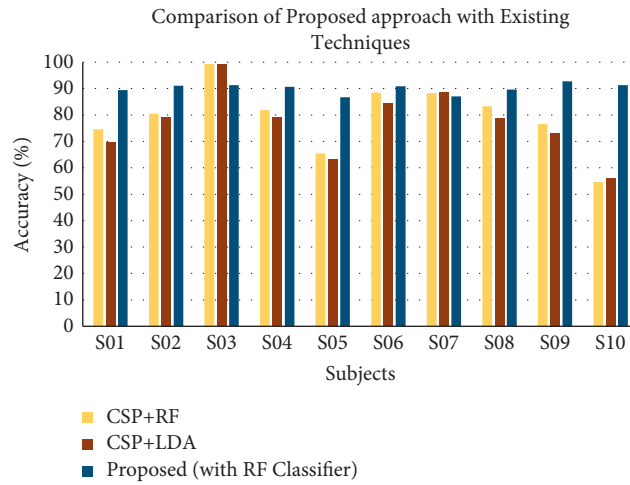


FIGURE 8: Comparison of the proposed approach with the existing approaches.

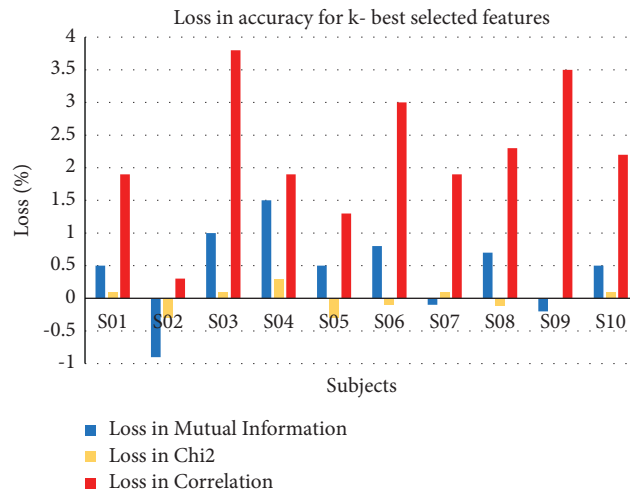


FIGURE 9: Average loss in accuracy for implemented methods of feature ranking.

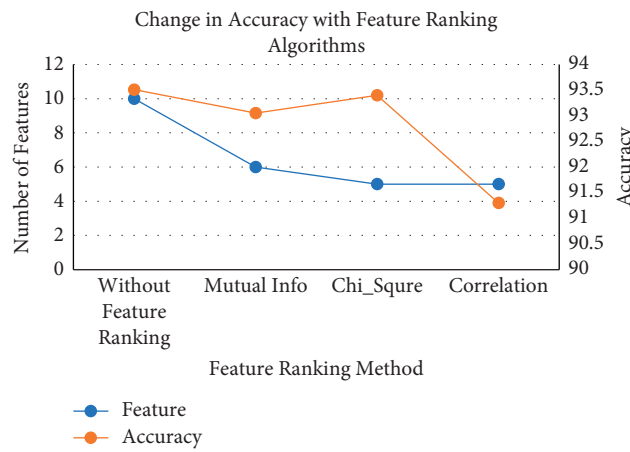


FIGURE 10: Change in accuracy with feature ranking algorithms.

5. Conclusion

The improvement in motor imagery based BCI is important in its sophisticated model and analysis. In the current study, MI signals are classified by the power of alpha, beta, and gamma frequencies along with the relative power. Five different classifiers are adopted to choose the better results from the variety of classifiers. The results of the classification suggest that all classifiers have competitive results but, based on accuracy, weighted KNN outperforms all the other classifiers. The proposed approach also outperforms the CSP-based feature extraction method for the given dataset. The proposed approach not only resolves the problem to maintain the tradeoff and complexity but also mitigates the intersubject variability problem. To judge the significance of new added features, feature ranking method is presented. Based on the mutual information, chi-square, and correlation methods of feature ranking, results are calculated and compared. The results of the feature ranking method suggest that when feature space and time complexity are the concerns, then chi-square method outperforms other feature ranking methods with reduction of 45% feature space. However, when the accuracy of the method is concerned for BCI, the addition of relative power feature improves the overall performance of the system. Future era will be based on BCI, which flourishes the scope of further research to meet the challenges of electrode placements and selection, signal to noise ratio improvement, and device dependency reduction.

Data Availability

Publicly available datasets are used in this research, which can be accessed via <http://bnci-horizon-2020.eu/database/data-sets>.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [2] E. C. Leuthardt, G. Schalk, J. R. Wolpaw, J. G. Ojemann, and D. W. Moran, "A brain-computer interface using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 1, no. 2, pp. 63–71, 2004.
- [3] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: initial tests by ALS patients," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 538–548, 2006.
- [4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.
- [5] S. Vaid, P. Singh, and C. Kaur, "EEG Signal Analysis for BCI Interface: A Review," in *Proceedings of the 2015 Fifth International Conference on Advanced Computing & Communication Technologies*, pp. 143–147, IEEE, Haryana, India, February 2015.
- [6] N. Kamel, "The Fundamentals of EEG Signal Processing," in *EEG/ERP Analysis*, CRC Press, Boca Raton, FL, USA, 1st edition, 2014.
- [7] M. R. Lakshmi, T. v Prasad, and D. V. C. Prakash, "Survey on EEG signal processing methods," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 1, 2014.
- [8] F. Lotte, "A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, Article ID 031005, 2018.
- [9] R. Chaudhari and H. J. Galiyawala, "A review on motor imagery signal classification for BCI," *Signal Processing: An International Journal*, vol. 11, no. 2, pp. 16–34, 2017.
- [10] J. Wang, Z. Feng, N. Lu, L. Sun, and J. Luo, "An information fusion scheme based common spatial pattern method for classification of motor imagery tasks," *Biomedical Signal Processing and Control*, vol. 46, pp. 10–17, 2018.
- [11] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proceedings of the 2007 3rd International IEEE/EMBS Conference On Neural Engineering*, pp. 204–207, IEEE, Kohala Coast, HI, USA, May 2007.
- [12] Y. S. Bahendwar and G. R. Sinha, "Efficient algorithm for denoising of medical images using discrete wavelet transforms," in *Proceedings of the 17th International Conference Computational and Mathematical Methods in Science and Engineering*, pp. 10–12, Cádiz, Spain, January 2015.
- [13] A. Devil, A. Misal, and G. R. Sinha, "Performance analysis of DWT at different levels for feature extraction of PCG signals," in *Proceedings of the 2013 Annual International Conference on Emerging Research Areas and 2013 International Conference on Microelectronics*, pp. 1–5, IEEE, Kanjirapally, India, June 2013.
- [14] N. Brodu, F. Lotte, and A. Lécuyer, "Comparative Study of Band-Power Extraction Techniques for Motor Imagery Classification," in *Proceedings of the 2011 IEEE Symposium On Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pp. 1–6, IEEE, Paris, France, April 2011.
- [15] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns," *Clinical Neurophysiology*, vol. 115, no. 12, pp. 2744–2753, 2004.
- [16] L. Qin and B. He, "A wavelet-based time-frequency analysis approach for classification of motor imagery for brain-computer interface applications," *Journal of Neural Engineering*, vol. 2, no. 4, p. 65, 2005.
- [17] C. Kim, J. Sun, D. Liu, Q. Wang, and S. Paek, "An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI," *Medical, & Biological Engineering & Computing*, vol. 56, no. 9, pp. 1645–1658, 2018.
- [18] J. Tidare, M. Leon, N. Xiong, and E. Astrand, "Discriminating EEG spectral power related to mental imagery of closing and opening of hand," in *Proceedings of the 2019 9th International IEEE/EMBS Conference On Neural Engineering (NER)*, pp. 307–310, IEEE, San Francisco, CA, USA, March 2019.
- [19] T. Yu, J. Xio, F. Wang et al., "Enhanced motor imagery training using a hybrid BCI with feedback," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 7, pp. 1706–1717, 2015.
- [20] L. Yao, X. Sheng, D. Zhang et al., "A stimulus-independent hybrid BCI based on motor imagery and somatosensory attentional orientation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1674–1682, 2017.

- [21] L. Yang, Y. Song, K. Ma, and L. Xie, "Motor imagery EEG decoding method based on a discriminative feature learning strategy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, 2021.
- [22] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.
- [23] X. Deng, B. Zhang, N. Yu, K. Liu, and K. Sun, "Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces," *IEEE Access*, vol. 9, pp. 25118–25130, 2021.
- [24] H. Akbulut, S. Güney, H. B. Çotuk, and A. D. Duru, "Classification of EEG signals using alpha and beta frequency power during voluntary hand movement," in *Proceedings of the 2019 Scientific Meeting On Electrical-Electronics & Biomedical Engineering And Computer Science (EBBT)*, pp. 1–4, IEEE, Istanbul, Turkey, April 2019.
- [25] D. Steyrl, R. Scherer, F. Oswin, and R. M. Gernot, "Motor imagery brain-computer Interfaces: random forests vs regularized LDA - non-linear beats linear," in *Proceedings of the 6th International Brain-Computer Interface Conference* Verlag der Technischen Universität Graz, Graz, Austria, September 2014.
- [26] D. Steyrl, R. Scherer, J. Faller, and G. R. Müller-Putz, "Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: a practical and convenient non-linear classifier," *Biomedical Engineering/Biomedizinische Technik*, vol. 61, no. 1, pp. 77–86, 2016.
- [27] I. W. Selesnick and C. S. Burrus, "Generalized digital Butterworth filter design," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1688–1694, 1998.
- [28] G. Pfurtscheller and A. Aranibar, "Event-related cortical desynchronization detected by power measurements of scalp EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 42, 1977.
- [29] S. Mohdiwale, M. Sahu, G. R. Sinha, and V. Bhateja, "Statistical wavelets with harmony search based optimal feature selection of EEG signals for motor imagery classification," *IEEE Sensors Journal*, vol. 21, 2020.
- [30] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," *Computers in Biology and Medicine*, vol. 112, Article ID 103375, 2019.
- [31] S. Cang and D. Partridge, "Feature ranking and best feature subset using mutual information," *Neural Computing & Applications*, vol. 13, no. 3, pp. 175–184, 2004.
- [32] N. Mantel, "Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure," *Journal of the American Statistical Association*, vol. 58, no. 303, pp. 690–700, 1963.
- [33] M. A. Hall, *Correlation-based Feature Selection for Machine Learning*, 1999.
- [34] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson Education India, Delhi, India, 2016.
- [35] M. Sahu, Y. Sharma, D. Sharma, and S. Bajpai, "Feature compression using PCA on motor imagery classifications," in *Proceedings of the 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, Malaviya National Institute of Technology, Jaipur, India, March 2018.
- [36] M. Sahu and S. Shukla, "Impact of feature selection on EEG based motor imagery," in *Information And Communication Technology For Competitive Strategies*, pp. 749–762, Springer, New York, NY, USA, 2019.
- [37] R. Kumar and M. Sahu, "Application of a particle swarm optimization technique in a motor imagery classification problem," in *Modern Optimization Methods for Science, Engineering and Technology*, IOP Publishing, Bristol, UK, pp. 2053–2563, 2019.

Research Article

Complexity and Entropy Analysis to Improve Gender Identification from Emotional-Based EEGs

Noor Kamal Al-Qazzaz ^{1,2}, Mohannad K. Sabir,¹ Sawal Hamid Bin Mohd Ali,²
Siti Anom Ahmad,^{3,4} and Karl Grammer⁵

¹Department of Biomedical Engineering, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad 47146, Iraq

²Department of Electrical Electronic & Systems Engineering, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, UKM, Bangi, Selangor 43600, Malaysia

³Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia, UPM, Serdang, Selangor 43400, Malaysia

⁴Malaysian Research Institute of Ageing (MyAgeing™), Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia

⁵Department of Evolutionary Anthropology, University of Vienna, Althan Strasse 14, A-1090 Vienna, Vienna, Austria

Correspondence should be addressed to Noor Kamal Al-Qazzaz; noorbme@kecbu.uobaghdad.edu.iq

Received 14 May 2021; Revised 29 August 2021; Accepted 1 September 2021; Published 22 September 2021

Academic Editor: Cosimo Ieracitano

Copyright © 2021 Noor Kamal Al-Qazzaz 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.

Investigating gender differences based on emotional changes becomes essential to understand various human behaviors in our daily life. Ten students from the University of Vienna have been recruited by recording the electroencephalogram (EEG) dataset while watching four short emotional video clips (anger, happiness, sadness, and neutral) of audiovisual stimuli. In this study, conventional filter and wavelet (WT) denoising techniques were applied as a preprocessing stage and Hurst exponent (Hur) and amplitude-aware permutation entropy (AAPE) features were extracted from the EEG dataset. k -nearest neighbors (kNN) and support vector machine (SVM) classification techniques were considered for automatic gender recognition from emotional-based EEGs. The main novelty of this paper is twofold: first, to investigate Hur as a complexity feature and AAPE as an irregularity parameter for the emotional-based EEGs using two-way analysis of variance (ANOVA) and then integrating these features to propose a new CompEn hybrid feature fusion method towards developing the novel WT_CompEn gender recognition framework as a core for an automated gender recognition model to be sensitive for identifying gender roles in the brain-emotion relationship for females and males. The results illustrated the effectiveness of Hur and AAPE features as remarkable indices for investigating gender-based anger, sadness, happiness, and neutral emotional state. Moreover, the proposed WT_CompEn framework achieved significant enhancement in SVM classification accuracy of 100%, indicating that the novel WT_CompEn may offer a useful way for reliable enhancement of gender recognition of different emotional states. Therefore, the novel WT_CompEn framework is a crucial goal for improving the process of automatic gender recognition from emotional-based EEG signals allowing for more comprehensive insights to understand various gender differences and human behavior effects of an intervention on the brain.

1. Introduction

Perceiving gender based on human emotions has gained lots of research interest to investigate personal characteristics in neuroscience and psychology [1]. Gender differences primarily based on processing emotions have attracted precise interest due to their attainable utility in understanding human psychopathology such as depression and

nervousness that might also be associated with the differential response of females and males to stress [2].

Thus far, few researchers have investigated gender variations primarily based on emotional changes [3], and most of them report substantial differences [2]. Accordingly, the kind of stimulus could be visual, auditory, or audiovisual stimuli. The visual stimuli and auditory stimuli are related to an increase or decrease in the sensorimotor rhythm

amplitude [4]. To reveal personal characteristics that would be valuable in recognizing individual gender accurately in daily life, visual and auditory stimuli are considered as two common ways for human beings to elicit different emotional states [2]. Recently, researchers indicated that, to provide the best environment for automatic emotion recognition, they need to get the combined effect of both visual and auditory stimuli to elicit a specific emotional state [5]. Audiovisual elicitations utilizing short film video clips are usually used to elicit various conditions of emotion better compared to the other modalities [4, 6–9]. Hence, in this work, emotions were precipitated with the aid of the use of short audiovisual video clips.

From the psychological point of view, the emotional state can be distinguished and grouped into two emotional models: the discrete model and the dimensional model. The discrete model comprises a lot of discrete emotional states that are identified to be one of the core emotions, and all other different emotions are considered part of these primary emotions (anger, fear, disgust, surprise, happiness, and sadness) or an aggregate of them [10, 11]. The dimensional model is a two-dimensional (2D) cognitive-emotional state model that is broadly utilized in mapping emotion recognition applications. It plots emotions on two scales, valence-arousal plots, where valence is in the horizontal axis and is considered as the polarity or the quality of an emotion ranging from unpleasant to pleasant and arousal is in the vertical axis and is considered as the intensity of emotion ranging from calm to excited [12]. Therefore, the 2D cognitive-emotional state model is the mapping of all emotions onto the valence-arousal graph, as portrayed in the circumplex model of emotion [13, 14]. Other researchers have proposed a three-dimensional (3D) cognitive-emotional state model which takes into consideration the attention-rejection property in addition to the 2D model [8, 15–17].

In this study, the conventional filter and wavelet (WT) denoising techniques were applied as a preprocessing stage to the EEG dataset. Hurst exponent (Hur) complexity feature and amplitude-aware permutation entropy (AAPE) irregularity parameter have been computed to investigate the gender changes of the emotional-based EEGs. Subsequently, the individual performances of these features were statistically examined using two-way analysis of variance (ANOVA) to recognize a gender-specific role in the brain-emotion relationship for females and males during four short emotional video clips (anger, happiness, sadness, and neutral) of audiovisual stimuli. Then, the used features were combined as a novel complexity and irregularity features (CompEn) hybrid feature fusion set towards developing the novel WT_CompEn framework for automated gender recognition system on EEG for gender identification. Finally, kNN and SVM classification techniques were used for automatic gender identification of emotional-based EEG datasets. The performances of these classifiers were examined on Hur and AAPE individually and on the CompEn feature set.

To the best of author's knowledge, the contribution of a gender-precise role in the brain-emotion relationship has been tended to in this work. Therefore, the main novelty of

this paper is threefold. First, it aims to propose an automated gender recognition system based on EEG data of different emotional states acquired using low-cost wireless EEG devices. This can be done by investigating the changes in complexity and irregularity features of the emotional-based EEGs using statistical analysis. Then, integrate the employed features as CompEn feature set towards developing the novel WT_CompEn framework as a core for automated gender recognition system feature set to be sensitive for identifying gender differences of emotional-based EEG signals. Third, the EEG elicitation convention and the EEG estimation system are utilized without precedent for this investigation for emotion data obtaining, and that may make gender contrast more articulated and may accomplish better performance.

2. Related Works

Over the last decade, studies have indicated that the possible adequacy of biomedical signs for recognizing people by exploring gender differences based on emotional changes would be elicited using different physiological measurements such as electrocardiogram (ECG) [18] and electroencephalogram (EEG) [6, 19]. Several studies illustrated the gender differences and classification from ECG signal analysis [20, 21], while in other studies, the gender has been classified based on using EEG signals [22, 23].

Characterized by wide availability, affordability, and lack of invasiveness, EEG is a clinical instrument capable of monitoring data processing in millisecond accuracy with a high level of temporal resolution [24]. Therefore, it has neurophysiology applications for the detection and differentiation of modifications in the brain [24, 25].

A wide range of brain disorders, including seizures, attention-deficit hyperactivity disorder (ADHD), and Alzheimer's disease (AD)/vascular dementia (VaD), have been detected based on EEG signals, while mental tasks and sleep stages have been classified based on such signals as well [23–26]. The latest research has employed EEG for high time-resolution evaluation of affective moods in people [26–29]. Recently, EEG has been generally utilized to assess human emotional states with high time resolution [6, 30, 31]. Given the important insight that it can provide in this regard, EEG may be a promising biomarker for the appraisal of different affective reactions from an EEG dataset with multiple channels across brain regions [32]. To give an example of such research, brain waveforms were used in [33] to develop a method of uninterrupted music emotion detection. Similarly, in [34], real-time techniques of human emotion detection based on EEG were employed to devise an integrated music therapy for the identification of present affective moods according to neurofeedback and patient-specific customization of treatment.

Besides being highly informative about brain physiology, EEG signals could potentially be biomarkers of brain linear and nonlinear behavior [26, 27, 35–37]. The Hurst exponent (Hur) [6, 38] and fractal dimension (FD) [39, 40] are among the nonlinear techniques that have been adopted for the representation of complex affective tasks and for the

examination of complicated dynamic data generated by the brain cortex [33, 41].

EEG signals are considered dynamic systems lacking stability, and the uncertainty of such systems can be determined by employing the nonlinear parametric index of entropy [42]. Research into cognitive mental states, sleep states, and approaches for categorizing affective levels has benefitted from the application of entropy to EEG signals [35, 43–45]. Furthermore, the use of a range of entropies for the identification of biological gender based on EEG could be useful in clinical analyses, especially on social emotion, individual identification, response to therapy, clinical effectiveness, and side effects [46]. To give an example, in [1], human emotions triggered by video clips were examined based on sample entropy (SampEn), approximate entropy (ApEn), and permutation entropy (PerEn), as these entropies are resistant to noise and can effectively measure time series complexity. In a different study, the analysis of EEG signals for clinical evaluation was conducted based on PerEn entropy and symbolic transfer entropy, hinting at the relational ability of the employed EEG entropy examination with clinical cases of different cognitive conditions [47]. Another type of entropy suggested for EEG examination is fuzzy entropy (FuzEn), involving the substitution of Heaviside functions with fuzzy membership functions [48, 49]. According to existing studies, the issue of entropy mutation is mitigated by FuzEn, but on the downside, the relevant information is lost when employing such entropy techniques because they entail single-scale analysis. Whereas the speed of SampEn is better compared to FuzEn, greater consistency and reduced reliance on data length are demonstrated by FuzEn [50]. ApEn [51], SampEn [52], FuzEn [53], and PerEn [54] constitute the four most popular entropy predictors within the context of EEG signal processing [50]. To identify how affective-based EEG signals across the brain differ between genders, the present work concentrates on EEG-derived indices.

Support vector networks (SVNs), artificial neural networks (ANNs), k -nearest neighbors (kNN) and support vector machine (SVM) classifiers [55], and hidden Markov models (HMM) have all been employed to investigate automatic algorithms for a system of gender categorization [56, 57]. For instance, the SVM classifier was used in [22] to devise an EEG signal-based automatic system of age and gender detection, while EEG data related to resting state were the basis of a model of automatic gender detection in [1]. In other research, EEG sensors with wavelet transform frequency breakdown for feature extraction and random forest classifier enabled the creation of an automatic system for detecting age and gender in resting state with eyes closed [58, 59].

Most gender detection studies using EEG signals based on emotional response focused on the linear analysis using spectral relative powers [30, 60, 61]. However, other researchers have used nonlinear features to investigate brain complexity [62–64]. In the current study, we aim to understand the role of EEG for gender identification using the integrated entropy and spatial features to characterize the emotional-based signals by examining different brain region

behaviors during audiovisual video clips. Integrated features are essential for an automatic gender detection system to perform effectively and be solidly reliable. In this context, the impact of gender discrepancies on the elements of EEG-based systems of affective reaction detection and the general performance of such systems are worth investigating. To this end, computation of entropy features was done to highlight the gender variability occurring in affective-based EEG systems.

3. Materials and Methods

Figure 1 illustrates the block diagram of the proposed study.

3.1. EEG Acquisition and Recording. A mobile and affordable Emotiv EPOC EEG 14-channel headset (Emotiv Systems, Inc., San Francisco, CA) was employed in this work to capture EEG signals labeled as AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, with the common mode sense (CMS) left mastoid and the driven right leg (DRL) being referenced as ground. The positioning of the sponge-based electrodes used by the headset was done according to the 10–20 system, while a band-pass filter of 0.5–70 Hz frequency facilitated the filtering of electrode data. The frequency of sampling was 128 Hz, with 0.51 mV resolution.

The study recruited a total of ten participants (6 males and 4 females; the age of 22.6 ± 2.75 years, mean \pm standard deviation (SD)), all university students, aged between 18 and 24 years. Before beginning the research, each participant underwent an evaluation to ensure no prior history of neurological or psychiatric issues and was then presented with an informed consent form (ICF) which they were requested to sign before participating in the study.

During the EEG recording procedure, subjects were asked to remain relaxed and calm for the entire EEG recording duration to minimize the data reading artifacts resulting from movements. The evaluation of the 3 emotion states (anger, sadness, and happiness) along with the neutral condition was conducted by allowing the participants to view various short emotionally stimulating video clips, with audio, following which the participants were allowed some time to evaluate and grade their responses to the clips employing a self-assessment questionnaire, followed by a break of 45 seconds before viewing the next video clip (Figure 2) [65].

The running time of the various video clips varied from one to the other, with the longest having a duration of four minutes. The emotional video clips used were selected based on those recommended by Rottenberg et al. [65]. As previously mentioned, participants were asked to evaluate the strength of their emotional response to each clip using a five-point-scale SAQ; participants were asked to select either 1 (very low), 2 (low), 3 (medium), 4 (high) or 5 (very high) to evaluate the degree of emotion experienced [31].

To enable the participants to view the affective video clips, the used video clips were in German language and the virtual emotion presenter (VEP) software from the University of Vienna was employed. This software was chosen

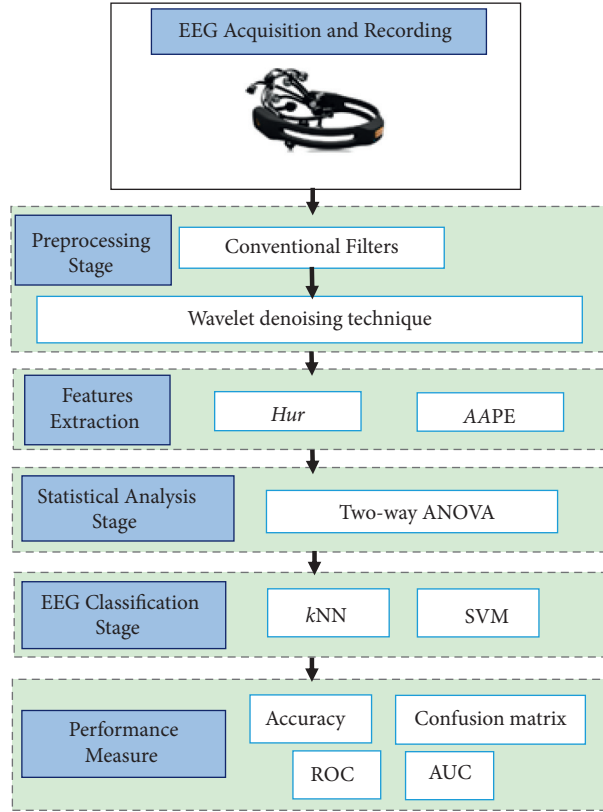


FIGURE 1: The block diagram of the proposed study.

because it not only permits arbitrary viewing but also documents extra sources of data. The experimental work was conducted in the Anthropology Research Laboratory, and besides the VEP software, the equipment used included the regular laboratory ambient lighting, LCD for screening the video clips, and stereo speakers so that the video clips could be accompanied by uninterrupted sound at a level acceptable to the participants (Figure 3). The Helsinki declaration and subsequent refinements were followed in every research procedure.

3.2. Preprocessing Stage. Since most artifacts occurring in EEG signals were overlapping with brain activity, preprocessing is essential in EEG signal processing.

3.2.1. Conventional Filtering. In this context, conventional filters were used as an initial stage to process each channel of the recorded EEG datasets. A notch filter at 50 Hz was used to remove the power line interference noise [32], and a fourth-order Butterworth bandpass filter was applied with a 0.5–64 Hz frequency range to limit the band of the recorded EEG signals [66].

3.2.2. Wavelet Analysis. WT has the ability in resolving EEG into specific time and frequency components by providing a good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. The DWT is a fast

nonredundant transform used in practice for analyzing both the low- and high-frequency components in the EEG signals because it requires less computational time than the continuous WT (CWT) [67]. The DWT can be processed by obtaining the discrete value of the parameters a and b , as in equation (1). It can be obtained as a set of decomposition functions of the correlation between the signal $f(t)$ and the shifting and dilating of one specific function called mother wavelet function $\psi(t)$. MWT is shifted by the location parameter (b) and dilated or contracted by frequency scaling parameter a , as in the following equation [8, 16, 17, 36, 68–70]:

$$\text{DWT}_{m,n}(f) = a_0^{-(m/2)} \int f(t) \psi(a_0^{-m}t - nb_0) dt, \quad (1)$$

where a_0 and b_0 values are set to 2 and 1, respectively.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad a \in \mathbb{R}^+, b \in \mathbb{R}. \quad (2)$$

SURE threshold is an adaptive soft thresholding method, which aims to determine the threshold limit for each level based on Stein's unbiased risk estimation [71] and commonly used value in [72–74].

In this study, the sampling frequency was 128 Hz and the EEG dataset was subjected to “sym9” from the Symlets family with a four-decomposition level of five subband EEG signals. Among the five subbands, cD1, cD2, cD3, and cD4 represented the decomposition detail coefficients and cA is the decomposition approximation coefficient. The SURE

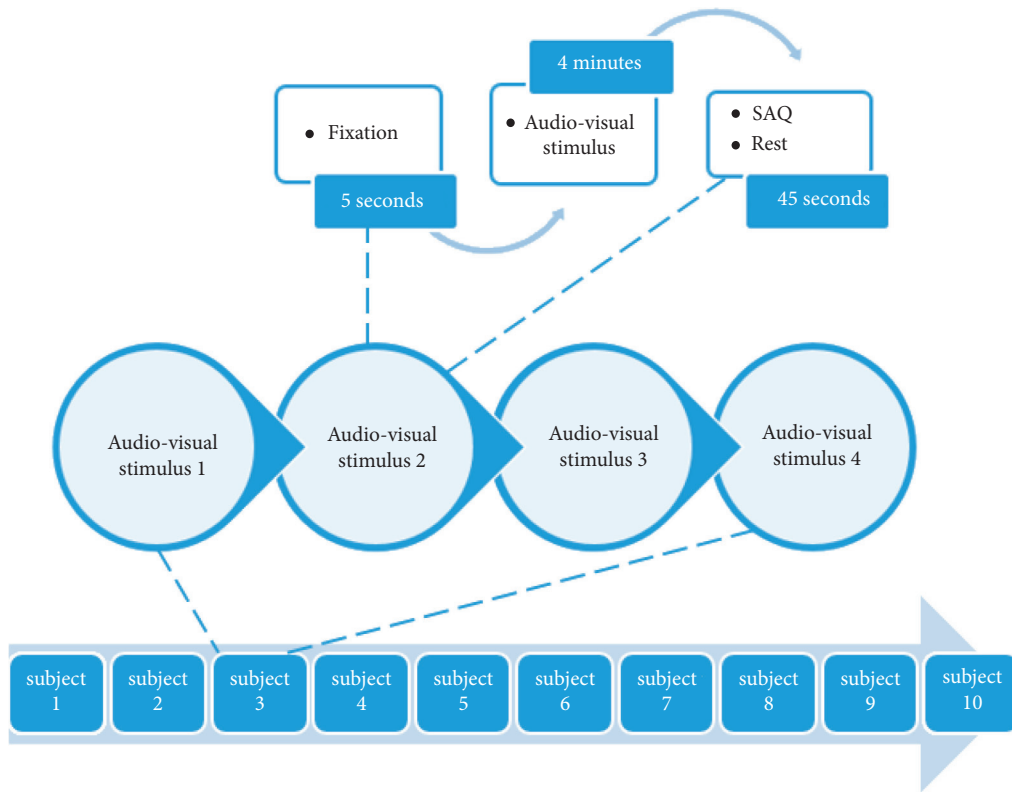


FIGURE 2: The experimental protocol of emotion [9].

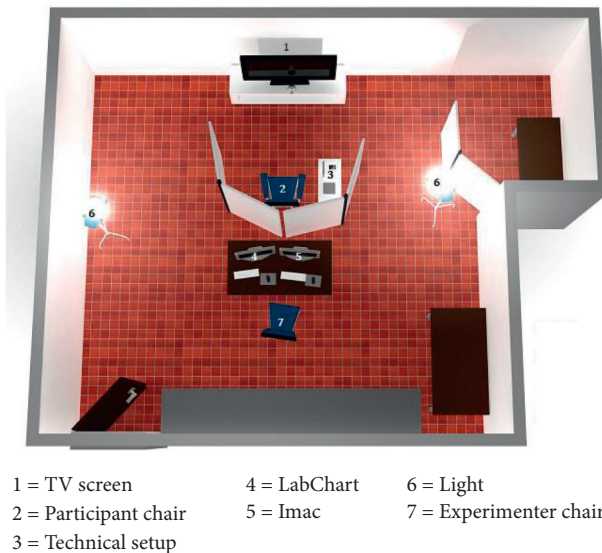


FIGURE 3: Setup of the experimental room with presentation TV and the recorders.

threshold is an adaptive soft thresholding method that is used to find the threshold limit for each level based on Stein’s unbiased risk estimation [70, 74].

3.3. Features Extraction Stage. Because of the complicated structure of the brain and its ability to perform multiple and complex sophisticated cognitive tasks, the brain neurons are

considered to be governed by nonlinear dynamic phenomena. EEG signals have been used to investigate the chaotic behavior of the brain from nonlinear time series [75, 76]. Since the EEG spectral-band analysis was unable to illustrate the electrical activity of the brain and the underlying mechanisms of the brain function, the nonlinear analysis based on dynamics information needs to be investigated. The present study was undertaken to examine the

gender differences from emotional-based EEG background activity with two different features: Hurst exponent (Hur) and amplitude-aware permutation entropy (AAPE) to illustrate the complexity and irregularity features in detecting gender differences [63, 77–79]. Indeed, the used features were selected based on previous studies due to their usefulness and effectiveness in discriminating the EEG signals [35, 80, 81]. In this stage, the filtered EEG datasets were segmented into 3 trials, and each trial includes 10 seconds of each video clip viewed (3×10 second period) with 1280 data points.

3.3.1. Hurst Exponent (Hur). Hur is a measure that has been widely used to evaluate the self-similarity and correlation properties of fractional Brownian noise and the time series produced by a fractional (fractal) Gaussian process. Hur is used to evaluate the presence or absence of long-range dependence and its degree in a time series. However, local trends (nonstationarities) are often present in physiological data and may compromise the ability of some methods to measure self-similarity. Hur is the measure of the smoothness of a fractal time series based on the asymptotic behavior of the rescaled range of the process. In time series analysis of EEG, Hur is used by [38, 80] to characterize the nonstationary behavior of the EEG signals. Hur is defined as

$$\text{Hur} = \frac{\log(R/S)}{\log(T)}, \quad (3)$$

where T is the duration of the sample of data and R/S is the corresponding value of the rescaled range. The above expression is obtained from Hurst's generalized equation of time series that is also valid for Brownian motion [82].

3.3.2. Amplitude-Aware Permutation Entropy (AAPE). AAPE has been proposed to consider the amplitude information from permutation entropy (PE) to overcoming the PE shortcoming of considering the order of the amplitude and discarding the information regarding the amplitude, besides the equal amplitude values in each embedded vector are not considered.

To estimate AAPE, assume $y = \{y_{t+(j_1-1)l}, y_{t+(j_2-1)l}, \dots, y_{t+(j_d-1)l}\}$ is the time series, where j is the time index of the element in the reconstruction vector, a vector including the $d!$ potential symbol patterns of π motifs, where d is the embedded dimension, which determines how much information is contained in each vector, and l is the time delay of the order pattern i , $i = 1, 2, \dots, d!$. To calculate AAPE, for each π_i , $p(\pi_k)$ demonstrates the relative frequency as follows [81]:

$$p(\pi_k) = \left\{ \begin{array}{ll} p(\pi_i^{d,l}) + \left[A/d \sum_{k=1}^d |x_{i+(k-1)l}| + 1 - A/d - 1 \sum_{k=2}^d |x_{i+(k-1)l} - x_{i+(k-2)l}| \right] & \text{if } p(\pi_i^{d,l}) = 0 \\ \frac{p(\pi_i^{d,l})}{\sum_{i=1}^{N-d+1} \left[A/d \sum_{k=1}^d |x_{i+(k-1)l}| + 1 - A/d - 1 \sum_{k=2}^d |x_{i+(k-1)l} - x_{i+(k-2)l}| \right]} & \text{otherwise} \end{array} \right\}, \quad (4)$$

$$\text{AAPE}(d, l, n) = - \sum_{\pi_k=1}^{\pi_k=d!} p(\pi_k) \ln p(\pi_k).$$

When all motifs have equal probability, the largest value of AAPE is obtained at $l = 1$. For 30 seconds, $N = 3840$ samples, 3 windows of 10-second length (1280 samples) were extracted from the original EEG time series for each 14 channels.

3.3.3. Complexity and Entropy Features Fusion. To get an efficient gender recognition model in terms of high accuracy recognition rates and to have more insights on the mental processes for females and males, the Hur index of complexity and AAPE index of irregularity have been combined to develop a new hybrid index of complexity-entropy (CompEn) set of feature.

3.4. Statistical Analysis Stage. This study intends to investigate the significance of Hur and AAPE features to be reliable indices in detecting gender differences in anger, happiness, sadness, and neutral emotional states. Therefore, statistical analysis has been conducted using SPSS statistical tool version 22. Two sessions of two-way analyses of variance (ANOVA) were performed to realize the significant differences among the emotions (i.e., anger, happiness, sadness, and neutral), and Hur was considered for the first session and AAPE was considered for the second session. Hur and AAPE were applied as dependent variables. The group factor (i.e., female and male) was the independent variable. The significance was set at $p < 0.05$. Moreover, the study was aimed to test the hypothesis that the gender differences from

emotional-based EEG performed from the complexity and entropy-based features would be different between females and males.

4. Classification Stage and Performance Measures

The last stage for identifying neurophysiological changes in females and males is using the classification model. In this study, k -nearest neighbors (kNN) and support vector machine (SVM) were used.

Given that the majority of the learning algorithms assume a balanced class distribution, their results typically favor the predominant class that gives poor class predictions. The class imbalance in the dataset highly affects the quality of the classification model. However, given that the minority class cannot be easily discriminated against, the classifier can simply classify each instance as the majority class. In this study, the minority class was represented by the females. A synthetic oversampling technique (SMOTE) was applied to overcome the data imbalance [83]. The classifier parameters and percentage of oversampling were determined via 10-fold cross-validation using a grid search approach to avoid overfitting and bias in the classification analysis [84]. The available dataset was divided into 10 equal size disjoint subsets. One of these subsets was used as the test set, while the remaining nine subsets were combined into a training set to learn the classifier. This procedure was performed 10 times, which resulted in 10 accuracies. The average of these accuracies represented the 10-fold cross-validation accuracy of learning from this dataset [85]. Given that SMOTE changes the dataset, the percentage of oversampling were combined with the parameters. Therefore, those parameters that are found with different SMOTE percentages may not be the same. Using only the training set, the SMOTE was used to equalize the frequency of the classes [86, 87].

kNN is one of the most popular nonparametric classification algorithms, it is more robust when $k > 1$ particularly to reduce the influenced noisy points within the training set. In this study, the Euclidean distance was utilized as a similarity measure to classify each trial by kNN . The classifier was trained to obtain the best value of $k = 7$ that maximizes the overall classification performance evaluation. kNN with 7 neighbors classifiers were selected based on previous work [9, 28, 29, 43, 77, 79, 88].

Optimization of the complexity parameter C with a range of $-4 \leq \log_{10}(C) \leq 4$ in C values $C \in \{0.0001, 0.001, 0.01, 0.1, 0, 10, 100, 1000, 10000\}$ on the training set via ten-fold cross-validation yielded ideal outcomes for the SVM classifier. During testing, C corresponding to 10 gave optimal results for C values. The multiclass SVM classifiers were applied based on the radial basis function (RBF) kernel. Furthermore, the training dataset was used to determine the minimum misclassification rate, which in turn helped to obtain the smoothing parameter σ in the context of SVM training. Methodical variation of σ value in different training episodes is the only way of determining the ideal σ . Hence, in this work, variation of the σ value was done in the range of 0.1–1 at 0.1

intervals. A σ value of 0.5 was established to be associated with the minimum misclassification rate.

The performance of the proposed framework was evaluated using the values of average classification accuracy, confusion matrix, receiver operating characteristic curve (ROC), and area under the curve (AUC).

5. Results and Discussions

5.1. Results of Preprocessing Stage. As previously described, the EEG signal datasets were filtered by conventional filters and subjected to the WT denoising technique. Figure 4 illustrates the data obtained from channel 7, representing the frontal brain area when subjected to the emotional state of anger. Observation shows that the artifactual signal elements (blue lines) present in the raw EEG signal were successfully blocked during signal denoising, resulting in the clean EEG signal (red line).

5.2. Results of Statistical Analysis. The statistical characterization of the differences in Hur and AAPE females and males will be discussed in the following sections.

5.2.1. Results of Hurst Exponent (Hur). The boxplots of Figure 5 indicate the overall pattern of Hur feature response for the two group factor distribution (i.e., female and male) from emotional-based EEG signals. It can be observed that Hur provides a significant variation with a useful way to visualize the characteristics of responses for the female and male group factors. Furthermore, boxplot analysis demonstrates the median value, as the value inside the boxplots is the median value of the distribution. The typical boxplot has lines at the upper median and lower quartile values. Figure 5 confirms the suitability of the feature for pattern classification.

Moreover, to recognize the importance of the complexity feature Hur method for the pattern classification, statistical analysis using two-way ANOVA was conducted on the Hur features. In this analysis, the group factor (i.e., female and male) was the independent variable, whereas the Hur features were the dependent variable. The significance for all statistical tests was set at $p < 0.05$. Normality was then assessed using the Kolmogorov–Smirnov test, whereas homoscedasticity was verified using Levene’s test. The post hoc comparison was performed through Duncan’s test.

Figure 6 illustrates the comparative plot of Hur which was estimated to discriminate between females and males based on anger, happiness, sadness, and neutral emotional states based on EEG signal complexity. Anger, happiness, and neutral were statistically significant from sadness, particularly for females, whereas anger, happiness, and sadness were statistically significant from neutral for males. One can see that the females had significantly lower Hur values at the four different emotional states compared to males $((Hur_{\text{anger,happiness,sadness,neutral}}(\text{Females}) < Hur_{\text{anger,happiness,sadness,neutral}}(\text{Males}))$ with significant differences ($p < 0.05$). These results suggest that the EEG activities of females are significantly less complex for males.

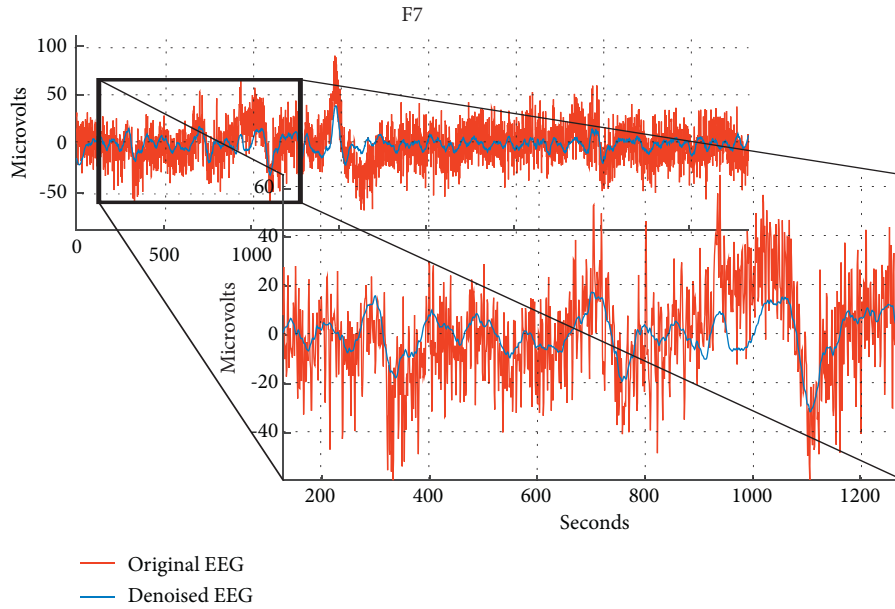


FIGURE 4: The denoising results after preprocessing stage for channel F7.

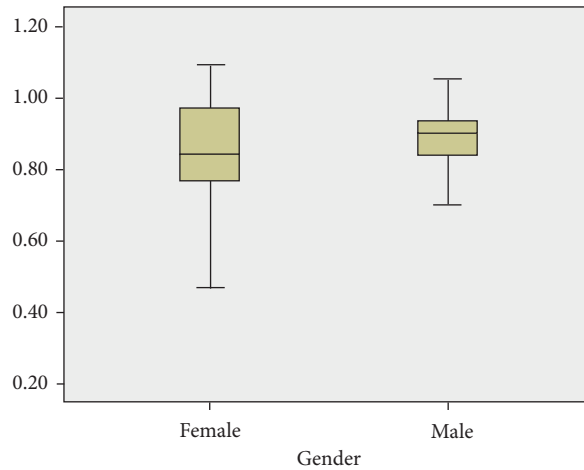


FIGURE 5: Boxplots for the Hur feature extraction method for gender distribution from emotional-based EEG signals. The dark black line represents the median values.

5.2.2. *Results of Amplitude-Aware Permutation Entropy (AAPE)*. The boxplots of Figure 7 indicate the overall pattern of AAPE feature response for the two group factor distribution (i.e., female and male) from emotional-based EEG signals. It can be observed that AAPE provides a significant variation with a useful way to visualize the characteristics of responses for the female and male group factors. Furthermore, boxplot analysis demonstrates the median value, as the value inside the boxplots is the median value of the distribution. The typical boxplot has lines at the upper median and lower quartile values. Figure 7 confirms the suitability of the feature for pattern classification.

Moreover, to recognize the importance of the complexity feature AAPE method for the pattern classification, statistical analysis using two-way ANOVA was conducted on the AAPE features. In this analysis, the group factor (i.e., female

and male) was the independent variable, whereas the AAPE features were the dependent variable. The significance for all statistical tests was set at $p < 0.05$. Normality was then assessed using the Kolmogorov–Smirnov test, whereas homoscedasticity was verified using Levene’s test. The post hoc comparison was performed through Duncan’s test.

In this study, AAPE has been used for discriminating females from males based on anger, happiness, sadness, and neutral emotional states based on EEG signal irregularities. Figure 8 illustrates the comparative plot of AAPE; it can be observed that sadness was statistically significant from neutral anger and happiness. Notably, EEG significantly had lower AAPE values in happiness and sadness for females compared to males ($AAPE_{\text{happiness, sadness}}(\text{Females}) < AAPE_{\text{happiness, sadness}}(\text{Males})$), whereas the females had higher AAPE values for anger and neutral

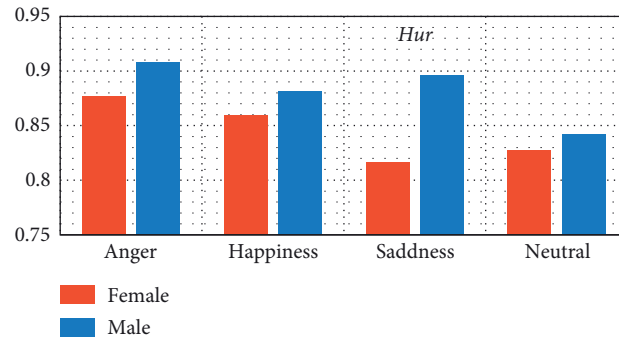


FIGURE 6: Comparative plot of the four tested emotional states for females and males using Hurst exponent complexity feature.

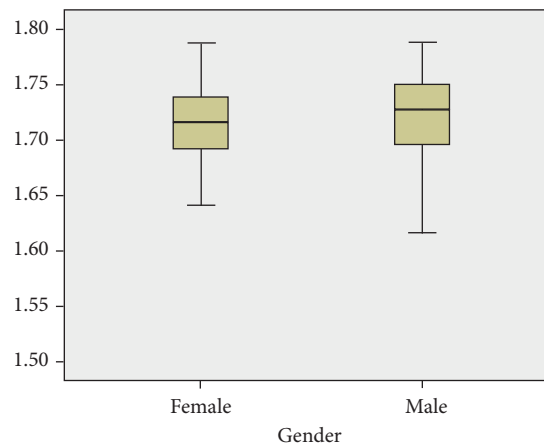


FIGURE 7: Boxplots for the AAPE feature extraction method for gender distribution from emotional-based EEG signals. The dark black line represents the median values.

emotional states compared to males ($AAPE_{\text{anger,neutral}}(\text{Females}) < AAPE_{\text{anger,neutral}}(\text{Males})$) ($p < 0.05$). These results suggest that EEG had regular behavioral activities for both females and males.

5.3. Results of Classification Stage. This study has dealt with emotional-based EEG signals for gender identification problems. The key design decisions for kNN and SVM used in the classification are the training process, as they depend on the size of the training set and the test set. However, to comparatively evaluate the performance of the proposed classifiers, the classifiers employed in this work were trained on the same training data set and tested on the testing data set.

5.3.1. Results of Hurst Exponent and Classification Performance. Tables 1 and 2 display the confusion matrix for female and male identification from emotional-based EEG signals using Hur complexity index with kNN and SVM classifiers, respectively, in which correct recognition is shown on the diagonal and substitution errors are off-diagonal.

In Table 1, the two diagonal cells show the percentage of correct classification using kNN classifier. For example,

females are correctly classified with 58.3%; similarly, 100% are correctly classified as males, whereas 41.7% of females are incorrectly classified as males.

The results show that kNN classifier can differentiate females and males from emotional-based EEG signals with a high accuracy of 83%. Moreover, Figure 9 illustrates the ROC curve and the AUC value obtained from the investigation of the Hur features.

In Table 2, the two diagonal cells show the percentage of correct classification using SVM classifier. For example, females are correctly classified with 80%; similarly, 90% are correctly classified as males. Moreover, 20% of females are incorrectly classified as males, whereas 10% of males are incorrectly classified as females.

The results show that SVM classifier can differentiate females and males from emotional-based EEG signals with a high accuracy of 86.7%. Moreover, Figure 10 illustrates the ROC curve and the AUC value obtained from the investigation of the Hur features.

5.3.2. Results of Amplitude-Aware Permutation Entropy and Classification Performance. Tables 3 and 4 display the confusion matrix for female and male identification from emotional-based EEG signals using AAPE entropy index with kNN and SVM classifiers, respectively, in which correct

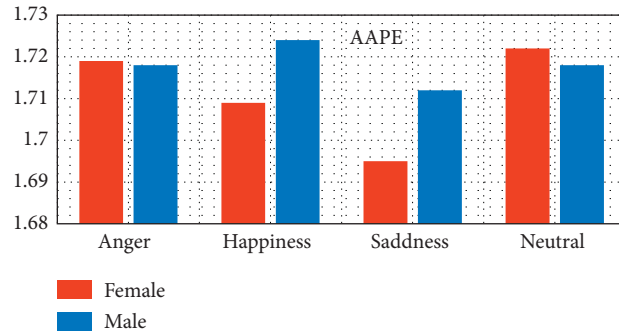


FIGURE 8: Comparative plot of the four tested emotional states for females and males using amplitude-aware permutation entropy feature.

TABLE 1: Confusion matrix calculations for gender classification from emotional-based EEGs using Hurst exponents and kNN classifier.

Predicted	Gender	Actual	
		Females (%)	Males (%)
Hur	Females	58.3	41.7
	Males	0	100

TABLE 2: Confusion matrix calculations for gender classification from emotional-based EEGs using Hurst exponents and SVM classifier.

Predicted	Gender	Actual	
		Females (%)	Males (%)
Hur	Females	80	20
	Males	10	90

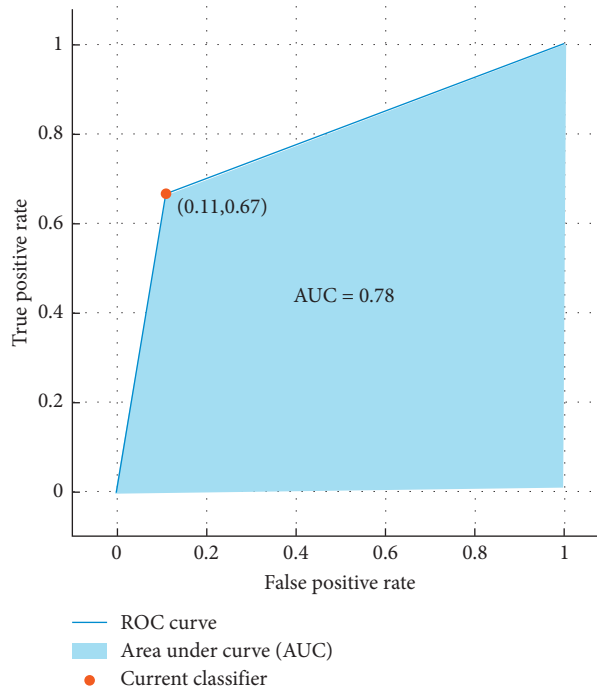


FIGURE 9: ROC curve and the AUC values of gender classification from emotional-based EEGs using Hurst exponent features and kNN classifier.

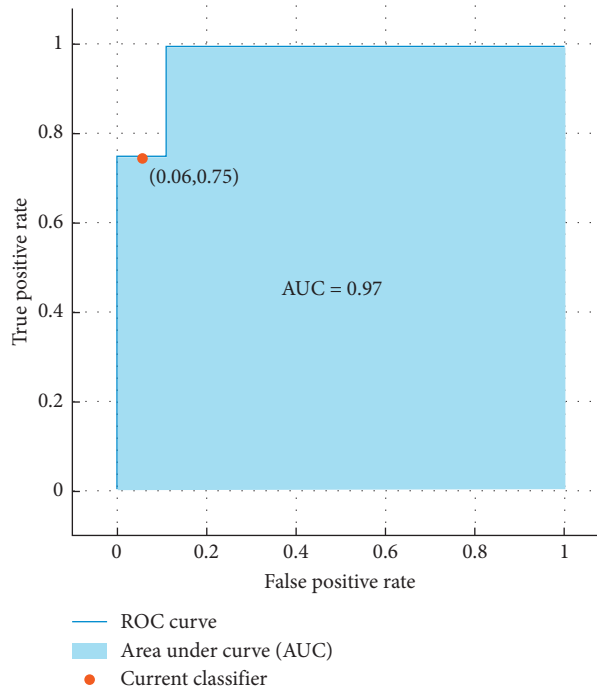


FIGURE 10: ROC curve and the AUC values of gender classification from emotional-based EEGs using Hurst exponent features and SVM classifier.

TABLE 3: Confusion matrix calculations for gender classification from emotional-based EEGs using amplitude-aware permutation entropy and kNN classifier.

Predicted	Gender	Actual	
		Females (%)	Males (%)
AAPE	Females	100	0
	Males	22.2	77.8

TABLE 4: Confusion matrix calculations for gender classification from emotional-based EEGs using amplitude-aware permutation entropy and SVM classifier.

Predicted	Gender	Actual	
		Females (%)	Males (%)
AAPE	Females	90	10
	Males	10	90

recognition is shown on the diagonal and substitution errors are off-diagonal.

In Table 3, the two diagonal cells show the percentage of correct classification using kNN classifier. For example, females are correctly classified with 100%; similarly, 77.8% are correctly classified as males, whereas 22.2% of males are incorrectly classified as females.

The results show that kNN classifier can differentiate females and males from emotional-based EEG signals with a high accuracy of 86.7%. Moreover, Figure 11 shows the ROC curve and the AUC value obtained from the investigation of the AAPE features.

In Table 4, the two diagonal cells show the percentage of correct classification using the SVM classifier. For example,

females are correctly classified with 90%; similarly, 90% are correctly classified as males. Moreover, 10% of females are incorrectly classified as males, whereas 10% of males are incorrectly classified as females.

The results show that the SVM classifier can differentiate females and males from emotional-based EEG signals with a high accuracy of 90%. Moreover, Figure 12 shows the ROC curve and the AUC value obtained from the investigation of the AAPE features.

5.3.3. Results of CompEn Hybrid Index and Classification Performance. Tables 5 and 6 display the confusion matrix for female and male identification from emotional-based EEG

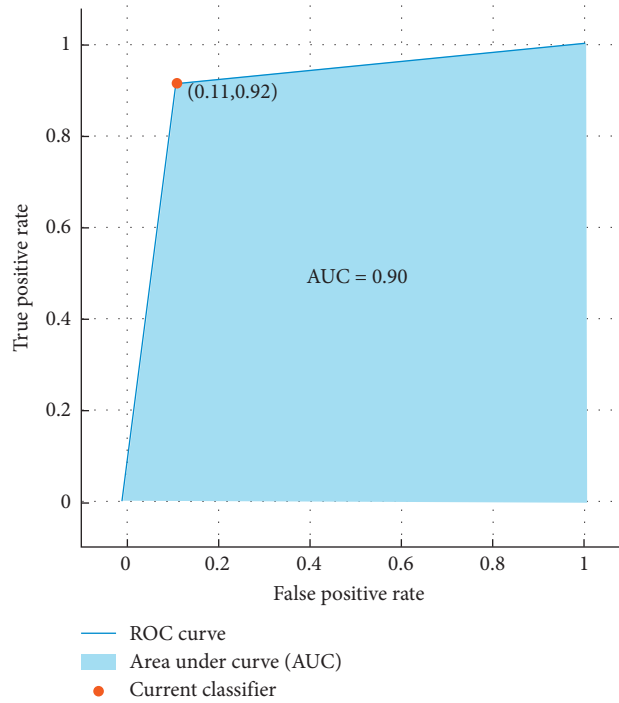


FIGURE 11: ROC curve and the AUC values of gender classification from emotional-based EEGs using amplitude-aware permutation entropy and kNN classifier.

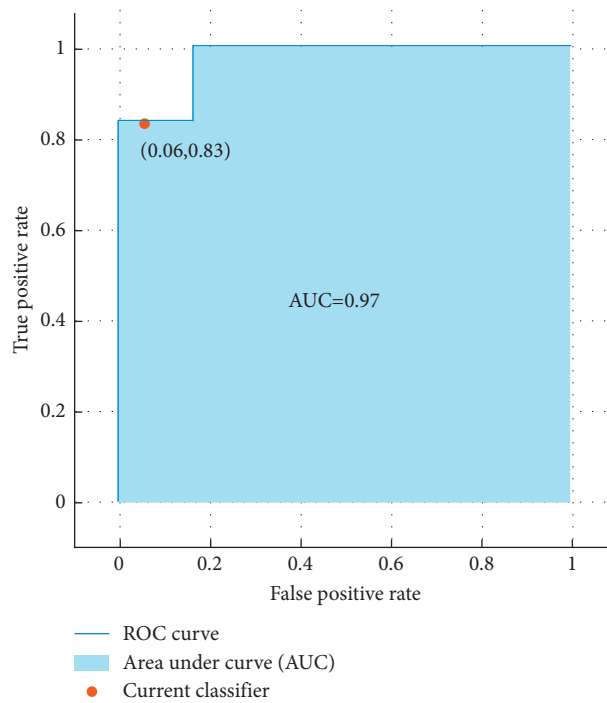


FIGURE 12: ROC curve and the AUC values of gender classification from emotional-based EEGs using amplitude-aware permutation entropy and SVM classifier.

signals using CompEn hybrid index with kNN and SVM classifiers, respectively, in which correct recognition is shown on the diagonal and substitution errors are off-diagonal.

From Table 5, the two diagonal cells show the percentage of correct classification using kNN classifier. The females are correctly classified with 91.7%; similarly, 100% are correctly

TABLE 5: Confusion matrix calculations for gender classification from emotional-based EEGs using CompEn hybrid fusion index and kNN classifier.

Predicted	Gender	Actual	
		Females (%)	Males (%)
CompEn	Females	91.7	8.3
	Males	0	100

TABLE 6: Confusion matrix calculations for gender classification from emotional-based EEGs using amplitude-aware permutation entropy and SVM classifier.

Predicted	Gender	Actual	
		Females (%)	Males (%)
CompEn	Females	100	0
	Males	0	100

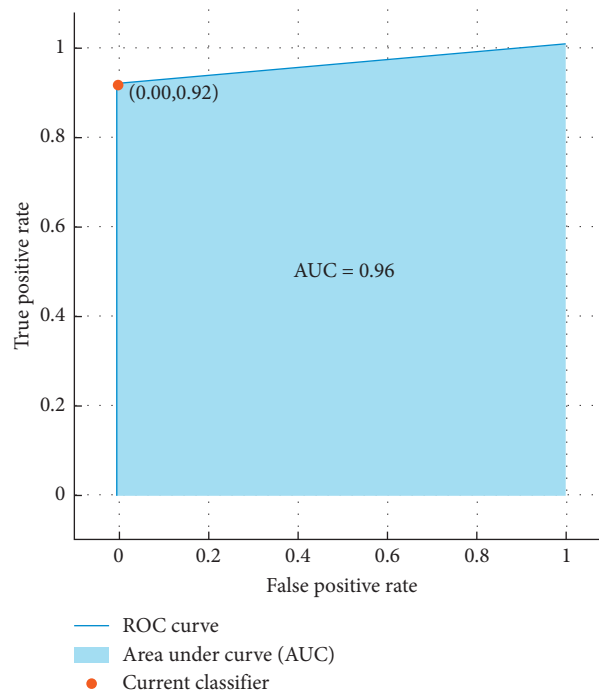


FIGURE 13: ROC curve and the AUC values of gender classification from emotional-based EEGs using proposed CompEn hybrid features and kNN classifier.

classified as males, whereas 8.3% of males are incorrectly classified as females.

The results show that females can be differentiated with a high accuracy of 96.7% using kNN classifier to discriminate females and males from emotional-based EEG signals. Moreover, Figure 13 illustrates the ROC curve and the AUC obtained from the investigation of the CompEn features; the AUC was 0.96 and indicates that the proposed CompEn hybrid index exhibits robust classification performance in discriminating females and males from emotional-based EEGs.

In Table 6, the two diagonal cells show the percentage of correct classification using the SVM classifier. For example, females are correctly classified with 100%; similarly, 100% are correctly classified as males.

The results show that females can be differentiated with a high accuracy of 100% using SVM classifier as a benchmark technique to discriminate females and males from emotional-based EEG signals. Moreover, Figure 14 illustrates the ROC curve and the AUC obtained from the investigation of the CompEn features; the AUC was 1 and indicates that the proposed CompEn hybrid index exhibits robust classification performance in discriminating females and males from emotional-based EEGs.

Therefore, the results showed that the proposed WT_CompEn framework significantly increases the classification accuracy. Indeed, the results emphasize the crucial role played by the novel proposed WT_CompEn framework in the EEG signal processing chain, particularly in the classification results.

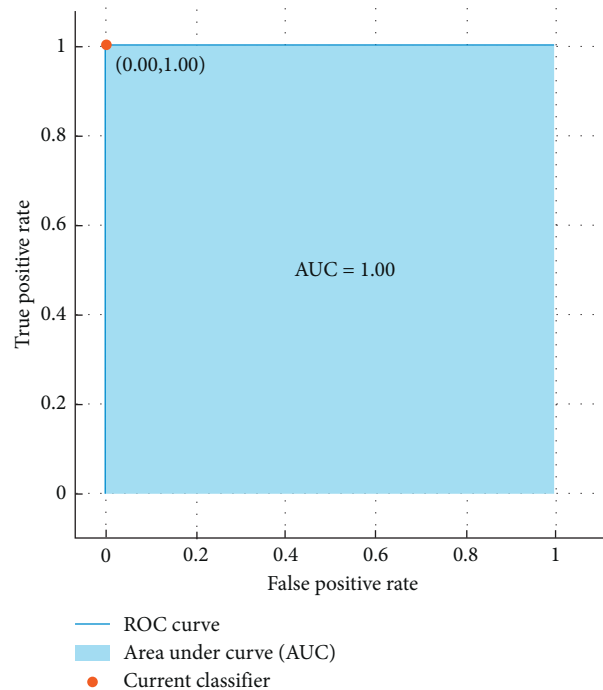


FIGURE 14: ROC curve and the AUC values of gender classification from emotional-based EEGs using proposed CompEn hybrid features using SVM classifier.

Gender recognition framework using emotional-based EEG signals has been performed under MATLAB R2021a on a laptop with processor Intel Core i7-8550U CPU @ 1.80 GHz and 1.99 GHz using 16.0 GB RAM and 64-bit operating system.

However, some limitations also need to be considered in this study; for instance, the sample size was small and an additional analysis with a large database should be performed in the future. Despite this, the different attributes of offline and online categorizations call for additional investigations based on real-time online experiments to validate the results obtained. Such limitations notwithstanding, there is an agreement between the results reported by this work and those of other studies, which confirmed the ability of EEG signals to identify the most gender discrepancies regarding anger, sadness, happiness, and neutral emotions and those discrepancies were reflected in the EEG bands as well [8, 17, 63, 77–79].

6. Conclusion

Conventional filters and WT techniques were used in the preprocessing stage to denoise the EEG datasets of 10 subjects while watching four short emotional video clips (anger, happiness, sadness, and neutral) of audiovisual stimuli. In the second stage, Hur complexity feature and AAPE irregularity parameter have been computed to investigate the gender changes of the emotional-based EEGs. Moreover, ANOVA has been used to statistically examine the individual performance of the used features to recognize a gender-specific role in the brain-emotion relationship for females and males during four short emotional video clips.

Then, the used features were combined as novel complexity and irregularity features CompEn hybrid feature set towards developing the novel WT_CompEn framework as a core for an automated gender recognition system on EEG for gender identification. Finally, kNN and SVM classification techniques have been used for automatic gender identification of emotional-based EEG datasets. The performances of these classifiers were examined on Hur and AAPE individually and on the CompEn hybrid feature set. Potentially, the novel WT_CompEn framework can be used to identify gender differences from emotional-based EEG signals with high classification results.

This study has a primary limitation of the small sample size examined during the experiment. Therefore, further investigations will be carried out on a larger database in the future. Like every work, this study has advantages and weak points. However, gender detection has many advantages as well as applications such as health care, human-computer based interaction, knowing consumer preferences for online retailers, and biometric. Our findings approve the effectiveness of using complexity and irregularity features and CompEn hybrid feature set towards developing the novel WT_CompEn framework as an automated gender recognition system on EEG for gender identification. This study reveals useful insights about gender detection from emotion-based EEG classification. More investigation can be performed to describe the physiological meaning of the extracted features. Other classification approaches can be employed in further studies. In the future, researchers can decrease the computation cost and processing time. It is worth mentioning that the advantages of the current study outweigh the drawbacks.

Data Availability

All data included in this study are available from the corresponding author upon request.

Conflicts of Interest

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

Acknowledgments

The authors would like to express their gratitude to the 10 volunteers who participated in this experiment. This work was supported by Universiti Kebangsaan Malaysia and Ministry of Education, Malaysia, Grant Code GUP-2020-009.

References

- [1] P. Wang and J. Hu, "A hybrid model for EEG-based gender recognition," *Cognitive neurodynamics*, vol. 13, no. 6, pp. 541–554, 2019.
- [2] C. Lithari, C. A. Frantzidis, C. Papadelis et al., "Are females more responsive to emotional stimuli? a neurophysiological study across arousal and valence dimensions," *Brain Topography*, vol. 23, no. 1, pp. 27–40, 2010.
- [3] J. S. Stevens and S. Hamann, "Sex differences in brain activation to emotional stimuli: a meta-analysis of neuroimaging studies," *Neuropsychologia*, vol. 50, no. 7, pp. 1578–1593, 2012.
- [4] C. Maaoui and A. Pruski, "Emotion recognition through physiological signals for human-machine communication," in *Cutting Edge Robotics 2010*, V. Kordic, Ed., IntechOpen, London, UK, 2010.
- [5] D. O. Bos, "EEG-based emotion recognition," *The Influence of Visual and Auditory Stimuli*, vol. 56, no. 3, pp. 1–17, 2006.
- [6] J. Selvaraj, M. Murugappan, K. Wan, and S. Yaacob, "Classification of emotional states from electrocardiogram signals: a non-linear approach based on hurst," *BioMedical Engineering Online*, vol. 12, no. 1, p. 44, 2013.
- [7] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers," *Cognition & Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
- [8] N. K. Al-Qazzaz, M. K. Sabir, S. Ali, S. A. Ahmad, and K. Grammer, "Effective EEG channels for emotion identification over the brain regions using differential evolution algorithm," in *Proceedings of 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2019.
- [9] N. K. Al-Qazzaz, M. K. Sabir, S. H. M. Ali, S. A. Ahmad, and K. Grammer, "The role of spectral power ratio in characterizing emotional EEG for gender identification," in *Proceedings of 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 334–338, IEEE, Langkawi Island, Malaysia, March 2021.
- [10] L. F. Barrett, "Discrete emotions or dimensions? The role of valence focus and arousal focus," *Cognition & Emotion*, vol. 12, no. 4, pp. 579–599, 1998.
- [11] P. Ekman, "An argument for basic emotions," *Cognition & Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [12] I. B. Mauss and M. D. Robinson, "Measures of emotion: a review," *Cognition & Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [13] A. Mehrabian, "Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, no. 4, pp. 261–292, 1996.
- [14] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.
- [15] E. A. Kensinger, "Remembering emotional experiences: the contribution of valence and arousal," *Reviews in the Neurosciences*, vol. 15, no. 4, pp. 241–252, 2004.
- [16] N. K. Al-Qazzaz, M. K. Sabir, and K. Grammer, "Gender differences identification from brain regions using spectral relative powers of emotional EEG," in *Proceedings of the 2019 7th International Work-Conference on Bioinformatics and Biomedical Engineering*, pp. 38–42, LNBI, Granada, Spain, May 2019.
- [17] N. K. Al-Qazzaz, M. K. Sabir, and K. Grammer, "Correlation indices of electroencephalogram-based relative powers during human emotion processing," in *Proceedings of the 2019 9th International Conference on Biomedical Engineering and Technology*, pp. 64–70, ACM, Tokyo, Japan, March 2019.
- [18] A. J. Moss, "Gender differences in ECG parameters and their clinical implications," *Annals of Noninvasive Electrocardiology*, vol. 15, no. 1, pp. 1–2, 2010.
- [19] J. Hu, "An approach to EEG-based gender recognition using entropy measurement methods," *Knowledge-Based Systems*, vol. 140, pp. 134–141, 2018.
- [20] J. Xue and R. M. Farrell, "How can computerized interpretation algorithms adapt to gender/age differences in ECG measurements?," *Journal of Electrocardiology*, vol. 47, no. 6, pp. 849–855, 2014.
- [21] A. Acharya and S. Kumar Choudhary, "Gender classification from ECG signal analysis using least square support vector machine," *American Journal of Signal Processing*, vol. 2, no. 5, pp. 145–149, 2012.
- [22] P. Nguyen, D. Tran, X. Huang, and W. Ma, "Age and gender classification using EEG paralinguistic features," in *Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1295–1298, IEEE, San Diego, CA, USA, November 2013.
- [23] D. Q. Phung, D. Tran, W. Ma, P. Nguyen, and T. Pham, "Using shannon entropy as EEG signal feature for fast person identification," in *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence*, pp. 413–418, ESANN, Bruges, Belgium, April 2014.
- [24] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 832–839, 2007.
- [25] F. Vecchio, C. Babiloni, R. Lizio et al., "Resting state cortical EEG rhythms in Alzheimer's disease: toward EEG markers for clinical applications: a review," *Supplements to Clinical Neurophysiology*, vol. 62, pp. 223–236, 2012.
- [26] 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.
- [27] D. Abásolo, R. Hornero, P. Espino, D. Álvarez, and J. Poza, "Entropy analysis of the EEG background activity in Alzheimer's disease patients," *Physiological Measurement*, vol. 27, no. 3, pp. 241–253, 2006.
- [28] N. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. Ahmad, M. Islam, and J. Escudero, "Automatic artifact removal in EEG of normal and demented individuals using ICA-WT during

- working memory tasks,” *Sensors*, vol. 17, no. 6, Article ID 1326, 2017.
- [29] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, “Discrimination of stroke-related mild cognitive impairment and vascular dementia using EEG signal analysis,” *Medical, & Biological Engineering & Computing*, vol. 56, pp. 137–157, 2017.
- [30] M. Murugappan, N. Ramachandran, and Y. Sazali, “Classification of human emotion from EEG using discrete wavelet transform,” *Journal of Biomedical Science and Engineering*, no. 4, pp. 390–396, 2010.
- [31] R. Yuvaraj, M. Murugappan, N. Mohamed Ibrahim et al., “s disease during emotion processing,” *Behavioral and Brain Functions: BBF*, vol. 10, no. 1, p. 12, 2014.
- [32] N. K. Al-Qazzaz, S. H. B. Ali, S. A. Ahmad, K. Chellappan, M. S. Islam, and J. Escudero, “Role of EEG as biomarker in the early detection and classification of dementia,” *Science World Journal*, vol. 2014, Article ID 906038, 16 pages, 2014.
- [33] N. Thammasan, K. Moriyama, K.-i. Fukui, and M. Numao, “Continuous music-emotion recognition based on electroencephalogram,” *IEICE - Transactions on Info and Systems*, vol. E99.D, no. 4, pp. 1234–1241, 2016.
- [34] O. Sourina, Y. Liu, and M. K. Nguyen, “Real-time EEG-based emotion recognition for music therapy,” *Journal on Multimodal User Interfaces*, vol. 5, no. 1-2, pp. 27–35, 2012.
- [35] N. K. Al-Qazzaz, S. Ali, M. S. Islam, S. A. Ahmad, and J. Escudero, “EEG markers for early detection and characterization of vascular dementia during working memory tasks,” in *Proceedings of the Biomedical Engineering and Sciences (IECBES), 2016 IEEE EMBS Conference on*, pp. 347–351, IEEE, Kuala Lumpur, Malaysia, December 2016.
- [36] N. K. Al-Qazzaz, S. H. M. Ali, S. Islam, S. A. Ahmad, and J. Escudero, “EEG wavelet spectral analysis during a working memory tasks in stroke-related mild cognitive impairment patients,” in *Proceedings of the International Conference for Innovation in Biomedical Engineering and Life Sciences*, pp. 82–85, Springer, Putrajaya, Malaysia, December 2016.
- [37] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, “Entropy-based markers of EEG background activity of stroke-related mild cognitive impairment and vascular dementia patients,” in *Proceedings of the 2nd International Conference on Sensors Engineering and Electronics Instrumental Advances*, SEIA, Barcelona, Spain, September 2016.
- [38] C. W. Yean, W. Khairunizam, M. I. Omar et al., “Emotional states analyze from scaling properties of EEG signals using hurst exponent for stroke and normal groups, Lecture Notes in Mechanical Engineering,” in *Proceedings of the Symposium on Intelligent Manufacturing and Mechatronics*, pp. 526–534, Springer, 2019.
- [39] O. Sourina and Y. Liu, “A fractal-based algorithm of emotion recognition from EEG using arousal-valence model,” in *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing-BIOSIGNALS*, pp. 209–214, Rome, Italy, January 2011.
- [40] W. Choong, W. Khairunizam, M. Omar et al., “Eeg-based emotion assessment using detrended flunctuation analysis (dfa),” *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 10, no. 1–13, pp. 105–109, 2018.
- [41] B. García-Martínez, A. Martínez-Rodrigo, R. Zangróniz Cantabrana, J. Pastor García, and R. Alcaraz, “Application of entropy-based metrics to identify emotional distress from electroencephalographic recordings,” *Entropy*, vol. 18, no. 6, p. 221, 2016.
- [42] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, “Identifying stable patterns over time for emotion recognition from EEG,” *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2017.
- [43] N. K. Al-Qazzaz, S. H. M. Ali, and S. A. Ahmad, “Differential evolution based channel selection algorithm on EEG signal for early detection of vascular dementia among stroke survivors,” in *Proceedings of the IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 239–244, IEEE, Sarawak, Malaysia, December 2018.
- [44] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, and J. Escudero, “Optimal EEG channel selection for vascular dementia identification using improved binary gravitation search algorithm,” in *Proceedings of the International Conference for Innovation in Biomedical Engineering and Life Sciences*, pp. 125–130, Springer, Penang, Malaysia, December 2017.
- [45] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, and J. Escudero, “Classification enhancement for post-stroke dementia using fuzzy neighborhood preserving analysis with QR-decomposition,” in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3174–3177, IEEE, Jeju South Korea, July 2017.
- [46] A. Freeman, P. Stanko, L. N. Berkowitz et al., “Inclusion of sex and gender in biomedical research: survey of clinical research proposed at the University of Pennsylvania,” *Biology of Sex Differences*, vol. 8, no. 1, p. 22, 2017.
- [47] A. Thul, J. Lechinger, J. Donis et al., “EEG entropy measures indicate decrease of cortical information processing in Disorders of Consciousness,” *Clinical Neurophysiology*, vol. 127, no. 2, pp. 1419–1427, 2016.
- [48] J. Tian and Z. Luo, “Motor imagery EEG feature extraction based on fuzzy entropy,” *Journal of Huazhong University of Science and Technology*, vol. 41, pp. 92–95, 2013.
- [49] Y. Cao, L. Cai, J. Wang et al., “Characterization of complexity in the electroencephalograph activity of Alzheimer’s disease based on fuzzy entropy,” *Chaos*, vol. 25, no. 8, Article ID 083116, 2015.
- [50] H. Azami, A. Fernández, and J. Escudero, “Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis,” *Medical, & Biological Engineering & Computing*, vol. 55, no. 11, pp. 2037–2052, 2017.
- [51] S. M. Pincus, “Approximate entropy as a measure of system complexity,” *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [52] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [53] W. Chen, J. Zhuang, W. Yu, and Z. Wang, “Measuring complexity using fuzzyen, apen, and sampen,” *Medical Engineering & Physics*, vol. 31, no. 1, pp. 61–68, 2009.
- [54] E. Reyes-Sanchez, A. Alba, M. Mendez, G. Milioli, and L. Parrino, “Spectral entropy analysis of the respiratory signal and its relationship with the cyclic alternating pattern during sleep,” *International Journal of Modern Physics C*, vol. 27, no. 12, Article ID 1650140, 2016.
- [55] M. Murugappan, “Human emotion classification using wavelet transform and KNN,” in *Proceedings of the International Conference on Pattern Analysis and Intelligence Robotics*, pp. 148–153, IEEE, Kuala Lumpur, Malaysia, June 2011.
- [56] P. Shen, Z. Changjun, and X. Chen, “Automatic speech emotion recognition using support vector machine,” in *Proceedings of the Electronic and Mechanical Engineering and*

- Information Technology (EMEIT), 2011 International Conference on*, pp. 621–625, IEEE, Harbin, China, August 2011.
- [57] Y. Pan, P. Shen, and L. Shen, “Speech emotion recognition using support vector machine,” *International Journal of Smart Home*, vol. 6, no. 2, pp. 101–108, 2012.
- [58] B. Kaur, D. Singh, and P. P. Roy, “Age and gender classification using brain-computer interface,” *Neural Computing & Applications*, vol. 31, no. 10, pp. 5887–5900, 2019.
- [59] H. Shahabi and S. Moghimi, “Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity,” *Computers in Human Behavior*, vol. 58, pp. 231–239, 2016.
- [60] K.-E. Ko, H.-C. Yang, and K.-B. Sim, “Emotion recognition using EEG signals with relative power values and Bayesian network,” *International Journal of Control, Automation and Systems*, vol. 7, no. 5, pp. 865–870, 2009.
- [61] M. Murugappan, R. Nagarajan, and S. Yaacob, “Combining spatial filtering and wavelet transform for classifying human emotions using EEG Signals,” *Journal of Medical and Biological Engineering*, vol. 31, no. 1, pp. 45–51, 2011.
- [62] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [63] N. K. Al-Qazzaz, M. K. Sabir, S. H. B. M. Ali, S. A. Ahmad, and K. Grammer, “Electroencephalogram profiles for emotion identification over the brain regions using spectral, entropy and temporal biomarkers,” *Sensors*, vol. 20, no. 1, p. 59, 2020.
- [64] X. Jie, R. Cao, and L. Li, “Emotion recognition based on the sample entropy of EEG,” *Bio-Medical Materials and Engineering*, vol. 24, no. 1, pp. 1185–1192, 2014.
- [65] J. Rottenberg, J. J. Gross, F. H. Wilhelm, S. Najmi, and I. H. Gotlib, “Crying threshold and intensity in major depressive disorder,” *Journal of Abnormal Psychology*, vol. 111, no. 2, pp. 302–312, 2002.
- [66] D. Abásolo, J. Escudero, R. Hornero, C. Gómez, and P. Espino, “Approximate entropy and auto mutual information analysis of the electroencephalogram in Alzheimer’s disease patients,” *Medical, & Biological Engineering & Computing*, vol. 46, no. 10, pp. 1019–1028, 2008.
- [67] S. R. Messer, J. Agzarian, and D. Abbott, “Optimal wavelet denoising for phonocardiograms,” *Microelectronics Journal*, vol. 32, no. 12, pp. 931–941, 2001.
- [68] Z. German-Sallo and C. Ciufudean, “Waveform-adapted wavelet denoising of ECG signals,” *Advances in Mathematical and Computational Methods*, vol. 172175, 2012.
- [69] A. Shoeb and G. Clird, “Chapter 16—wavelets; multiscale Activity in physiological signals,” *Biomedical Signal and Image Processing*, 2005.
- [70] N. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. Ahmad, M. Islam, and J. Escudero, “Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task,” *Sensors*, vol. 15, no. 11, pp. 29015–29035, 2015.
- [71] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *Annals of Statistics*, pp. 1135–1151, 1981.
- [72] R. Romo-Vazquez, R. Ranta, V. Louis-Dorr, and D. Maquin, “EEG ocular artefacts and noise removal,” in *Proceedings of the Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 5445–5448, IEEE, Lyon, France, August 2007.
- [73] E. Estrada, H. Nazeran, G. Sierra, F. Ebrahimi, and S. K. Setarehdan, “Wavelet-based EEG denoising for automatic sleep stage classification,” in *Proceedings of the Electrical Communications and Computers (CONIELECOMP), 2011 21st International Conference on*, pp. 295–298, IEEE, San Andres Cholula, Mexico, March 2011.
- [74] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, M. S. Islam, and M. I. Ariff, “Selection of mother wavelets thresholding methods in denoising multi-channel EEG signals during working memory task,” in *Proceedings of the Biomedical Engineering and Sciences (IECBES), 2014 IEEE Conference on*, pp. 214–219, IEEE, Kuala Lumpur, Malaysia, December 2014.
- [75] J. Jeong, “Nonlinear dynamics of EEG in Alzheimer’s disease,” *Drug Development Research*, vol. 56, no. 2, pp. 57–66, 2002.
- [76] 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.
- [77] N. K. Al-Qazzaz, M. K. Sabir, S. H. B. M. Ali, S. A. Ahmad, and K. Grammer, “Multichannel optimization with hybrid spectral-entropy markers for gender identification enhancement of emotional-based EEGs,” *IEEE Access*, vol. 9, 2021.
- [78] N. Al-Qazzaz, M. Sabir, S. Ali, S. A. Ahmad, and K. Grammer, “Electroencephalogram Brain Mapping for revealing the emotional changes over the brain regions using Entropy biomarker,” *Sensors*, vol. 20, no. 1, p. 59, 2021.
- [79] N. K. Al-Qazzaz, M. K. Sabir, and K. Grammer, “Gender differences identification from brain regions using spectral relative powers of emotional EEG,” in *Proceedings of the 7th International Work-Conference on Bioinformatics and Biomedical Engineering*, IWBBIO, Granada, Spain, May 2019.
- [80] S. Madan, K. Srivastava, A. Sharmila, and P. Mahalakshmi, “A case study on Discrete Wavelet Transform based Hurst exponent for epilepsy detection,” *Journal of Medical Engineering & Technology*, vol. 42, no. 1, pp. 9–17, 2018.
- [81] H. Azami and J. Escudero, “Amplitude-aware permutation entropy: illustration in spike detection and signal segmentation,” *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 40–51, 2016.
- [82] K. Natarajan, R. Acharya U, F. Alias, T. Tiboleng, and S. K. Puthusserypady, “Nonlinear analysis of EEG signals at different mental states,” *BioMedical Engineering Online*, vol. 3, no. 1, p. 7, 2004.
- [83] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [84] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *’95: Proceedings of the 14th international joint conference on Artificial intelligence*, pp. 1137–1145, Montreal Quebec, Canada, August 1995.
- [85] Y. Song and J. Zhang, “Discriminating preictal and interictal brain states in intracranial EEG by sample entropy and extreme learning machine,” *Journal of Neuroscience Methods*, vol. 257, pp. 45–54, 2016.
- [86] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlington, MA, USA, 2005.
- [87] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software,” *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [88] N. K. Al-Qazzaz, S. H. M. Ali, and S. A. Ahmad, “Comparison of the effectiveness of AICA-WT technique in discriminating vascular dementia EEGs,” in *Proceedings of the 2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, pp. 109–112, IEEE, Kuching, Malaysia, July 2018.

Research Article

Stress Classification by Multimodal Physiological Signals Using Variational Mode Decomposition and Machine Learning

Nilima Salankar ¹, Deepika Koundal ¹ and Saeed Mian Qaisar ^{2,3}

¹Department of Virtualization, School of Computer Science, University of Petroleum & Energy Studies (UPES), Dehradun, India

²Electrical and Computer Engineering Department, Effat University, Jeddah 22332, Saudi Arabia

³Communication and Signal Processing Lab, Energy and Technology Research Centre, Effat University, Jeddah 22332, Saudi Arabia

Correspondence should be addressed to Saeed Mian Qaisar; sqaisar@effatuniversity.edu.sa

Received 6 June 2021; Revised 22 July 2021; Accepted 12 August 2021; Published 26 August 2021

Academic Editor: Siti Anom Ahmad

Copyright © 2021 Nilima Salankar 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 pandemic situation, importance and awareness about mental health are getting more attention. Stress recognition from multimodal sensor based physiological signals such as electroencephalogram (EEG) and electrocardiography (ECG) signals is a very cost-effective way due to its noninvasive nature. A dataset, recorded during the mental arithmetic task, consisting of EEG + ECG signals of 36 participants is used. It contains two categories of performance, namely, “Good” (nonstressed) and “Bad” (stressed) (Gupta et al. 2018 and Eraldeir et al. 2018). This paper presents an effective approach for the recognition of stress marker at frontal, temporal, central, and occipital lobes. It processes the multimodality physiological signals. The variational mode decomposition (VMD) strategy is used for data preprocessing and for the decomposition of signals into various oscillatory mode functions. Poincare plots (PP) are derived from the first eight variational modes and features from these plots have been extracted such as mean, area, and central tendency measure of the elliptical region. The statistical significance of the extracted features with $p < 0.5$ has been performed using the Wilcoxon test. The multilayer perceptron (MLPN) and Support Vector Machine (SVM) algorithms are used for the classification of stress and nonstress categories. MLPN has achieved the maximum accuracies of 100% for frontal and temporal lobes. The suggested method can be incorporated in noninvasive EEG signal processing based automated stress identification systems.

1. Introduction

Short-term mental fatigue results in reduced efficiency in workspace, whereas long-term mental fatigue may result into brain damage. Therefore, timely awareness about reasonable rate of mental fatigue is very crucial. Stress management is very necessary for successful and happy leading life. The population who can easily manage stress does exhibit their behavior in brain as well and can be easily captured by noninvasive approach. Acquisition of data in real time environment is very tough; thus, induced technique plays a very important role in behavioral study. Three levels of mental arithmetic tasks that are mostly used in literature are easy, medium, and hard and it is an appropriate technique for inducing the stress in virtual environment [1].

Nowadays, in every work domain and culture, performance setting is marked and stress management is a key to succeed and nonmanagement of stress not only leads to failure but is a major reason of depression, frustration, and negative approach towards life. Thus, stress management is an important skill to learn and to help this fraternity by early identification of markers which is very essential.

In literature, lot of attempts have been made in this direction. Firstly, in order to understand the markers of fraternity stress and nonstress, various types of studies have been conducted where arithmetic test is the most common technique which has been widely adopted. However, mental arithmetic task is an appropriate stress inducing technique but it has certain limitation and according to age it has shown different impact [2]. To capture the response from

different types of subjects in response to various stress induced techniques, multimodality approach has exhibited significant role and its impact is noteworthy [3]. Various types of studies have been carried out in this domain and quite interesting study has been carried out on the arithmetic task where experimentation has been done on numbers with respect to ordered and unordered [4]. As brain has various lobes and it gives response to various activities, sometimes, it might be subject dependent and independent and this correlation has been identified by studying different regions of brain while performing activity abacus [5] and hemisphere activation has been studied [6]. Thinking also acts as a vital role in the generation of brain signals; thus, silent text reading and study of brain signals in response to silent text reading have been carried out [7–9].

For capturing the physiological signals in response to induced techniques, electroencephalogram (EEG), electromyography (EMG), electrooculography (EOG), and electrocardiography (ECG) have mostly been used whereas photoplethysmography (PPG) also serves a very important role in capturing physiological signals [3, 10]. Single channel study though focuses on single task whereas multimodal ones focus and capture the responses from various parts; thus, multimodality signals and their correlation with mental workload have been studied [11, 12]. Every significant activity contains the significance of various brain regions and its association details are presented [13]. Even though behavioral impact is available all over, certain lobes have shown prominent changes; majorly frontal region exhibits prominent changes in signals while performing mental arithmetic task [14]. Data acquisition by inducing relevant technique/protocol and its systematic analysis lead to designing of an appropriate study protocol and popularly used methods are appropriate selection of channels [15]. Analysis of an EEG signal from only visualization is an empirical science and needs expertise in neurological domain and thus, it is very time-consuming and tedious process [16]. As there are stress induced techniques, similarly stress relieving techniques and their cognitive impact are also great to understand [17]. Stress has significant impact to invoke and subsidize emotions [18] and generate different behavioral response in epilepsy patients [19, 20]. Physiological signals also help to learn behavioral pattern in special category [21]. In order to better visualize signals decomposition method, feature extraction plays very crucial role. Decomposition of signals in empirical way has been widely used [22], but high frequency study is well supported by VMD; thus authors have proposed the use of VMD technique to better visualize the signal in time domain while retaining its frequency components. This method has proven its significant role in area of seizure [23, 24]. Understanding of connectivity among brain regions gives clear insight about origin and exit of electrical connectivity between regions [20]. However, convolution techniques have been used [21] for reading EEG signals but they have no flexibility of reading signals in time-frequency domain and sometimes because of nonstationary behavior of the EEG signal they need to compromise on accuracy [25]. There are unlimited areas where stress gets evoked and reason for it could be

noise trigger or unpleasant vision [26]. As it is said every task is time bounded and it is proven in studying correlation of activity and time in [24]. Mere clean data acquisition does not solve the purpose unless relevant features have been extracted and its importance is viable [1, 23]. Capturing of signal from throughout brain region is very tedious; thus, study of only frontal region has been carried out in [27]. EEG signals are very effective in carrying correlation between various rhythmic signals [28]. Activation of specific region and band is dependent upon types of activities performed by subjects [29, 30]. Because of scarcity of professional automation and semiautomation, the analysis of multimodal signals such as EEG + ECG is very important [31]. In real life, time requirement to induce stress and analyze its resistant capability has various constraints [32–34].

The main objective of this work is to extract relevant features from the multimodality physiological signals and to design a classifier which can easily detect the stress (bad) and nonstress (good) performer where signals have been captured while subjects have performed silent math activity as well as during getting acquainted with an environment. Therefore, key contributions of this research work are as follows:

- (i) Proposing an effective method for the automated classification of stress resistant capability while conducting short time mental arithmetic task.
 - (1) The VMD is used for decomposing the multimodal physiological signals.
 - (2) The PPs plots are derived from the first eight variational modes.
 - (3) The discriminating features have been identified such as area, mean, and central tendency measure from each PP.
 - (4) The extracted features are passed to the considered classifiers for automated identification of good and bad classes.

Remainder of the paper is organized as follows: Section 2 describes materials and methods, Section 3 discusses the results, and conclusion is presented in Section 4.

2. Materials and Methods

This section has contributed for the discussion of methodology which consists of four components: (i) description of dataset used for an experimentation, (ii) selection of channels for an experiment purpose, (iii) VMD, (iv) PPs and features extraction, (v) classifiers, and (vi) evaluation measures.

The suggested work flow is shown in Figure 1. Different stages are described in the following subsections.

2.1. Dataset. In this work, dataset used for the purpose of an experimentation and evaluation of stress classifier is available publicly [15]. For each subject, two trails have been conducted where physiological signals (EEG + ECG) have been captured. Trail 1 is baseline activity for 180 seconds to get subjects acquainted with an

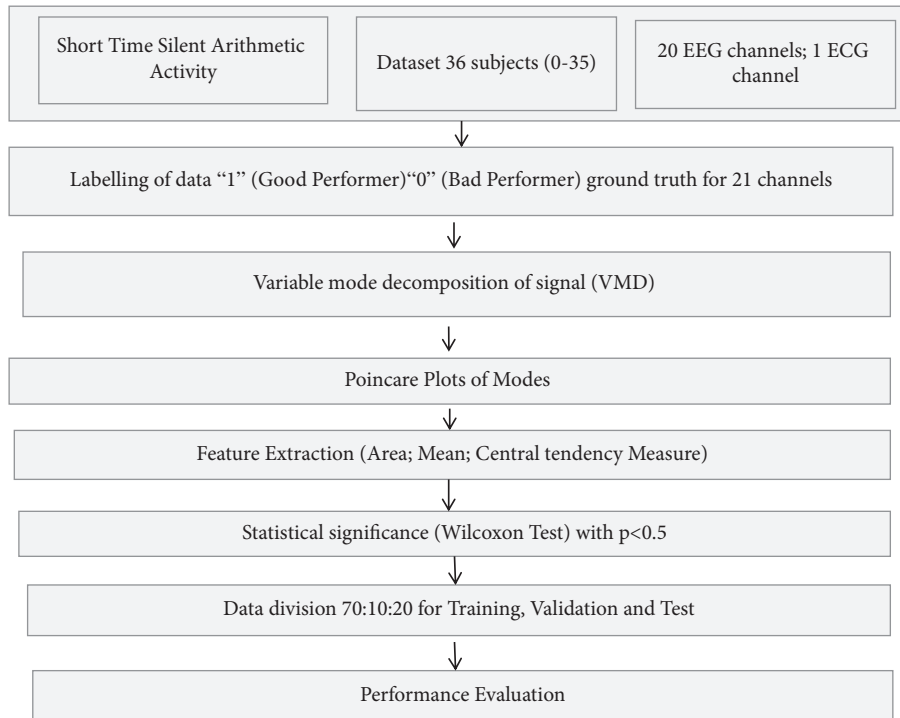


FIGURE 1: The overall workflow of proposed method.

environment and 60 seconds' data for actual cognitive task performance. Sampling rate for data acquisition is 516 Hz. Thus, each subject cognitive task has raw file with data dimension (21 channels \times 516 samples \times time duration results) into (21 \times 516 \times 60 = 650160) data points and (21 \times 516 \times 180 = 1950480) data points for baseline activity. Independent component analysis (ICA) has been done for the purpose of noise removal. However, dataset is captured while subjects were performing silent math activity where no muscle movement was expected but noise generally get introduced and elimination of artifacts like eyes movement and cardiac overlapping of cardiac position has been done. Tasks which was performed by the subjects include subtraction of two numbers without making any movement. Each trial has been commenced with the communication orally of 2-digit (subtrahend) and 4-digit (minuend) numbers like 42 and 3141, respectively. More details of the dataset can be found in [15]. As a ground truth labelling of dataset is done on the basis of performance report card which is available in the form of excel sheet with dataset. During data acquisition subjects were asked to perform arithmetic activity and nonstressed performer has performed 21 subtractions approximately and stressed performer has performed 7 subtractions in the given time. Number of nonstress performers in dataset is 26 and that of stress performers was 10. Dummy participants have been added to the dataset for nonstress category by replicating the data channelwise for an experimentation purpose with goal of balancing data for both categories. Age group of subjects is in range of 16–26; both male and female categories were included.

The total data of 36 subjects has been evaluated and has been given to the classifier by a robust and appropriate feature extraction approach. For labels in documentation for dataset [1], in the performance report of subjects with notation subject 0 to subject 35 which consist of name, age, gender, number of subtractions, and count quality “G” indicates good and “B” indicates bad, as mentioned in the excel file. Joining subject file with EEG data and name attribute created a labelled dataset file and those labels have been used as a ground truth for binary classification.

2.1.1. Channel Selection. Channels included into study cover complete skull hemisphere ranging from frontal to occipital region: frontal position, 6 + 1 channels; temporal position, 4 channels; central position, 2 + 1 channels; parietal position, 2 + 1 channels; occipital position, 2 channels; behind the ear, 1 channel; and 1 ECG channel. Channel selection is done with objective in mind to investigate the affected areas so that precise marker can be identified in each category which definitely helps for identification of discrimination purpose of stress and nonstress biomarkers. Channel positions considered are illustrated in Figure 2.

2.2. Variational Mode Decomposition (VMD). This decomposition method is robust for noise handling [35]. It is a process of decomposition of real valued input signal f into discrete number of subsignals also known as mode u_k . Each mode is densely oriented towards its central frequency w_k , which is determined during decomposition process. Each mode has a sparsity property which is being used while reconstructing the signal. Before decomposition sparsity of

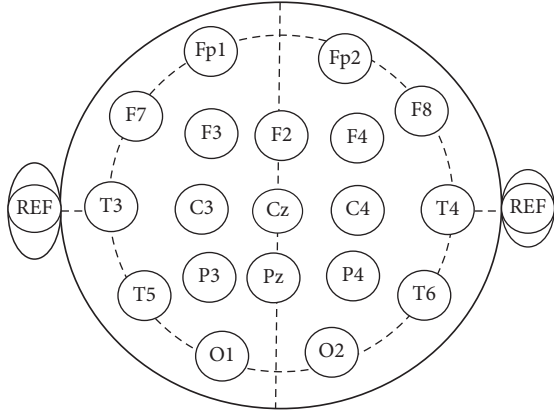


FIGURE 2: Channel positions considered in study: F =Frontal, C=Central, P=Parietal, O=Occipital, and T = Temporal.

each mode has to be determined by bandwidth in spectral domain. To identify the bandwidth, three steps need to be followed which are given as follows:

- (1) From each mode, Hilbert transform has been used to obtain the unilateral frequency domain.
- (2) Subsequently, shift the frequency spectrum which has been obtained in step 1 to the baseband by integrating with an exponential tune to the respective centre frequency.
- (3) Afterwards, apply *H1* Gaussian smoothness to the demodulated signal for obtaining bandwidth given in

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_k \partial_t \left[\partial(t) + \frac{j}{\pi t} * u_k(t) \right] e^{-jw_k t} \right\}, \quad (1)$$

where $u_k = \{u_1, \dots, u_k\}$ and $w_k = \{w_1, \dots, w_k\}$ $\sum_k u_k = f$.

The detailed behavior of variational mode decomposition (VMD) is as shown in Algorithm 1.

2.3. Poincare Plots (PPs). PP of the modes obtained after VMD of EEG signals can provide favourable characteristic patterns for the classifications purpose. The PP of signal is defined as given in the equation below:

$x(t)$ is a plotting of $X(t)$ against $Y(t)$ as shown in the following equations:

$$X(t) = x(t+1) - x(t), \quad (2)$$

$$Y(t) = x(t+2) - x(t+1). \quad (3)$$

This plot indicates the successive proportions against each other [36]. The resultant elliptical shape of plots portrayed from mode signifies the strong positive association between the consecutive data points, variability, and stochastic nature. As modes derived from signal are deviated towards central frequency, asymmetric area coverage is visible in PPs. Region coverage for the plot of the first eight modes is significantly higher than that of the other modes.

Ten modes are obtained from the signal and evaluated but only first eight modes are considered for experimentation and the remaining two are excluded on the basis of significant area coverage as compared to the rest at central modes.

2.4. Features Extraction. Owing to the no stationary behavior of captured physiological signals, features obtained from the PPs are the area of the elliptic region, mean distance, and central tendency measure. PPs are designed from each mode obtained from variational mode decomposition, and for the calculation of area coverage of the elliptic shape of plots, the equations used are as follows:

$$S_X = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} X(t)^2}, \quad (4)$$

$$S_Y = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} Y(t)^2}, \quad (5)$$

$$S_{XY} = \frac{1}{N} \sum X(t)Y(t), \quad (6)$$

$$\text{Area}_{\text{Ellipse}} = \pi ab, \quad (7)$$

where a = major radius of poicare plots ($\sqrt{S_X^2 + S_Y^2 + D}$),
 b = minor radius of poicare plots ($\sqrt{S_X^2 + S_Y^2 - D}$), and
 $D = \sqrt{(S_X^2 + S_Y^2)4(S_X^2 S_Y^2 + S_{XY}^2)}$.

Area computed from the PPs of EEG and ECG signals is used as a discriminative feature with 95% confidence. The details of area computation includes calculation of mean values $X(t)$ and $Y(t)$ as mentioned in equations (2) and (3). To compute plot parameter D compute mean values from equations (4)–(6). Ellipse area can be computed by equation (7). Mean and central tendency measure (CTM) of the PPs have been derived.

2.5. Classifier Used in Study. For the analysis of the robustness of the proposed approach, two classifiers have been tested which are multilayer perceptron neural network (MLPN) and support vector machine (SVM). First experimentation has been carried out with SVM. It is of category supervised learning and can be used for the purpose of data classification either at binary or at multiclass type. Each data point is indicated as n_i and requires n -dimensional space for plotting all data points in consideration. For classification purpose objective of an algorithm is to find an appropriate hyperplane where discrimination and segregation of correct data points are possible. In the condition where classification is not easily achievable kernelling is an option opted by SVM, which is a method to elevate lower dimensional input space to higher level. Unlike SVM MLPN has a capability of performing more complex operations and it had a tendency

```

Initialize  $\{\hat{u}_k^1\}, \{\hat{w}_k^1\}, \hat{\lambda}^1, n \leftarrow 0$ 
Repeat
 $n \leftarrow n + 1$ 
for  $k = 1: K$  do
update  $\hat{u}_k$  for all  $w \geq 0$ :  $\hat{u}_k^{(n+1)}(w) \leftarrow (\hat{f}(w) - \sum_{i < k} \hat{u}_i^{(n+1)}(w) - \sum_{i > k} \hat{u}_i^n(w) + \hat{\lambda}^n(w)/2/1 + 2\alpha(w - w_k^n)^2)$ 
Update  $w_{kb}, w_k^{(n+1)} \leftarrow (\int_0^\infty w |\hat{u}_k^{(n+1)}(w)|^2 dw) / \int_0^\infty |\hat{u}_k^{(n+1)}(w)|^2 dw$ 
End for
Dual ascent for all  $w \geq 0$ :  $\hat{\lambda}^{(n+1)}(w) \leftarrow (w) + \tau(\hat{f}(w) - \sum_k \hat{u}_k^{(n+1)}(w))$ 
Until convergence  $\sum_k \hat{u}_k^{(n+1)} - \hat{u}_{k_2}^{n_2} \hat{u}_{k_2}^{n_2} < \epsilon$ 

```

ALGORITHM 1: Variational mode decomposition (VMD).

to approximate the input as it is integrated with nonlinear activation function.

As data is not separable linearly all the time MLPN does the error correction by using a backpropagation approach where initial set weights and biased get corrected in order to reduce the difference level between obtained and expected results. It is very useful for image classification approach as well. Authors have converted the data into labelled version by using the ground truth value as described in Section 2. Different numbers of layers have been experimented and suitable layers fixed are 10 neurons and 3 hidden layers at each level. Activation function employed is rectified linear unit (ReLU), various learning parameters are set to regularization, and alpha is decided to be at value 0.0001 as most studies have reported this value and it deems fit for our experimentation purpose as well. Number of iterations is set to 200. Adam optimizer is used for an optimization purpose and

it has helped us to reach to the expected accuracy level for most of the subsets. In order to avoid very fast and very slow learning process we have selected the value 0.001 and kept it constant for experimentation. For SVM details the parameters selected are Cost C = 0.90, Epsilon = 0.1, RBF kernel = $\exp(-g|x - y|2)$, $g = 0.02$, Iteration limit = 100, and Numerical Tolerance = 0.0010.

2.6. Performance Evaluation Measures. Performance evaluators mostly used are sensitivity, specificity, and accuracy which gives insight about training, validation, and testing phase in order to compute the variance and bias level of the classifier. In this work, performance of the proposed classifier has been evaluated by using three evaluation metrics such as accuracy, specificity, and sensitivity as represented in the following equations:

$$\text{sensitivity (SEN)} = \frac{TP}{TP + FN} * 100, \quad (8)$$

$$\text{specificity (SPE)} = \frac{TN}{TN + FP} * 100, \quad (9)$$

$$\text{accuracy (ACC)} = \frac{TTP + TN}{TP + TN + FP + FN} * 100, \quad (10)$$

$$F1 \text{ measure} = \frac{TP}{TP + (1/2)(FP + FN)}, \quad (11)$$

$$\text{Kappa statistics} = \frac{(\text{percent agreement observed}) - (\text{percent agreement expected by chance alone})}{100 - (\text{percent expected by chance alone})}, \quad (12)$$

where TN and TP are the indicators to notify about how many data points have been correctly predicted by the classifier and FP and FN are the indicators to notify about how many data points have been incorrectly classified by the proposed classifier.

3. Results and Discussion

In this work, dataset used for the purpose of an experimentation and evaluation of stress classifier is available publicly [15]. These

are recordings of EEG signals while performing the mental arithmetic task of finite duration. Figure 3 shows the decomposition of signal by using VMD approach and its modes which are deviated towards the central frequency of the original input signal that proves the significance of an approach for better noise handling and its appropriateness for handling the lower frequency as compared to higher frequency components.

The PPs of first eight variational modes are shown in Figure 4. Plots for modes are clearly representing the varying area covered and it is more inclined towards lower frequency

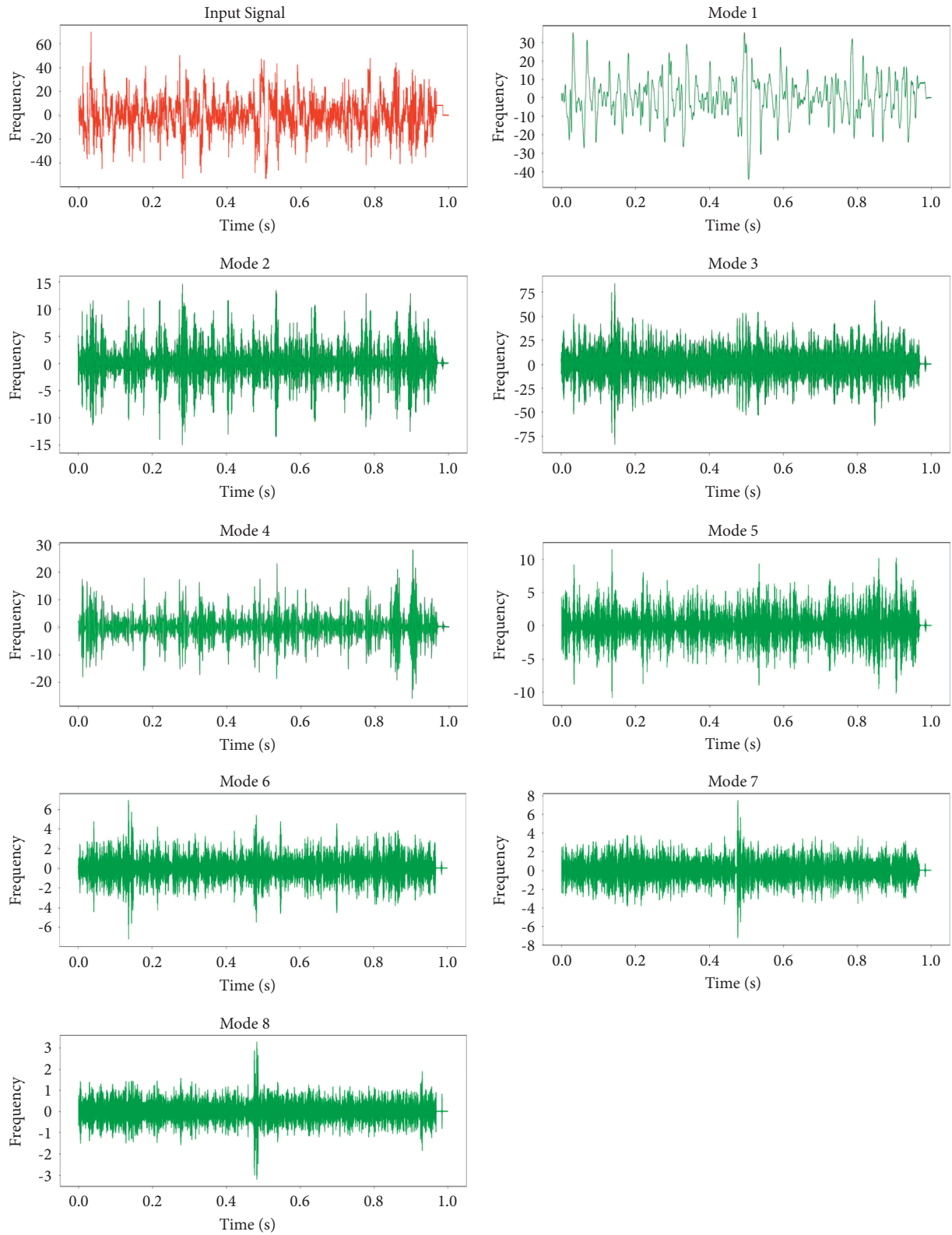


FIGURE 3: VMD of the signal while performing mental arithmetic activity of 1 min duration for subject 1 (female, good performer), channel FP1.

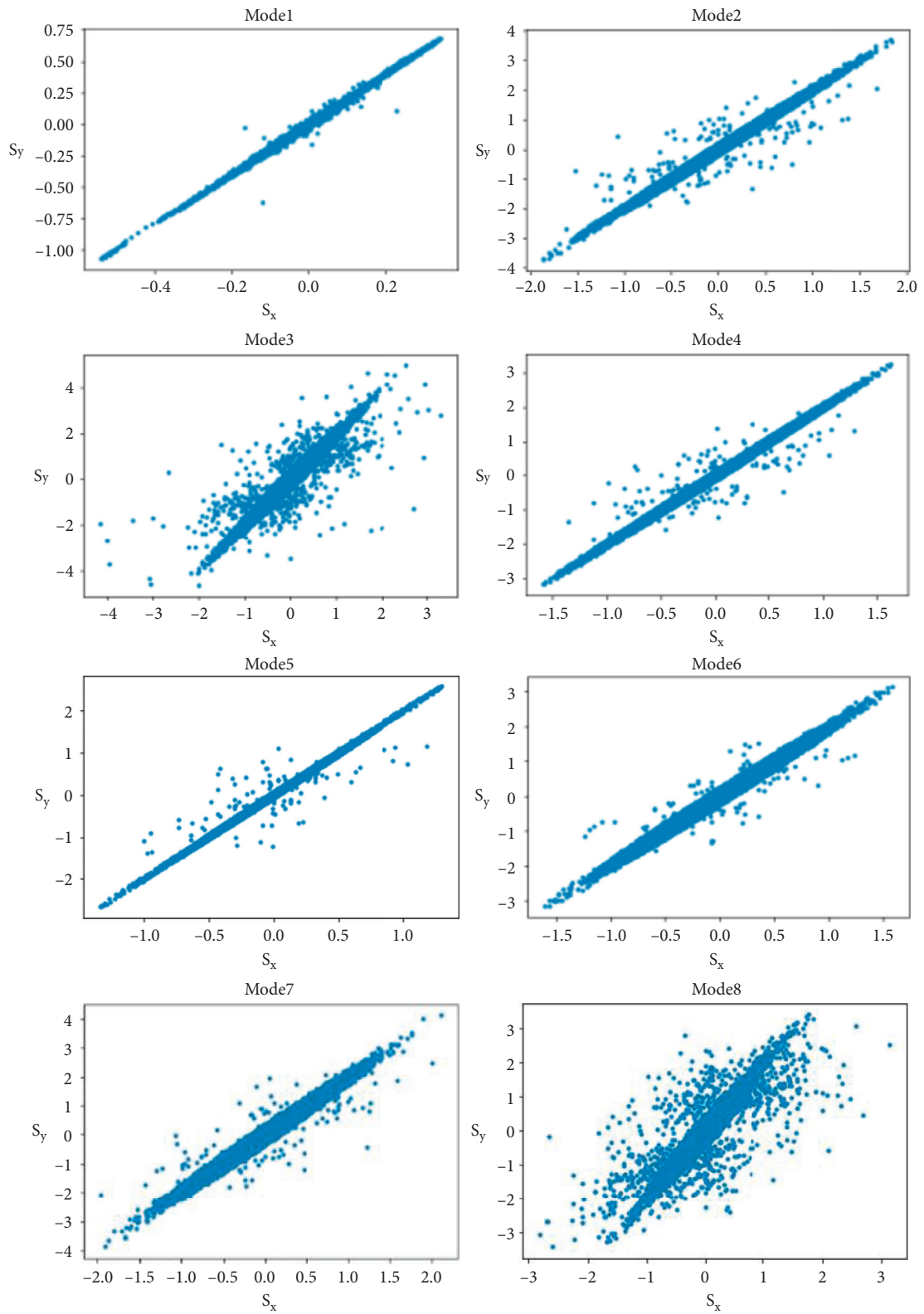


FIGURE 4: PPs of the modes while performing mental arithmetic activity of 1 min duration for subject 1 (female, good performer), channel FP1.

component mode as compared to higher frequency component mode. These are PPs of an EEG signal, for subject 1 (female, good performer) recorded for channel FP1, while performing mental arithmetic activity of 1 minute duration.

Features extracted described in Section 2, “Materials and Methods,” contributed for the designing of feature space. Extracted features from PPs contributed for the design of feature space and it is provided as an input to the SVM and multilayer perceptron neural network. Statistical significance of the extracted feature has been performed with Wilcoxon statistical significance test with $p < 0.5$. Extracted features are central tendency measure, mean distance, and area. The data generated for trail 1 and trail 2 for good performer female for channels FP1 and FP2 as listed in Table 1. Area coverage is the highest for middle modes as compared to initial and end position.

The Wilcoxon signed-rank test has been performed to verify the statistical significance of the features extracted with a confidence interval of 95 %.

To check robustness and more reliable performance evaluation of classifier data is provided to classifier in two ways for the first time, where 70%, 10%, 20% training, validation, and testing have been used by the split strategy to split the data and later k-fold data where value of $k = 10$ is selected. Only first eight modes have been used for the extraction of results. By considering all three features on 20% of test data from 36 subjects, the first to eight modes, binary classification results according to group selection of channels are listed in Table 2. Table 3 summarizes results for the case of k-fold cross validation.

Modes extracted from EEG signals have displayed variable frequency behavior and because of tendency towards central frequency modes are fluctuating from lower to higher and vice versa. The central mode exhibited higher frequency components in comparison to initial and ending. Initial signal is decomposed into 10 modes but after 8th mode it has stopped exhibiting any variation in behavior and appears to be static in nature. Thus, only first 8 modes have been considered for an experimentation purpose. Remaining 8 nodes also exhibited some grouping characteristics; thus 3 groups with combination of 1–4 modes, 5–8 modes, and 1–8 modes are created and processed accordingly. As each subject consists of 21 channels decomposition has been done subjectwise and trialwise. For the decomposition of signals into modes it took approximately 2–3 mins and for designing PPs of mode it took approximately 3–4 mins for each subject. The range mentioned is because of different size of data points for trial 1 and trial 2 as explained in Section 2. For classification purpose two algorithms have been used, SVM and MLPN. The time taken for SVM is much more as compared to MLPN for binary classification of different subsets. By using k-fold, performance of classifier is improved compared to the intended percentage split base study. It is particularly notable for modes 1–4 and 1–8 but marginal for modes 5–8 except for the first set, good performer vs. bad performer. The better performance, in the case of k-fold cross validation, is attained because a higher percentage of data is used for the training purpose compared to the percentage split case. Authors main objective were to

TABLE 1: Format of the feature map used in experimentation for channels FP1 and FP2 for trail 2.

Channel	Mode	Area	m ($r=0.5$)	ctm ($r=0.5$)	Target
EEG Fp1	1	2.53	0.93	0.19	1
EEG Fp1	2	37.39	0.20	0.26	1
EEG Fp1	3	63.25	0.13	0.29	1
EEG Fp1	4	11.44	0.31	0.25	1
EEG Fp1	5	4.78	0.41	0.24	1
EEG Fp1	6	48.82	0.16	0.28	1
EEG Fp1	7	12.03	0.28	0.28	1
EEG Fp1	8	33.35	0.22	0.26	1
EEG Fp2	1	3.19	0.85	0.21	1
EEG Fp2	2	4.70	0.56	0.24	1
EEG Fp2	3	22.77	0.26	0.25	1
EEG Fp2	4	27.51	0.28	0.24	1
EEG Fp2	5	27.43	0.23	0.25	1
EEG Fp2	6	22.22	0.24	0.26	1
EEG Fp2	7	29.60	0.21	0.28	1
EEG Fp2	8	1.77	0.42	0.27	1

identify specific region of brain which exhibits significant behavior and can be used as marker for discrimination and thus subsets have been designed accordingly.

First subset is good vs. bad (combination of 20 EEG and 1 ECG channel) and subsequently 4 subsets are at regions frontal, temporal, central, and occipital. In addition to these five subsets 2 more subsets were tested for good vs. bad male and female but have not achieved any good classification accuracy as per gender and thus concluded that gender discrimination for performance is not possible through the designed approach and needs some other approach for the same as what also happened in case of subset for parietal region and has been excluded from an experimentation. Extracted modes exhibited the inconsistent and abrupt upsurge and/or fall, which can be taken care of by detection and removal of outliers before determining the PPs. Elliptic nature of plots varies/diminishes for different modes but only those modes have been considered for which plots have shown good elliptic curve and the rest were excluded for an experimentation purpose. Area covered by plots emerges as a combination of real imaginary numbers and we treated it in a form of absolute version. The reason for complex nature emerge is correct as plots of equations (4)–(6) involve root function and it is quite possible that the root results into negative number which emerges as a complex number.

Dataset consists of 21 channels but for an experimentation purpose initially only frontal channels were considered and subsequently temporal, central, and parietal ones were evaluated. At the end 20 EEG channels and 1 ECG channel were considered for experimentation. The feature map consists of channel, trail, mode, area, mean, and CTM of dimension 8×3 for each channel and subject. The first eight modes and their plots were generated for each of the undertaken channels. Later 3 features have been extracted from each consecutive plot. Wilcoxon signed-rank test has been performed to validate the statistical significance of the derived features with $p < 0.5$. The plot area has been considerably reduced at initial and end position of mode, which is a sign of less frequency contents of the underlying signal.

TABLE 2: Classification results for the percentage split.

Sets	Statistical parameters	MPLN	SVM	MPLN	SVM	MPLN	SVM
		Modes 1-4	Modes 1-4	Modes 5-8	Modes 5-8	Modes 1-8	Modes 1-8
Good performer vs. bad performer	Sensitivity	97.2	61.23	98	68.67	98	70.65
	Specificity	96.3	65.12	98	68.67	98	70.65
	Accuracy	97.2	61.23	98	68.67	98	70.65
	F-measure	0.95	0.65	0.97	0.57	1	0.69
	Kappa statistics	0.94	0.63	0.98	0.56	0.99	0.69
Good performer vs. bad performer (female)	Sensitivity	78.56	62.34	81.67	68.67	83.78	69.67
	Specificity	78.9	62.37	82.23	68.67	83.67	69.67
	Accuracy	76.56	62.78	81.78	68.67	83.67	69.67
	F-measure	0.67	0.64	0.82	0.65	0.87	0.67
	Kappa statistics	0.68	0.63	0.82	0.65	0.88	0.69
Good performer vs. bad performer (male)	Sensitivity	79.56	63.43	80.21	68.67	82.67	70.02
	Specificity	78.78	63.56	80.12	68.67	82.56	70.02
	Accuracy	78.67	63.56	80.32	68.67	82.67	70.02
	F-measure	0.87	0.65	0.79	0.68	0.83	0.69
	Kappa statistics	0.87	0.65	0.79	0.67	0.83	0.69
Good performer vs. bad performer (frontal region)	Sensitivity	98	70.78	98.45	70.67	99	72.67
	Specificity	98	78.78	98.45	70.67	99	72.67
	Accuracy	98	67.89	98.45	70.67	99	72.67
	F-measure	0.98	0.74	0.99	0.69	1	0.71
	Kappa statistics	1	0.78	1	0.7	1	0.71
Good performer vs. bad performer (temporal region)	Sensitivity	99.8	67.78	99.99	61.23	99.99	75.56
	Specificity	97.78	67.78	99.99	65.12	99.99	75.56
	Accuracy	99.8	67.78	99.99	61.23	99.99	75.56
	F-Measure	0.99	0.74	0.99	0.59	1	0.72
	Kappa statistics	1	0.74	1	0.59	1	0.76
Good performer vs. bad performer (occipital region)	Sensitivity	78.56	60	80.23	62.34	89.34	76.7
	Specificity	78.9	60	80.34	62.37	89.43	76.7
	Accuracy	76.56	60	80.32	62.78	89.34	76.7
	F-measure	0.88	0.58	0.89	0.62	0.85	0.74
	Kappa statistics	0.85	0.58	0.86	0.62	0.85	0.74
Good performer vs. bad performer (central region)	Sensitivity	75.78	64.67	79.78	63.43	80.56	68.78
	Specificity	76.89	64.67	79.56	63.56	80.56	68.78
	Accuracy	78.67	64	79.34	63.56	80.56	68.78
	F-measure	0.76	0.67	0.75	0.62	0.83	0.56
	Kappa statistics	0.78	0.65	0.78	0.62	0.83	0.56

In comparison, the mean derived from the PPs has shown even rise and steadiness that can be visible in case of central tendency measure.

The experimental results are noticeable for approximately all the subsets utilized for the classification. Most stress related studies reported for specific/limited channels [37–39]. The proposed approach for the stress classification has outperformed other existing methods [3, 40–43] by achieving 100% accuracy with minimum time of mental arithmetic activity and has also given an insight for the identification of marker lobewise (frontal, temporal, central, and occipital) rather than selection of channels in a generalized way. Our brain activity consists of interchange of ions between neurons which results into current flow through synaptic mode. Stress and emotions either generated naturally or in induced environment have a tendency to retain in the form of current flow. EEG device is meant to measure the voltage fluctuation that occurs because of the movement of the neuron and as it has tendency to retain some time; this is very effective way

for the measure of positive and negative impact of any environmental situation on human brain [44–46]. Therefore, in this research work, an efficient and accurate classifier has been proposed with exceptional results for stress classification from EEG signals employing VMD, SVM, and multilayer perceptron. The maximum accuracy achieved at temporal and frontal lobe and in [47] was reported as category activation and discriminating area is observed at temporal lobe which is closely related with speech and nonspeech activity and as dataset [15] used study prototype which includes silent mental counting activity without any movement; the extracted results are relevant. Extracted results are more prominent at frontal and temporal region which is closely associated with concentration and focused mode of nature. The approach works nice for the intended dataset. In future the performance of the proposed method will be tested for other biomedical signals. The incorporation of event-driven methods can improve the performance of suggested solution in terms of computational effectiveness,

TABLE 3: Classification results for the k-fold cross validation.

Sets	Statistical parameters	K-fold	K-fold	K-fold	K-fold	K-fold	K-fold
		MPLN	SVM	MPLN	SVM	MPLN	SVM
		Modes 1-4		Modes 5-8		Modes 1-8	
Good performer vs. bad performer	Sensitivity	98.12	65.45	98.1	69	99.12	72.34
	Specificity	97	66	98	69	99	72.34
	Accuracy	97.56	63	98	69	99	72.34
	<i>F</i> -measure	0.98	0.7	0.97	0.59	1	0.7
	Kappa statistics	0.97	0.66	0.98	0.56	1	0.71
Good performer vs. bad performer (female)	Sensitivity	79.01	63.12	81.67	68.67	85.56	71.34
	Specificity	79.12	63.34	82.23	68.67	84	70
	Accuracy	77.45	63.45	81.78	68.67	84.02	71.78
	<i>F</i> -measure	0.69	0.68	0.82	0.65	0.89	0.68
	Kappa statistics	0.7	0.65	0.82	0.65	0.89	0.7
Good performer vs. bad performer (male)	Sensitivity	80.01	64.34	80.21	68.67	83	71
	Specificity	79.9	65.67	80.12	68.67	84	72
	Accuracy	79.12	64	80.32	68.67	84	71.34
	<i>F</i> -measure	0.89	0.65	0.79	0.68	0.86	0.72
	Kappa statistics	0.88	0.66	0.79	0.67	0.85	0.72
Good performer vs. bad performer (frontal region)	Sensitivity	98.23	71.01	98.45	70.67	100	75.23
	Specificity	98.45	79.23	98.45	70.67	100	73.23
	Accuracy	99.12	69	98.45	70.67	100	73.12
	<i>F</i> -Measure	0.99	0.76	0.99	0.69	1	0.72
	Kappa statistics	1	0.79	1	0.7	1	0.76
Good performer vs. bad performer (temporal region)	Sensitivity	99.82	68	99.99	61.23	100	76
	Specificity	98	68	99.99	65.12	100	78
	Accuracy	100	68	99.99	61.23	100	78
	<i>F</i> -measure	1	0.78	0.99	0.59	1	0.75
	Kappa statistics	1	0.76	1	0.59	1	0.78
Good performer vs. bad performer (occipital region)	Sensitivity	79.12	62.67	80.23	62.34	90	77
	Specificity	79	62	80.34	62.37	90	77
	Accuracy	77	61	80.32	62.78	90	77
	<i>F</i> -measure	0.9	0.6	0.89	0.62	0.87	0.78
	Kappa statistics	0.87	0.6	0.86	0.62	0.89	0.76
Good performer vs. bad performer (central region)	Sensitivity	76	65	79.78	63.43	82.12	69
	Specificity	77.78	65	79.56	63.56	82	69.12
	Accuracy	79.23	65	79.34	63.56	82	69
	<i>F</i> -measure	0.79	0.68	0.75	0.62	0.87	0.59
	Kappa statistics	0.79	0.66	0.78	0.62	0.86	0.59

compression, and power consumption [48–51]. Investigation of this approach is another prospect.

4. Conclusions

In this work, an attempt has been made to propose and explore the VMD approach and its Poincare plots for the classification of stress managing capability from the silent mental arithmetic activity. The VMD is a promising method for extracting the relevant features from the EEG + ECG signals. The resultant region of the Poincare plots has exhibited discriminating nature and varies widely for stress and nonstress category. Only the first six or seven modes provided the better classification accuracy for the comparative analysis of the stress. Signals accompanying with the activity have shown significant variability in comparison to the baseline activity for good performer while it has shown stability in case of bad performer and thus had a straightway more extensive influence on the Poincare plots. The area of

good performer female has been significantly higher. The devised method has achieved the maximum accuracies of 100% for frontal and temporal lobes. The proposed scheme can be beneficial for the clinical identification of low- and high-dominance regions in the subjects. In future scope, the proposed method can be extended to study the classification of other brain conditions such as epilepsy, Alzheimer's, and depression. Because of the identified marker in frontal and temporal lobe this approach can be used as promising approach to implement in real time situation.

Data Availability

The dataset used in this paper is publicly available at <https://physionet.org/content/eegmat/1.0.0/>.

Conflicts of Interest

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

References

- [1] C. Schiweck, D. Piette, D. Berckmans, S. Claes, and E. Vrieze, "Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. a systematic review," *Psychological Medicine*, vol. 49, no. 2, pp. 200–211, 2019.
- [2] M. G. Marciani, M. Maschio, F. Spanedda, C. Caltagirone, G. L. Gigli, and G. Bernardi, "Quantitative EEG evaluation in normal elderly subjects during mental processes: age-related changes," *International Journal of Neuroscience*, vol. 76, no. 1–2, pp. 131–140, 1994.
- [3] Y. Ding, Y. Cao, V. G. Duffy, Y. Wang, and X. Zhang, "Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning," *Ergonomics*, vol. 63, no. 7, pp. 896–908, 2020.
- [4] S. E. Vogel, N. Koren, S. Falb et al., "Automatic and intentional processing of numerical order and its relationship to arithmetic performance," *Acta Psychologica*, vol. 193, pp. 30–41, 2019.
- [5] C. Wang, T. Xu, F. Geng et al., "Training on abacus-based mental calculation enhances visuospatial working memory in children," *Journal of Neuroscience*, vol. 39, no. 33, pp. 6439–6448, 2019.
- [6] I. Papousek and G. Schuster, "Manipulation of frontal brain asymmetry by cognitive tasks," *Brain and Cognition*, vol. 54, 2004.
- [7] M. T. Arslan, S. G. Eraldemir, and E. Yıldırım, "Subject-dependent and subject-independent classification of mental arithmetic and silent reading tasks," *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, vol. 9, no. 3, pp. 186–195, 2017.
- [8] X. Yu, J. Zhang, D. Xie, J. Wang, and C. Zhang, "Relationship between scalp potential and autonomic nervous activity during a mental arithmetic task," *Autonomic Neuroscience*, vol. 146, no. 1–2, pp. 81–86, 2009.
- [9] L. Bréchet, D. Brunet, G. Birot, R. Gruetter, C. M. Michel, and J. Jorge, "Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with EEG and fMRI," *NeuroImage*, vol. 194, pp. 82–92, 2019.
- [10] H. Sharma, "Heart rate extraction from PPG signals using variational mode decomposition," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 1, pp. 75–86, 2019.
- [11] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *International Journal of Industrial Ergonomics*, vol. 35, no. 11, pp. 991–1009, 2005.
- [12] W. H. Guo, J. G. Zhang, X. F. Jiang, J. T. Shen, Z. J. Jia, and S. L. Xu, "The obsessive-compulsive brain 99mTc-ECD image analysis for different brain area with the functional extraction method," *Shanghai Medical Imaging*, vol. 15, no. 3, pp. 234–237, 2006.
- [13] T. Harmony, T. Fernández, J. Silva et al., "Do specific EEG frequencies indicate different processes during mental calculation?" *Neuroscience Letters*, vol. 266, 1999.
- [14] W. K. Y. So, S. W. H. Wong, J. N. Mak, and R. H. M. Chan, "An evaluation of mental workload with frontal EEG," *PLoS One*, vol. 12, no. 4, Article ID e0174949, 2017.
- [15] I. Zyma, S. Tukaev, I. Seleznev et al., "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, 2019.
- [16] A. Gupta, P. Singh, and M. Karlekar, "A novel signal modeling approach for classification of seizure and seizure-free EEG signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 925–935, 2018.
- [17] A. D. Duru and M. Assem, "Investigating neural efficiency of elite karate athletes during a mental arithmetic task using EEG," *Cognitive Neurodynamics*, vol. 12, no. 1, pp. 95–102, 2018.
- [18] N. Salankar, P. Mishra, and L. Garg, "Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot," *Biomedical Signal Processing and Control*, vol. 65, Article ID 102389, 2021.
- [19] N. Salankar, "Functional connectivity and classification of actual and imaginary motor movement," *International Journal of Engineering and Advanced Technology*, vol. 9, 2019.
- [20] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 1, pp. 31–42, 2019.
- [21] P. W. Fuller, "Attention and the EEG alpha rhythm in learning disabled children," *Journal of Learning Disabilities*, vol. 11, no. 5, pp. 303–312, 1978.
- [22] P. Gaur, K. McCreddie, R. B. Pachori, H. Wang, and G. Prasad, "An automatic subject specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation," *Biomedical Signal Processing and Control*, vol. 68, Article ID 102574, 2021.
- [23] S. G. Eraldemir, M. T. Arslan, and E. Yildirim, "Investigation of feature selection algorithms on A cognitive task classification: a comparison study," *Balkan Journal of Electrical and Computer Engineering*, vol. 6, no. 2, pp. 31–36, 2018.
- [24] S.-Y. Cheng, "Evaluation of effect on cognition response to time pressure by using EEG," *Advances in Intelligent Systems and Computing*, Springer, Berlin, Germany, pp. 45–52, 2018.
- [25] N. U. Rehman and H. Aftab, "Multivariate variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 67, no. 23, pp. 6039–6052, 2019.
- [26] N. Salankar, N. Ahuja, S. Chaurasia, and N. Singh, "Fuzzy logic based approach for automation of emotion detection in misophonia," *International Journal of Computer Science and Information Technology*, vol. 9, no. 1, pp. 37–55, 2017.
- [27] R. Rifai Chai, S. H. Sai Ho Ling, G. P. Hunter, and H. T. Nguyen, "Mental task classifications using prefrontal cortex electroencephalograph signals," in *Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, USA, August 2012.
- [28] J. Zhang, X. Yu, and D. Xie, "Effects of mental tasks on the cardiorespiratory synchronization," *Respiratory Physiology & Neurobiology*, vol. 170, no. 1, pp. 91–95, 2010.
- [29] T. Fernández, T. Harmony, M. Rodríguez et al., "EEG activation patterns during the performance of tasks involving different components of mental calculation," *Electroencephalography and Clinical Neurophysiology*, vol. 94, 1995.
- [30] A. Glass, "Mental arithmetic and blocking of the occipital alpha rhythm," *Electroencephalography and Clinical Neurophysiology*, vol. 16, 1964.
- [31] B. Fatimah, A. Javali, H. Ansar, B. G. Harshitha, and H. Kumar, "Mental arithmetic task classification using fourier decomposition method," in *Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSPP)*, Chennai, India, July 2020.

- [32] R. T. Toll, W. Wu, S. Naparstek et al., "An electroencephalography connectomic profile of posttraumatic stress disorder," *American Journal of Psychiatry*, vol. 177, no. 3, pp. 233–243, March 2020.
- [33] Y. Zhang, W. Wu, R. T. Toll et al., "Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography," *Nature Biomedical Engineering*, vol. 5, no. 4, pp. 309–323, 2021.
- [34] W. Wu, Y. Zhang, J. Jiang et al., "An electroencephalographic signature predicts antidepressant response in major depression," *Nature Biotechnology*, vol. 38, no. 4, pp. 439–447, 2020.
- [35] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [36] R. A. Thuraisingham, "A classification system to detect congestive heart failure using second-order difference plot of RR intervals," *Cardiology Research and Practice*, vol. 2009, Article ID 807379, 7 pages, 2009.
- [37] J. Kwon, J. Shin, and C.-H. Im, "Toward a compact hybrid brain-computer interface (BCI): performance evaluation of multi-class hybrid EEG-fNIRS BCIs with limited number of channels," *PLoS One*, vol. 15, no. 3, Article ID e0230491, 2020.
- [38] D. Gurve, D. Delisle-Rodriguez, M. Romero-Laiseca et al., "Subject-specific EEG channel selection using non-negative matrix factorization for lower-limb motor imagery recognition," *Journal of Neural Engineering*, vol. 17, no. 2, Article ID 026029, 2020.
- [39] Y. Pan, I. W. Tsang, A. K. Singh, C.-T. Lin, and M. Sugiyama, "Stochastic multichannel ranking with brain dynamics preferences," *Neural Computation*, vol. 32, no. 8, pp. 1499–1530, 2020.
- [40] F. Al-Shargie, M. Kiguchi, N. Badruddin, S. C. Dass, A. F. M. Hani, and T. B. Tang, "Mental stress assessment using simultaneous measurement of EEG and fNIRS," *Biomedical Optics Express*, vol. 7, no. 10, p. 3882, 2016.
- [41] F. Al-shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Towards multilevel mental stress assessment using SVM with ECOC: an EEG approach," *Medical, & Biological Engineering & Computing*, vol. 56, no. 1, pp. 125–136, 2018.
- [42] P. Murugesu Pandiyan and S. Yaacob, "The assessment of developed mental stress elicitation protocol based on heart rate and EEG signals," *International Journal of Computer Theory and Engineering*, vol. 7, 2015.
- [43] T. H. Priya, P. Mahalakshmi, V. Naidu, and M. Srinivas, "Stress detection from EEG using power ratio," in *Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, India, February 2020.
- [44] R. Bernays, J. Mone, P. Yau et al., "Lost in the dark," in *Proceedings of the Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology—UIST Adjunct Proceedings'12*, Cambridge, MA, USA, October 2012.
- [45] A. Morozov and W. Ito, "Plasticity of amygdala neurons underlying fear learning and extinction," *Handbook of Behavioral Neurobiology*, vol. 26, pp. 127–139, 2020.
- [46] Y. Y. Huang and E. R. Kandel, "Postsynaptic induction and PKA-dependent expression of LTP in the lateral amygdala," *Neuron*, vol. 21, 1998.
- [47] M. Delazer, A. Gasperi, L. Bartha, E. Trinka, and T. Benke, "Number processing in temporal lobe epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 6, pp. 901–903, 2004.
- [48] S. M. Qaisar and A. Subasi, "Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, pp. 1–13, 2020.
- [49] S. Mina Qaisar and A. Subasi, "Cloud-based ECG monitoring using event-driven ECG acquisition and machine learning techniques," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 623–634, 2020.
- [50] S. Mina Qaisar, D. Sidiya, M. Akbar, and A. Subasi, "An event-driven multiple objects surveillance system," *International Journal of Electrical and Computer Engineering Systems*, vol. 9, no. 1, pp. 35–44, 2018.
- [51] S. M. Qaisar, "Signal-piloted processing and machine learning based efficient power quality disturbances recognition," *PLoS One*, vol. 16, no. 5, Article ID e0252104, 2021.

Research Article

Assessment of Acoustic Features and Machine Learning for Parkinson's Detection

Moumita Pramanik ¹, Ratika Pradhan ¹, Parvati Nandy ², Saeed Mian Qaisar ^{3,4}
and Akash Kumar Bhoi ⁵

¹Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India

²Department of Medicine, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Tadong 737102, Sikkim, India

³Department of Electrical and Computer Engineering, Effat University, Jeddah 22332, Saudi Arabia

⁴Communication and Signal Processing Lab, Energy and Technology Research Center, Effat University, Jeddah 22332, Saudi Arabia

⁵Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India

Correspondence should be addressed to Saeed Mian Qaisar; sqaisar@effatuniversity.edu.sa

Received 15 June 2021; Revised 22 July 2021; Accepted 13 August 2021; Published 23 August 2021

Academic Editor: Siti Anom Ahmad

Copyright © 2021 Moumita Pramanik 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.

This article presents a machine learning approach for Parkinson's disease detection. Potential multiple acoustic signal features of Parkinson's and control subjects are ascertained. A collaborated feature bank is created through correlated feature selection, Fisher score feature selection, and mutual information-based feature selection schemes. A detection model on top of the feature bank has been developed using the traditional Naïve Bayes, which proved state of the art. The Naïve Bayes detector on collaborative acoustic features can detect the presence of Parkinson's magnificently with a detection accuracy of 78.97% and precision of 0.926, under the hold-out cross validation. The collaborative feature bank on Naïve Bayes revealed distinguishable results as compared to many other recently proposed approaches. The simplicity of Naïve Bayes makes the system robust and effective throughout the detection process.

1. Introduction

Parkinson's disease (PD) is an inherent disease among elderly individuals. The disease appears when the dopamine neurons significantly fall in the human brain [1, 2]. The PD symptoms start with voice impairments at its early stage, tremor, and loss of memory, and the subject shows an inability to walk, run, and even perform regular day-to-day duties. The situation worsens at a late age, where the subject suffers huge memory loss and cannot move and lean to perform minor activities. The worst part is that the disease is not curable and not reversible [3], so all efforts have been made to its early detection and preventive measures to

suppress its adverse effects. Medical science reveals that Parkinson's disease mainly causes gradual reduction of dopamine hormone in the human brain as this hormone acts as the transmitter of signals among various neurons [4]. Insufficient amount of dopamine hormone leads to non-transmission of signals and various neurorelated disorders and symptoms being started in human beings, and Parkinson's disease is one of them. Symptoms of PD can be nonmotor and motor-related. Nonmotor symptoms include sleep disorder, speech variation, problem in swallowing, and loss of smell, whereas motor symptoms were connected to slow movement, e.g., bradykinesia, tremor, rigidity, and postural instability [5]. These symptoms also vary from

patient to patient over different time periods, and the appearance of symptoms is often lately observed by the patient due to the casual ignorance of early symptoms.

The effect of Parkinson's disease varies from person to person, and all the symptoms may not be evident by every PD patient and even may not appear in the same order and same combination. However, subjects suffering from idiopathic rapid eye movement sleep behavior disorder (iRBD) are more prone to PD. Speech changes are the first motor symptom that appears even ten years before the actual diagnosis starts [6]. Therefore, assessing speech signals provides a better scope for detecting chances of Parkinson's in the early stage. For instance, the time domain amplitude of both controls and Parkinson's has been visualized in Figure 1. Each block of Figure 1 represents a subject, where the green color plots represent controls and the red color plots represent the subjects suffering Parkinson's. The subjects' specific plots are generated on the pronunciation of sustaining vowel/a/in Italian language [7].

From Figure 1, the amplitude of Parkinson's subjects appears to be abnormal, where the disorder can be identified. On the other hand, the amplitude of the Non-Parkinson's Disease subject is uniformly in a decreasing trend. The disorder signal of Parkinson's subjects is the dysphonia and hypokinetic dysarthria that a subject suffers at various stages of PD [8]. Dysphonia refers to the inability to produce normal phonation due to impaired functioning of the phonatory system. Reduction of pitch variation often denotes monotonicity and reduces loudness, breathless voice, and tiny speech formation [9]. Approximately 90% of the PD patients are affected by this combined sign of hypokinetic dysarthria [9]. In the context of acoustic voice analysis, it is difficult to identify the slight variation of a sound wave through the naked ears. In such a situation, the power of machine learning techniques can be employed to discriminate Parkinson's from the other signal [10, 11].

As PD is a nonreversible disease, the only option left with the clinical practitioners is to reduce the speed of the effect. In this way, the subject feels confident and cured if the diagnosis process starts early. PD shows only a few symptoms at the early stage on the flip side of the coin, like voice disorder and mild tremors. However, these symptoms also resemble other symptoms of an average person. This is why diagnostic technicians and clinical practitioners are nowadays exploring machine learning and artificial intelligence approaches [12–14] to predict the presence and severity of disease among their subjects.

The main contribution of this article is as follows:

- (a) A collaborative feature bank consisting of seven vocal features has been created from Baseline Features (BF), Vocal Fold Features (VFF), and Time Frequency (TFF) with the help of Correlated Feature Selection (CFS) [15], Fisher Score Feature Selection (FSFS) [16], and Mutual Information-Based Feature Selection (MIFS) [17].
- (b) The traditional Naïve Bayes has been trained and tested on the seven features of the collaborative feature bank, which shows the robustness and

effectiveness of our system as compared to other recent approaches of Parkinson's disease detection.

The rest of the article is as follows. Section 2 deals with literature reviews, Section 3 outlines the materials and methods, and Section 4 briefly discusses the results, followed by a conclusion at Section 5.

2. Literature Review

Many recent machine learning techniques, including Naïve Bayes, proved useful in segregating subjects suffering PD from the controls. For instance, Avuçlu and Elen [18] proposed Parkinson's detection through multiple classifiers. Their experiment was conducted on various training and testing instances spanned over 22 vocal features of 195 sound samples. The k-Nearest Neighbor, Random Forest, Support Vector Machine, along with Naïve Bayes, have been used to detect Parkinson's. It has been observed that the Naïve Bayes detects the Parkinson's subjects with 70.26% accuracy with a precision of 0.64. Bourouhou et al. [19] compared many classifiers to predict the presence of Parkinson's among subjects. Their experiment was conducted on 40 subjects comprising 20 Parkinson's and control subjects. The experimental results Naïve Bayes detector revealed a detection accuracy of 65%, the sensitivity of 63.6%, and specificity of 66.6%, respectively. On a similar note, Zhang et al. [20] used Naïve Bayes along with other machine learning techniques to detect Parkinson's disease. Their approach employed signal processing techniques to extract relevant features from the acoustic signal of Parkinson's and control subjects. At the next stage Naïve Bayes, Support Vector Machine (SVM), Logistic Regression (LR), and single and double-layered neural networks have been used to segregate Parkinson's and control subjects. With the 22 vocal features, the Naïve Bayes reveals 69.24% of detection accuracy with a 96.02% of the precision rate. Meghraoui et al. [21] proposed Bernoulli and Multinomial Naïve Bayes (BMNB) on harmonicity, pitch, and pulse features. The BMNB approaches are proved to be a better solution to detect the presence of Parkinson's. A test on 28 samples comes across with a 62.5% detection accuracy on Bernoulli Naïve Bayes (BNB) with 0.375 Mean Squared Error (MSE). Kadiri et al. [22] proposed a method of Parkinson's disease detection using SVM on Single Frequency Filtering Cepstral Coefficients (SFFCC) and Shifted Delta Cepstral (SDC) features exacted from voice signals of Parkinson's and control subjects. The SFFCC + SDC features witnessed 9% of performance improvements as compared to traditional MFCC + SDC features. The traditional SVM on SFFCC + SDC features shows 73.33% detection accuracy with 73.32% F1-score.

Apart from Naïve Bayes, many other supervised techniques, including but not limited to famous deep learning techniques, have been proposed to detect Parkinson's among subjects. Recently Jain et al. [23] proposed a Parkinson's detection method using multiple classifier ensembles. The authors used Synthetic Minority Oversampling Technique (SMOTE) to generate artificial samples for prediction. Their

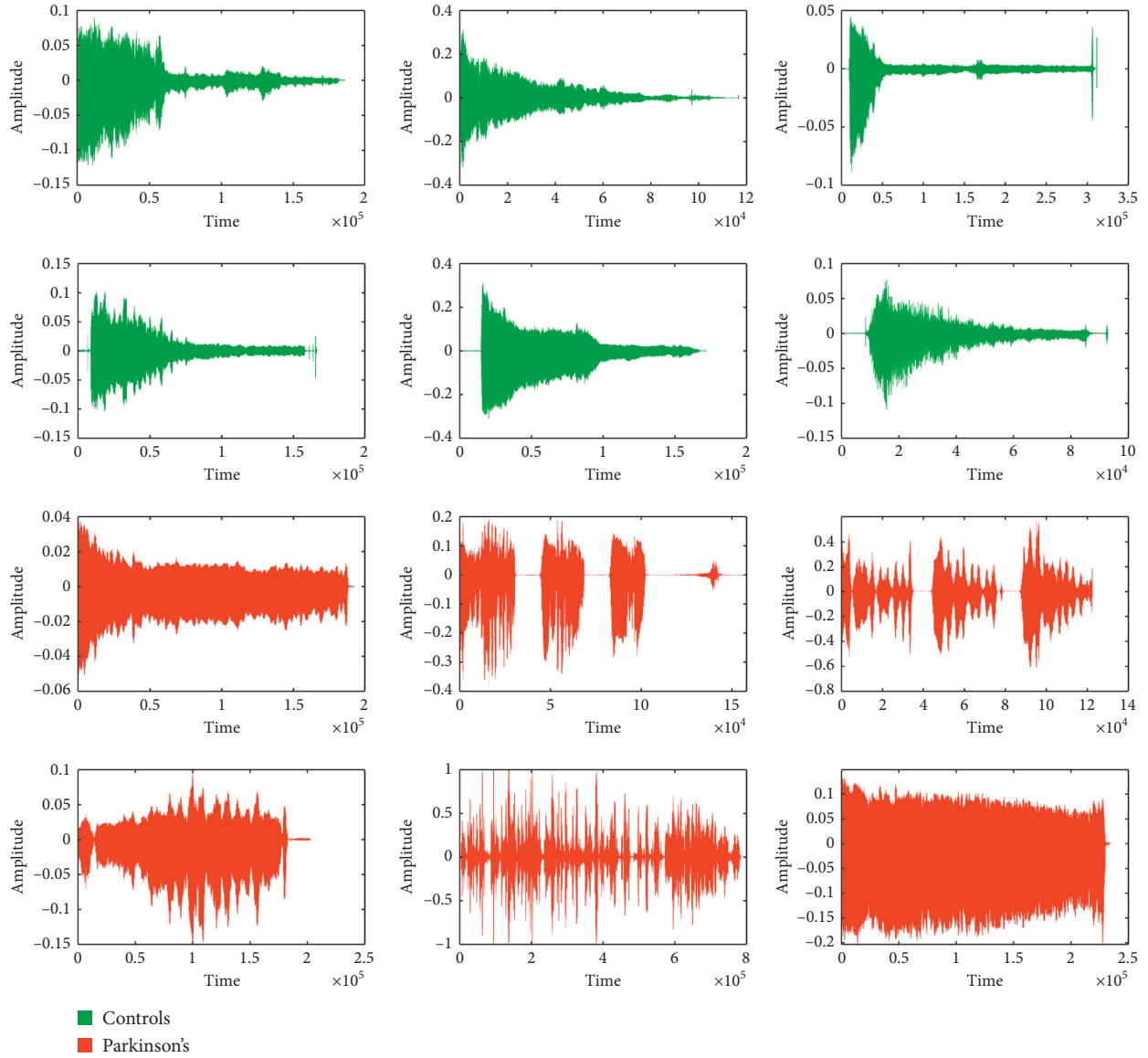


FIGURE 1: Amplitudes of controls and Parkinson's subjects.

proposed approach on Deep Neural Network (DNN) detects Parkinson's with a detection accuracy of 91.47%. Though the result seems impressive, their approach does not appear practical for many reasons. The authors used the dataset proposed by Sakar et al. [24], and the dataset contains replicated speech information of 252 subjects resulting in 756 instances. Machine learning methods cannot be directly applied to these instances as each subject has three readings of the speech signal. These instances need to be consolidated before the actual classification starts. Moreover, creating a Parkinson's detection system on 754 features is not convincing. The Performance of DNN, as claimed by the authors, may vary on consolidated instances. Further, their system may not be practically effective on synthetic samples generated by SMOTE. Similarly, Polat and Nour [25] use multiple classifiers ensemble to detect Parkinson. The One Against All (OAA) sampling technique plays a pivotal role in the detection process. The Logistic Regression (LR) on OAA

samples proved to be a brilliant Parkinson's detector. Multiple supervised classifiers are also used on vocal features selected through Adaptive Grey Wolf Optimization Algorithm (AGWOA) and Sparse Auto Encoder (SAE) [26]. The Naïve Bayes classifier on AGWOA and SAE features reveals a detection accuracy of 72%. In the recent past, decision trees are gaining popularity in biomedical data classification [27]. Classification and Regression Tree (CART) have been used to detect the presence of Parkinson's [28], where the CART detector detects Parkinson's with 75.19% through 8 optimum features of vowel /a/.

3. Materials and Methods

3.1. Dataset. The idea behind the proposed approach is the feature collaboration to detect Parkinson's disease. For feature collaboration, the Baseline Features (BF), Vocal Fold Features (VFF), and the Time Frequency Features (TFF)

about the acoustic signal of both Parkinson's and control patients have been considered. All the BF, VFF, and TFF are extracted from a recent Parkinson's detection database publicly available at the UCI machine learning repository [24], prepared at the Department of Neurology in Cerrahpaşa, Faculty of Medicine, Istanbul. The database contains 752 acoustic features of 252 subjects, including control and Parkinson's. Data is prepared with a 44.1 kHz microphone setting followed by a physician's examination. The sustained phonation of the vowel /a/ was collected from each subject with three repetitions.

The vast 752 features also include 22 VFF, 11 TFF, and 21 BF. These features are extracted using Praat acoustic analysis software [24]. The number of features available under VFF, TFF, and BF of the dataset has been presented in Table 1. Gender-specific control and sick subjects are outlined in Table 2. The detailed characteristics of these features segments and corresponding features can be found at [24, 27].

The Istanbul acoustic database [24] used here comprises 252 subjects, where 64 are controls, and 188 subjects are suffering from Parkinson's. Similarly, the dataset contains vocal information of 122 female (41 controls and 81 Parkinson's) and 130 male subjects (41 controls and 81 Parkinson's).

3.2. Features Selection. For effective collaboration, a Features Bank (FB) is created using the best features of BF, VFF, and TFF. The identification of best features has been established through three prominent feature selection techniques [29, 30]— Correlated Feature Selection (CFS) [15], Fisher Score Feature Selection (FSFS) [16], and Mutual Information-based Feature Selection (MIFS) [17]. These feature selection schemes initially ranked the features (based on their contribution towards the classification). They selected the most suitable features from the ranked features (features having the highest contribution towards the classification process). All three CFS, FSFS, and MIFS techniques use distinct proven mechanisms for feature ranking. The CFS calculates correlation among attributes to understand the variable similarity. For two attributes $A = \{a_1, a_2, a_3, \dots, a_n\}$ and $B = \{b_1, b_2, b_3, \dots, b_n\}$, CFS calculates correlation r as follows:

$$r = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}}, \quad (1)$$

where \bar{a} = mean of attribute A and \bar{b} = mean of attribute B . The higher the value of r , the more the underlying attributes correlated and the lower the value of r the underlying attributes have far deviated from each other. After calculating the correlation score for each attribute, the attributes are arranged in the ascending order of the correlation score. Arranging attributes based on correlation score provides a scope to move the highly uncorrelated attributes to the front and perfectly correlated attributes at the rear, thus supporting the classifiers for enhanced detection. Similarly, FSFS calculates the fisher score of individual features of the underlying Parkinson's dataset. The feature weights are

TABLE 1: Features of Istanbul acoustic dataset.

Features group	Number of features
<i>Vocal fold features</i>	
Glottis quotient	3
Glottal to noise excitation	6
Empirical mode decomposition	6
Vocal fold excitation ratio	7
<i>Time frequency features</i>	
Voice intensity	3
Bandwidth	4
Formant frequencies	4
<i>Baseline features</i>	
Entropy of recurrence period density	1
Detrended fluctuation	1
Entropy of pitch period	1
Harmonicity	2
Variants of jitter	5
Fundamental frequency	5
Variants of shimmer	6

TABLE 2: Gender-specific controls and subjects suffering from Parkinson's in the dataset.

Genders (↓)/Classes (→)	Controls	Parkinson's	Total
Female (—)	41	81	122
Male (—)	23	107	130
Total	64	188	252

calculated based on the sample size and number of class labels. FSFS are tested for binary and multiclass datasets, but it is widely used for binary datasets [31]; hence, a suitable feature ranker is proposed for the current work. For a given set of features $f = \{f_1, f_2, f_3, \dots, f_p\}$ having a set of classes $K = \{k_1, k_2, k_3, \dots, k_c\}$, the fisher score S of the feature f_i can be estimated as follows:

$$S = \frac{\sum_{j=1}^c n_j (\mu_{ij} - \mu_i)^2}{\sum_{j=1}^c n_j \rho_{ij}^2}, \quad (2)$$

where n_j is the number of instances in the j^{th} class, μ_i is the mean of the i^{th} feature, and μ_{ij} and ρ_{ij} are the mean and variance of the i^{th} feature and j^{th} class, respectively. In this way, the fisher score of each feature of the Parkinson's dataset has been calculated, allowing us to rank the features based on the score accumulated. It should be noted that the fisher score evaluates the score individually; i.e., no two features are taken simultaneously to calculate the feature's score [32]. The individual fisher score proved to be a limitation to identify the feature redundancy. However, since prominent features have been selected iteratively through Naïve Bayes classification, the limitation of identifying feature redundancy will not affect the evaluation process. With a similar guideline of CFS, the MIFS ranking algorithm estimated the relationship among features through mutual information and ranked the features based on the mutual information score of attributes. For any two given attributes a and b having values $\{1, \dots, p\}$ and $\{1, \dots, q\}$, respectively, a joint probability π_{ab} ensures the samples of attribute

$(a, b) \in \{1, \dots, p\} \times \{1, \dots, q\}$, then the dependency between a and b can be estimated [17] through mutual information as follows:

$$MI = \sum_{a=1}^p \sum_{b=1}^q \pi_{ab} \log \frac{\pi_{ab}}{\sum_b \pi_{ab} \sum_a \pi_{ab}}. \quad (3)$$

Like correlation score, mutual information places a crucial role in features ranking. All the three feature ranking algorithms CFS, FSFS, and MIFS can also be extended to select a subset of features. After ranking all ranked feature segments, the ranked features are passed to Naïve Bayes incrementally *one* feature at a time in an iterative fashion. The incremental feature classification allows selecting the suitable number of features from each segment where the Naïve Bayes shows the highest detection accuracy.

In a nutshell, all the three feature selection techniques CFS, FSFS, and MIFS work jointly to identify goodness scores for each attribute of the underlying Parkinson's dataset. The idea behind this incremental feature selection is to select only those attributes which are mainly close to class attributes and not close to each other. However, instead of depending on the practical way of identifying attributes, selecting attributes through incremental classification is emphasized. In a landscape, the incremental feature selection helps to identify potential attributes in the most realistic way. The selected features of BF, VFF, and TFF through CFS, FSFS, and MIFS provide the most relevant collaborative Parkinson's disease detection features. The entire process of Parkinson's detection process has been depicted in Figure 2.

The process of detecting subjects affected with Parkinson's follows three steps; viz., Feature Selection, Feature Collaboration, and Parkinson's Detection. As pointed earlier, in the feature selection stage, the BF, TFF, and VFF are ranked separately using CFS, FSFS, and MIFS techniques. As a result, nine feature blocks are realized. The feature collaboration stage's ranked feature blocks are passed, where Naïve Bayes play a crucial role in suitable feature identification. Features from each ranked feature block are fetched incrementally and sent to Naïve Bayes for classification. This process continues till all features are fetched from each ranked feature block. The incremental features for classification help identify the minimum number of features required to achieve maximum detection accuracy. The number of ranked features for which the maximum amount of detection accuracy has been received are identified. For each feature block, i.e., VFF, TFF, and BF, the best features are identified by comparing all three feature ranking schemes (i.e., CFS, FSFS, and MIFS).

3.3. Classification. The ranked features are collaborated and sent to Naïve Bayes for detection of Parkinson's. In this way, the entire detection process relies on a small number of collaborative features; thus, it appears to be a practical method of Parkinson's detection. The detection approach has been developed using the Weka machine learning repository [33, 34]. The implementation settings of the proposed model are outlined in Table 3.

The predictive model of Naïve Bayes uses estimator classes for prediction [35]. The numeric estimator precision values are chosen based on the analysis of the training data. The batch size indicates the desired number of instances to process for batch prediction of testing samples. The supervised discretization option ensures the conversion of numerical attributes to nominal ones. All the attributes remain numerical, so this option has been disabled during the training and testing process.

4. Results and Discussion

The results of the proposed work have been analyzed in three broad ways. At the first stage, the efficiency of feature ranking schemes, i.e., CFS, FSFS, and MIFS, has been analyzed. The individual ranking of features per feature selector helps identify the most potential VFF, TFF, and BF segments for effective collaboration. At the second stage, the performance of Naïve Bayes has been evaluated along with many other traditional supervised classifiers in the context of Parkinson's detection. Finally, the proposed collaborative feature-based Parkinson's detection system has been compared against other recent vibrant Parkinson's detection mechanisms.

4.1. Collaborative Features Identification. As the first stage of the collaborative Parkinson's detection scheme, a bank of collaborative features is prepared. The detection accuracy of Naïve Bayes on change in the vocal fold, time frequency, and baseline feature through CFS, FSFS, and MIFS ranking has been presented in Figures 3–5, respectively. The classification accuracy of Naïve Bayes was also recorded on original features to understand the power of feature ranking techniques.

It is to note that both the original and the ranked acoustic features are incrementally processed through Naïve Bayes to observe the performance enhancement with a change in the number of features. The performance of Naïve Bayes due to CFS, FSFS, and MIFS shows a satisfactory result as compared to original features. It can be seen from Figure 3 that the CFS shows the highest detection accuracy with just ten features in hand. In contrast, the same Naïve Bayes took 12 original features to produce similar detection accuracy. On the other hand, the three features of the FSFS ranked scheme help the Naïve Bayes attain the same CFS detection accuracy. On a similar note, the Naïve Bayes shows the same detection accuracy with 6 MIFS features. Therefore, all the three CFS, FSFS, and MIFS boost the performance of Naïve Bayes to the peak with the help of 10, 3, and 6 features, respectively. Therefore, the 3 FSFS features have been sent to the feature bank for collaboration.

With a similar guideline, when both the original TFF features and ranked CFS, FSFS, and MIFS features are processed incrementally, only the 3 features of CFS boost the performance of Naïve Bayes exceptionally well up to 75.79%. However, FSFS also boosts the Naïve Bayes' performance but not as that of CFS and MIFS. Both FSFS and MIFS reveal a satisfactory performance improvement with a detection

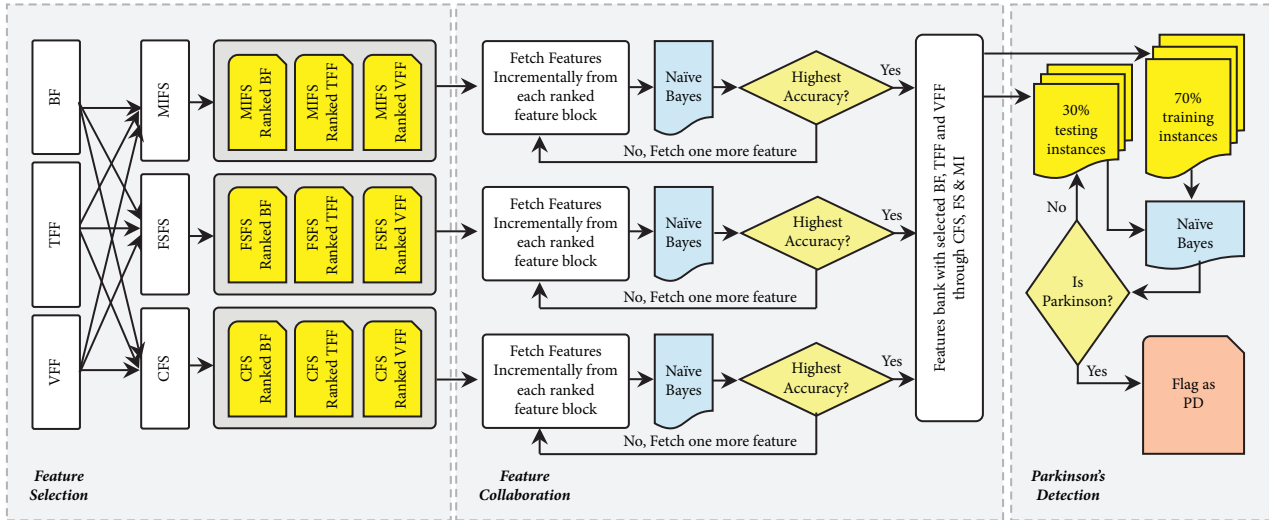


FIGURE 2: The process of collaborative Parkinson's detection.

TABLE 3: Settings used for Naïve Bayes.

Settings	Value
Batch size	100
Use kernel estimator	True
Use supervised discretization	False

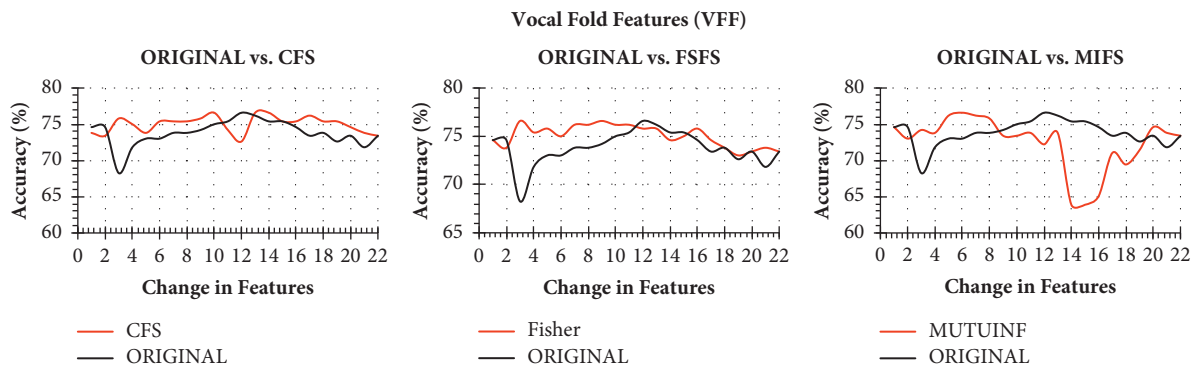


FIGURE 3: Classification accuracy of Naïve Bayes with change in original features and ranked features on VFF.

accuracy of 73.4% and 73.81%, respectively. Though the Naïve Bayes took only 1 MIFS feature, the first 3 features of CFS have been sent to the feature bank for collaboration due to the highest detection accuracy.

When the performance of Naïve Bayes is studied, the performance of the classifier due to rankers CFS, FSFS, and MIFS was found to be degraded. Nevertheless, the rankers show a similar result as that of original arrangements with minimal features. In this regard, the Naïve Bayes yields the highest accuracy of 76.59% with 3FSFS features. But instead of FSFS, we prefer to choose 1 CFS ranked baseline feature. The CFS enhances the performance of Naïve Bayes with the same detection accuracy parallel to the original order of features with a lesser number of features. Therefore, the first feature of baseline ranked through CFS ranker has been shortlisted and sent to feature bank for collaboration.

The performance of Naïve Bayes on CFS, FSFS, and MIFS and the original order of VFF, TFF, and BF features have been presented in Table 4. The feature threshold column indicates the minimum number of features identified to produce maximum detection accuracy under the concern settings. So, a total of 3 FRFS ranked vocal fold features. 3 CFS ranked time frequency features and 1 CFS ranked baseline features are identified for feature collaboration.

4.2. Performance Analysis of Collaborative Parkinson's Detection. As the first stage of collaborative Parkinson's detection scheme, a bank of 7 collaborative features comprising VFF, TFF, BF has been prepared. Those 7 features have been undergone 10-fold cross validation on Naïve Bayes classifier. The result obtained both for Parkinson's and control subjects has been presented in Table 5.

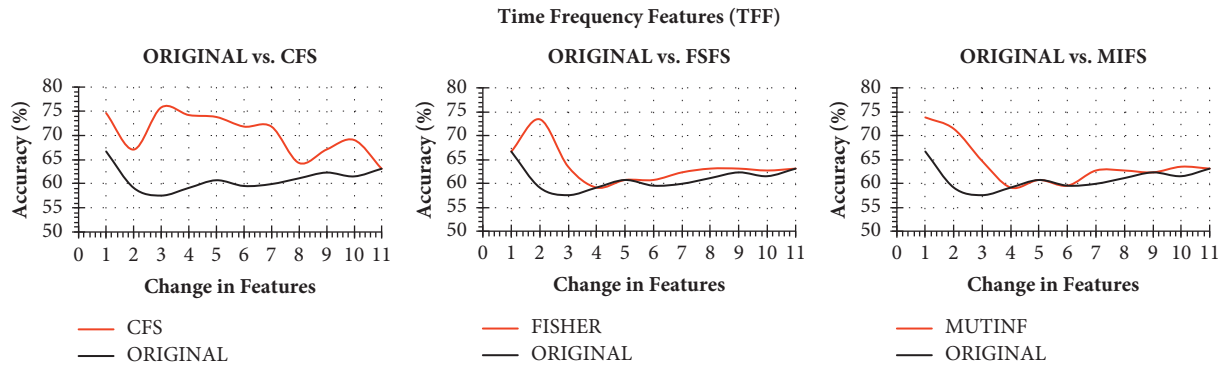


FIGURE 4: Classification accuracy of Naïve Bayes with change in original features and ranked features on TFF.

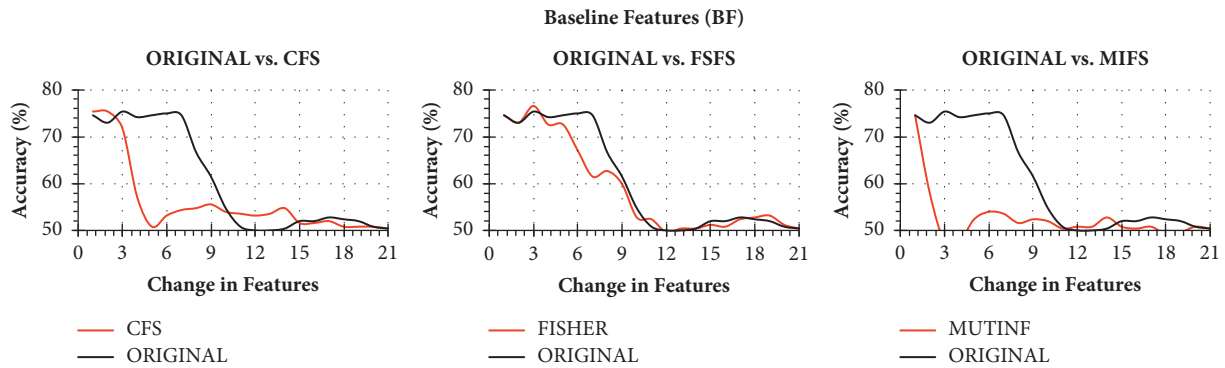


FIGURE 5: Classification accuracy of Naïve Bayes with change in original features and ranked features on BF.

TABLE 4: Highest detection accuracy of Naïve Bayes due to ranked acoustic features and original features.

Feature selection techniques	Vocal fold		Time frequency		Baseline	
	Feature threshold	Accuracy	Feature threshold	Accuracy	Feature threshold	Accuracy
CFS	10	76.59	3	75.79	1	75.40
FSFS	3	76.59	2	73.41	3	76.59
MIFS	6	76.59	1	73.81	1	74.60
Original	12	76.59	1	66.67	3	75.40

TABLE 5: Performance of collaborative Parkinson’s detection on Naïve Bayes.

Subjects	Sensitivity	Specificity	Precision	F-measure	MCC	ROC area	PRC area
Control	0.391	0.926	0.641	0.485	0.380	0.762	0.514
Parkinson’s	0.926	0.391	0.817	0.868	0.380	0.762	0.905

According to Table 5, the sensitivity of Parkinson’s subjects and specificity of control subjects are satisfactory. The specificity of 0.926 for control subjects indicates that the collaborative Parkinson’s detection model correctly detects negative results for 92.6% of control subjects who have undergone the test. Similarly, the sensitivity of 0.926 for Parkinson’s subjects pointed out that the model will correctly return a positive result for 92.6% of the disease subjects. Similarly, a precision of 0.817 indicates a total of 174 subjects are suffering from Parkinson’s out of all the subjects that are predicted as Parkinson’s, which is impressive in the context of medical diagnosis. On the other

hand, the Receiver Operating Curve (ROC) represents an excellent AUC (>71%). The Precision-Recall Curve (PRC) represents 0.905, which is again in an acceptable range. The ROC and the PRC of subjects predicted as control or Parkinson’s have been presented in Figure 6.

According to Figure 6(a), the ROC of both the Control and Parkinson’s subjects is entirely satisfactory. The curves are tending nicely towards the true positive rate. The curves claim 76.2% area of the plot both for Controls and Parkinson’s subjects. On the other hand, the PRC is convincing for Parkinson’s subjects, whereas for the control subjects, the PRC is not convincing (Figure 6(b)).

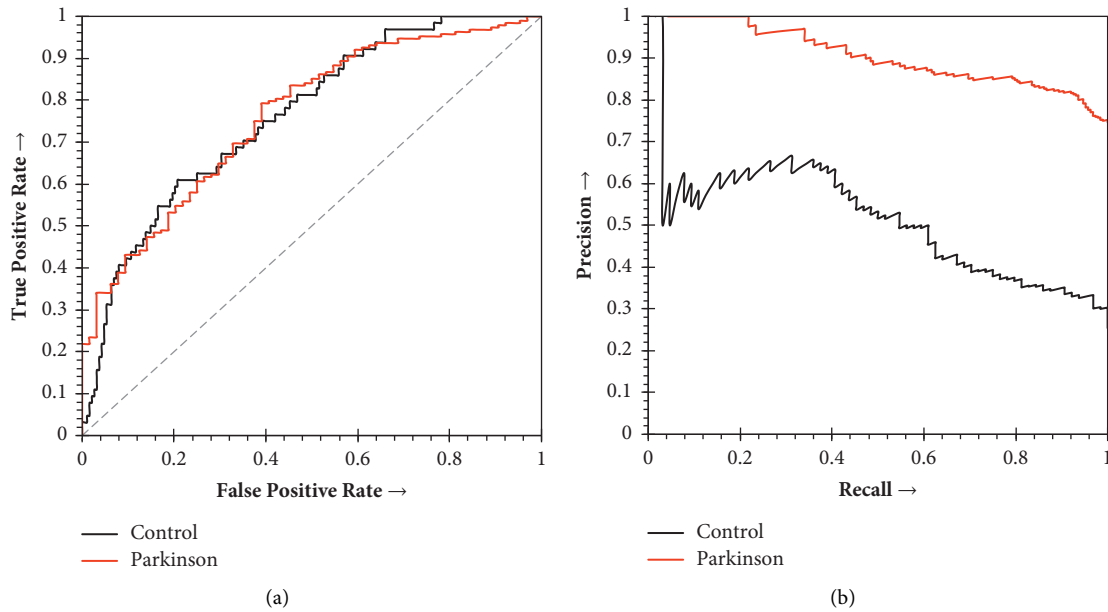


FIGURE 6: (a) Receiver Operating Curve of collaborative Parkinson's detection for Parkinson's and Control subjects. (b) Precision-Recall Curve of collaborative Parkinson's detection for Parkinson's and Control subjects.

4.3. Performance Comparison with Other State-of-the-Art Models. This section highlights the comparison of the proposed work with other similar classifiers for Parkinson's disease detection. The seven collaborative features used are also passed to the C4.5 decision tree, k-Nearest Neighbor, Logistic Regression, Neural Network, and Random Forest classifiers. The hold-out validation method has been employed to validate the proposed model with other state-of-the-art approaches. In the view of hold-out validation, the training instances are prepared with 30% of the subjects, and the testing instances are 70% of subjects randomly. It is observed that Naïve Bayes on collaborative features excels with 78.97% of detection accuracy with the lowest ever training time. The k-Nearest Neighbor suffers on the collaborative features with the lowest detection accuracy of 67.46%. However, the training time of k-Nearest Neighbor is at par with that of Naïve Bayes. On the other hand, Logistic Regression shows a close performance outcome of Naïve Bayes with a bit of training time of 0.03 s. The detailed performance outcomes of the proposed approach, along with others, are presented in Table 6.

In a subsequent attempt, errors generated by the proposed collaborative Parkinson's detection system have been observed along with peer supervised classifiers. The errors generated by the various classifiers along with collaborative features based on Naïve Bayes represent an inconclusive result. It is because the collaborative PDS shows better results for Mean Absolute Error (MAE). In contrast, it shows at par results with other classifiers in Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE). The outcome of error matrices such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE) have been presented in Table 7.

Similarly, the Naïve Bayes based on collaborative features is also compared with other classifiers through ROC and PRC. The results about the various classifiers have been outlined in Table 8.

In Table 8, Naïve Bayes represents exceptional ROC and PRC Values of 76% and 81%. The results appear to be far better than that of the k-Nearest Neighbor and C4.5 decision tree. The Logistic Regression is the only classifier that closely competes with Naïve Bayes. The ROC and PRC are visually represented for all classifiers, including Naïve Bayes in Figure 7 for control and Parkinson's subjects.

ROC of all the classifiers, including Naïve Bayes, can be seen more towards True Positive Rates. However, C4.5 and k-Nearest Neighbor suffers for controls but shows marginal results for Parkinson's subjects. In addition, with the progression of false positives, k-Nearest Neighbor reveals low true positive rates, and thus, results in low AUC. On the other hand, while evaluating PRC, it is found that Naïve Bayes outperforms with superior precision. Therefore, the proposed collaborative features on Naïve Bayes is a practical approach to Parkinson's detection. At the final stage of analysis, the proposed collaborative features-based Parkinson's detection system has been compared with the current state-of-the-art function-based methods, viz., Avuçlu and Elen [18], Bourouhou et al. [19], Zhang et al. [20], Meghraoui et al. [21], Kadirı et al. [22], Polat and Nour [25], Xiong and Lu [26] and Mekyska et al. [28]. Since our approach is based on a function-based approach, most of the methods taken for comparison belong to function-based approaches such as Naïve Bayes and Support Vector Machine (SVM). The comparison has been conducted in two different sets of performance matrices. At first, the standard detection accuracy has been used for the comparison (Table 9). Finally, the Naïve Bayes based Parkinson's detection mechanisms

TABLE 6: Detection accuracy and misclassification rate of collaborative Parkinson’s Detection using Naïve Bayes and other supervised classifiers.

Classifiers	Number of features	Training time (Sec)	Accuracy (%)	Misclassification rate (%)
C4.5 decision tree	7	0.03	73.81	26.19
k-Nearest Neighbor	7	0.01	67.46	32.54
Logistic Regression	7	0.03	77.38	22.62
Neural Network	7	0.14	75.40	24.60
Random Forest	7	0.21	76.98	23.02
Naïve Bayes	7	0.01	78.97	21.03

TABLE 7: Error matrices of collaborative Parkinson’s detection using Naive Bayes.

Classifier	Attributes	MAE	RMSE	RAE	RRSE
C4.5 decision tree	7	0.33	0.46	86.12	104.99
k-Nearest Neighbor	7	0.33	0.57	86.02	130.45
Logistic Regression	7	0.31	0.40	80.33	91.63
Neural Network	7	0.31	0.42	80.42	96.26
Random Forest	7	0.31	0.40	81.26	92.33
Naïve Bayes	7	0.26	0.41	68.47	95.24

TABLE 8: ROC area and PRC area of collaborative Parkinson’s detection using Naive Bayes.

Classifier	Attributes	ROC area	PRC area
C4.5 decision tree	7	0.60	0.69
k-Nearest Neighbor	7	0.56	0.65
Logistic Regression	7	0.75	0.80
Neural Network	7	0.73	0.79
Random Forest	7	0.74	0.78
Naïve Bayes	7	0.76	0.81

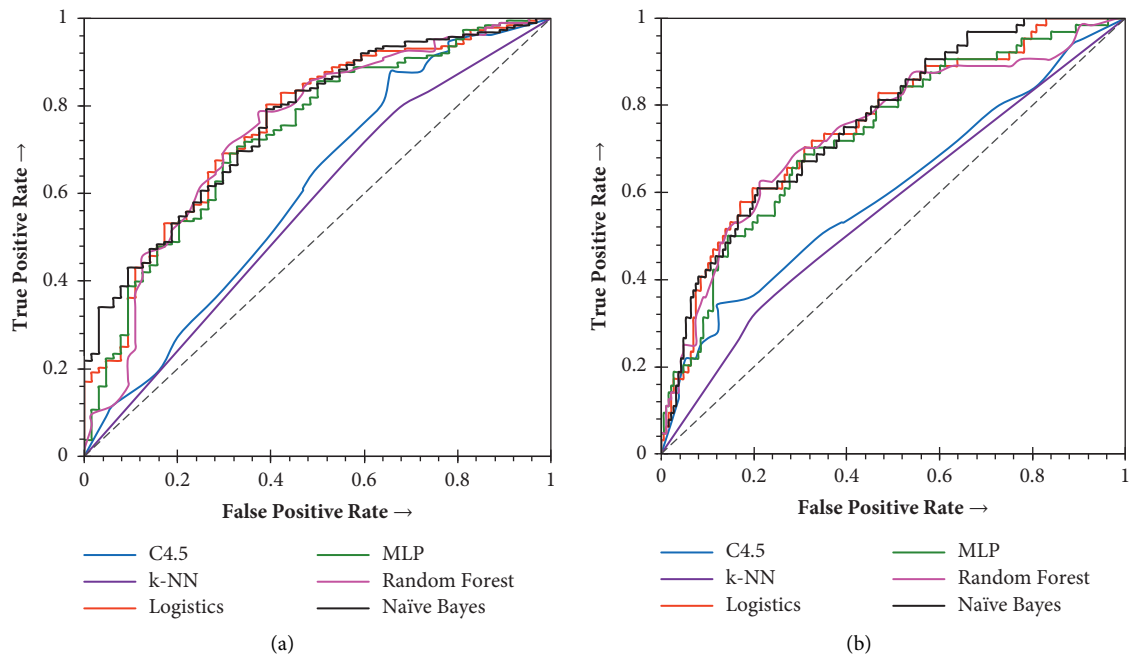


FIGURE 7: Continued.

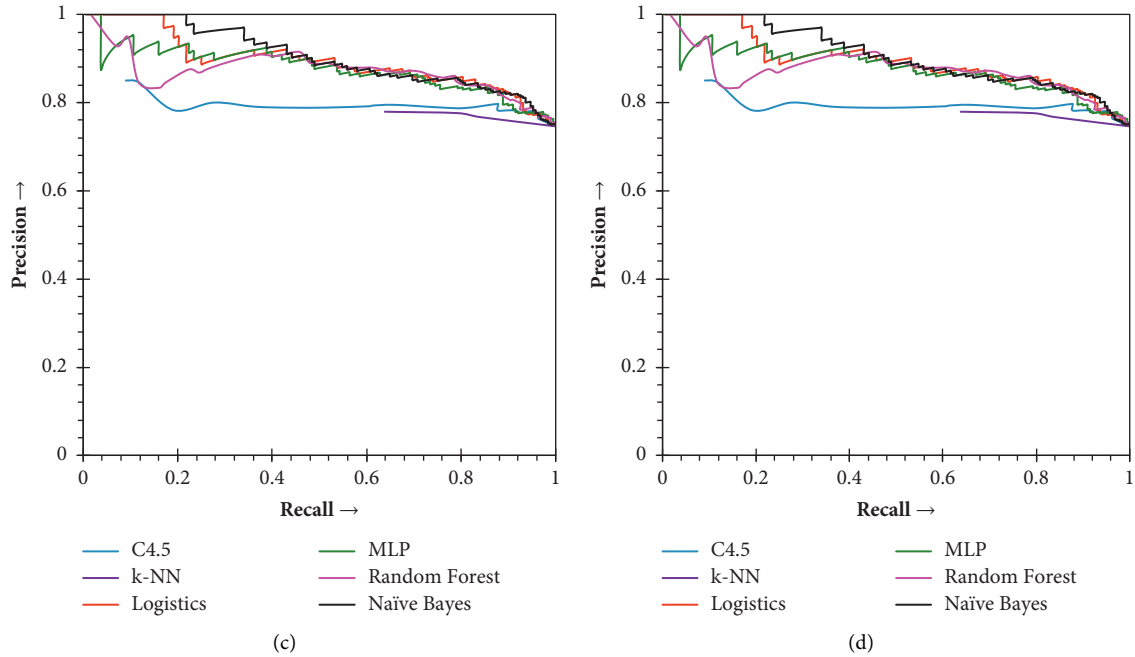


FIGURE 7: The receiver operating and precision-recall curves of the collaborative Parkinson's detection for Naive Bayes and its comparison with other supervised learning techniques. (a) Performance comparison, receiver operating curves, on control subjects. (b) Performance comparison, receiver operating curves, on Parkinson's subjects. (c) Performance comparison, precision-recall curves, on control subjects. (d) Performance comparison, precision-recall curves, on Parkinson's subjects.

are compared and analyzed using many other additional performance matrices and are presented in Table 10.

The detection result of five recent Parkinson's disease detection (PDD) schemes has been tabulated in Table 9 along with the proposed collaborative PDD scheme. All these methods used function-based approaches. It has been observed that the proposed collaborative approach claims the highest detection accuracy with the relatively lowest number of vocal features. Though the SVM approach of Kadiri et al. [22] shows 73.32% detection accuracy, which is close to our approach, but at the same time, the number of vocal features used is not clearly highlighted.

A detailed comparison through additional performance measures helps to visualize the capability of the proposed approach over other Naive Bayes approaches. For this comparison, the Avuçlu and Elen [18] and Bourouhou et al. [19] methods are taken into consideration. According to Table 10, the Avuçlu and Elen [18] method has the highest sensitivity score of 0.949. Therefore, the concerned method indicates that 94.9% of Parkinson's subjects are detected among all the Parkinson's subjects. On the other hand, our proposed PD detection model is more precise with a 0.926 precision rate. In addition, it shows the lowest false positive rate in detecting control subjects as Parkinson's.

5. Discussion, Limitations, and Future Works

Like any other detection model, the proposed method also suffers few limitations. The proposed model is based on a voice signal dataset provided by the Department of Neurology in

Cerrahpaşa, Faculty of Medicine, Istanbul. The pronunciation ascent of the sustained vowel /a/ is different for different geographical regions. As a result, the model may generate significant false positives or false negatives on the voice signals of subjects of other continents. Therefore, it is essential for further evaluation of other voice signal datasets. As future work, the proposed model can be extended to a graphical user interface mode which must have scope to be trained on varying Parkinson's signal datasets. Gender and age of subjects are other aspects that need a detailed investigation, which the proposed approach lacks. It should be noted that gender and age play a significant role in vocal performance both for control and Parkinson's subjects [36, 37]. An unbalanced dataset age and gender concerning disease pose considerable issues towards the detection process [36–39]. Therefore, the number of participants in the dataset should be balanced based on genders and age for both Parkinson's and control classes. The assessment of gender and age parameters is missing in this research work and will remain a limitation. The disease severity is another factor that allows a detector to determine the stage of the PD. In the future, the proposed work can be modeled to predict the severity of the disease.

A good Parkinson's detection dataset containing acoustic features of the subjects needs to address various factors such as the balance of gender concerning age, microphone quality, noise, the robustness of analysis procedure, number of subjects, disease severity, and influence of medication. Recently, Rusz et al. [40] presented a guideline for speech recording, which can prepare acoustic datasets for Parkinson's detection. The dataset considered here addresses and meets almost all the parameters

TABLE 9: Comparison of collaborative Parkinson's detection with other PDS through detection accuracy.

Parkinson's detection methods	Detector	Features	Accuracy (%)
Avuçlu and Elen [18]	Naïve Bayes	22	70.26
Bourouhou et al. [19]	Naïve Bayes	26	65.00
Zhang et al. [20]	Naïve Bayes	22	69.24
Meghraoui et al. [21]	Bernoulli Naïve Bayes	3	62.50
Kadiri et al. [22]	Support Vector Machine	—	73.32
Polat and Nour [25]	Linear Regression	45	77.50
Xiong and Lu [26]	Naïve Bayes	8	72.00
Mekyska et al. [28]	Classification and regression trees	8	75.19
Collaborative PD (proposed)	Naïve Bayes	7	78.97

TABLE 10: Comparison of collaborative Parkinson's detection with other PDS through additional performance measures.

Performance measures	Avuçlu and Elen [18]	Bourouhou et al. [19]	Proposed collaborative PDS
Number of vocal features selected	22	26	7
Balanced detection accuracy	0.699	0.650	0.729
Sensitivity	0.949	0.700	0.817
Specificity	0.448	0.600	0.641
Precision	0.639	0.636	0.926
False negative rate	0.051	0.300	0.183
False positive rate	0.552	0.400	0.359

stated above. However, it still fails to reveal the disease severity, which is a critical issue for any Parkinson's detection system that relies on the dataset used here. Therefore, the proposed work needs to be validated for disease severity prediction, which will make the application practical for clinical use.

Similarly, incorporating event-driven methods may improve the performance of suggested solutions in terms of computational effectiveness, compression, and power consumption [41–44]. Future work considering these aspects may be investigated.

6. Conclusion

In this article, a collaborative PDD model has been proposed. The model relies on the vocal fold, time frequency, and baseline features of both control and Parkinson's subjects. These vocal features are first ranked through correlation, fisher score, and mutual information-based feature selection schemes. The ranked features have been passed sequentially to many classifiers where Naïve Bayes evolved as the best classifier for the proposed model. The feature points are also identified based on the highest detection accuracy reported by Naïve Bayes. Relevant features are selected based on these feature points. A total of 7 ranked features has been selected from the vocal fold, time frequency, and baseline feature segments. The detection model based on the 7 ranked features shows promising detection accuracy of 78.97% and precision of 0.926, under the hold-out cross validation. The proposed model has also been compared with other function-based detection models, where our PD detection model proved to be accurate and precise. Finally, an extensive discussion has been carried out regarding the shortcoming and future direction of the proposed Parkinson's detection model.

Data Availability

The dataset used in this paper is publicly available via the UCI Machine Learning Repository with the labels and link as follows: (a) Parkinson's Disease Classification Data Set (<https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification>); (b) Italian Parkinson's voice and speech (<https://iee-dataport.org/open-access/italian-parkinsons-voice-and-speech#files>).

Conflicts of Interest

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

References

- [1] W. Poewe, K. Seppi, C. M. Tanner et al., "Parkinson disease," *Nature Reviews Disease Primers*, vol. 3, no. 1, Article ID 17013, Dec. 2017.
- [2] Q. Mao, W.-z. Qin, A. Zhang, and N. Ye, "Recent advances in dopaminergic strategies for the treatment of Parkinson's disease," *Acta Pharmacologica Sinica*, vol. 41, no. 4, pp. 471–482, Apr. 2020.
- [3] A. Agnihotri and O. I. Aruoma, "Alzheimer's disease and Parkinson's disease: a nutritional toxicology perspective of the impact of oxidative stress, mitochondrial dysfunction, nutrigenomics and environmental chemicals," *Journal of the American College of Nutrition*, vol. 39, no. 1, pp. 16–27, Jan. 2020.
- [4] F. N. Emamzadeh and A. Surguchov, "Parkinson's disease: biomarkers, treatment, and risk factors," *Frontiers in Neuroscience*, vol. 12, 2018.
- [5] L. Zahid, M. Maqsood, M. Y. Durrani et al., "A spectrogram-based deep feature assisted computer-aided diagnostic system for Parkinson's disease," *IEEE Access*, vol. 8, pp. 35482–35495, 2020.

- [6] J. Ruzs, J. Hlavnička, M. Novotný et al., "Speech biomarkers in rapid eye movement sleep behavior disorder and Parkinson disease," *Annals of Neurology*, vol. 90, 2021.
- [7] G. Dimauro, *Italian Parkinson's Voice and Speech*, IEEE Dataport, Austin, TX, US, 2019.
- [8] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, "Advances in Parkinson's Disease detection and assessment using voice and speech: a review of the articulatory and phonatory aspects," *Biomedical Signal Processing and Control*, vol. 66, Article ID 102418, Apr. 2021.
- [9] K. L. Lansford, J. M. Liss, J. N. Caviness, and R. L. Utianski, "A cognitive-perceptual approach to conceptualizing speech intelligibility deficits and remediation practice in hypokinetic dysarthria," *Parkinson's Disease*, vol. 2011, Article ID 150962, 9 pages, 2011.
- [10] R. Panigrahi, M. Pramanik, U. K. Chakraborty, and A. K. Bhoi, "Survivability prediction of patients suffering hepatocellular carcinoma using diverse classifier ensemble of grafted decision tree," *International Journal of Computer Applications in Technology*, vol. 64, no. 4, p. 349, 2020.
- [11] A. Reyana, V. T. Krishnaprasath, S. Kautish, R. Panigrahi, and M. Shaik, "Decision-making on the existence of soft exudates in diabetic retinopathy," *International Journal of Computer Applications in Technology*, vol. 64, no. 4, p. 375, 2020.
- [12] C. Quan, K. Ren, and Z. Luo, "A deep learning based method for Parkinson's disease detection using dynamic features of speech," *IEEE Access*, vol. 9, pp. 10239–10252, 2021.
- [13] P. R. Magesh, R. D. Myloth, and R. J. Tom, "An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery," *Computers in Biology and Medicine*, vol. 126, Article ID 104041, Nov. 2020.
- [14] R. Alkhatib, M. O. Diab, C. Corbier, and M. E. Badaoui, "Machine learning algorithm for gait analysis and classification on early detection of Parkinson," *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1–4, Jun. 2020.
- [15] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1/3, pp. 389–422, 2002.
- [16] Q. Gu, Z. Li, and J. Han, "Generalized Fisher score for feature selection," 2012, <https://arxiv.org/abs/1202.3725>.
- [17] M. Zaffalon and M. Hutter, "Robust feature selection using distributions of mutual information," in *Proceedings of the 18th International Conference on Uncertainty in Artificial Intelligence (UAI-2002)*, pp. 577–584, San Francisco, CA, US, June 2002.
- [18] E. Avcu and A. Elen, "Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements," *Medical, & Biological Engineering & Computing*, vol. 58, no. 11, pp. 2775–2788, Nov. 2020.
- [19] A. Bourouhou, A. Jilbab, C. Nacir, and A. Hammouch, "Comparison of classification methods to detect the parkinson disease," in *Proceedings of the 2016 International Conference on Electrical and Information Technologies (ICEIT)*, pp. 421–424, Tangiers, Morocco, May 2016.
- [20] L. Zhang, Y. Qu, B. Jin, L. Jing, Z. Gao, and Z. Liang, "An intelligent mobile-enabled system for diagnosing Parkinson disease: development and validation of a speech impairment detection system," *JMIR Medical Informatics*, vol. 8, no. 9, Article ID e18689, Sep. 2020.
- [21] D. Meghraoui, B. Boudraa, T. Merazi-Meksen, and M. Boudraa, "Parkinson's disease recognition by speech acoustic parameters classification," *Modelling and Implementation of Complex Systems*, vol. 1, pp. 165–173, 2016.
- [22] S. R. Kadiri, R. Kethireddy, and P. Alku, "Parkinson's disease detection from speech using single frequency filtering cepstral Coefficients," in *Proceedings of the Interspeech*, pp. 4971–4975, Shanghai, China, October 2020.
- [23] D. Jain, A. K. Mishra, and S. K. Das, "Machine learning based automatic prediction of Parkinson's disease using speech features," in *Proceedings of the International Conference on Artificial Intelligence and Applications*, pp. 351–362, New Delhi, India, July 2021.
- [24] C. O. Sakar, G. Serbes, A. Gunduz et al., "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," *Applied Soft Computing*, vol. 74, pp. 255–263, 2019.
- [25] K. Polat and M. Nour, "Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals," *Medical Hypotheses*, vol. 140, Article ID 109678, 2020.
- [26] Y. Xiong and Y. Lu, "Deep feature extraction from the vocal vectors using Sparse autoencoders for Parkinson's classification," *IEEE Access*, vol. 8, pp. 27821–27830, 2020.
- [27] M. Pramanik, R. Pradhan, P. Nandy, A. K. Bhoi, and P. Barsocchi, "Machine learning methods with decision forests for Parkinson's detection," *Applied Sciences*, vol. 11, no. 2, p. 581, 2021.
- [28] J. Mekyska, Z. Galáž, Z. Mzourek et al., "Assessing progress of parkinson's disease using acoustic analysis of phonation," in *Proceedings of the 2015 4th International Work Conference On Bioinspired Intelligence (IWOBI)*, pp. 111–118, June 2015.
- [29] G. Roffo, S. Melzi, and M. Cristani, "Infinite feature selection," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 4202–4210, Santiago, Chile, December 2015.
- [30] G. Roffo, S. Melzi, U. Castellani, and A. Vinciarelli, "Infinite latent feature selection: a probabilistic latent graph-based ranking approach," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1407–1415, Venice, Italy, October 2017.
- [31] C. Li and J. Xu, "Feature selection with the Fisher score followed by the Maximal Clique Centrality algorithm can accurately identify the hub genes of hepatocellular carcinoma," *Scientific Reports*, vol. 9, no. 1, p. 17283, Dec. 2019.
- [32] J. Tang, S. Alelyani, and H. Liu, "Data classification: algorithms and applications," *Data Mining and Knowledge Discovery*, pp. 37–64, CRC Press, Boca Raton, FA, US, 2014.
- [33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [34] R. Panigrahi, S. Borah, and U. K. Chakraborty, "WEKA result reader-A smart tool for reading and summarizing WEKA simulator files," *Evolution in Computational Intelligence*, Springer, vol. 1, , pp. 159–167, 2021.
- [35] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," 2013, <https://arxiv.org/abs/1302.4964>.
- [36] J. Ruzs, M. Novotný, J. Hlavnička, T. Tykalová, and E. Růžička, "High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design," *IEEE Transactions on Neural Systems*

- and Rehabilitation Engineering*, vol. 25, no. 8, pp. 1319–1321, 2016.
- [37] S. Skodda, W. Visser, and U. Schlegel, “Gender-related patterns of dysprosody in Parkinson disease and correlation between speech variables and motor symptoms,” *Journal of Voice*, vol. 25, no. 1, pp. 76–82, Jan. 2011.
- [38] J. Ruzs, J. Švihlík, P. Kryže, M. Novotný, and T. Tykalová, “Reproducibility of voice analysis with machine learning,” *Movement Disorders*, vol. 36, no. 5, pp. 1282–1283, May 2021.
- [39] P. Suphinnapong, O. Phokaewvarangkul, N. Thubthong et al., “Objective vowel sound characteristics and their relationship with motor dysfunction in Asian Parkinson’s disease patients,” *Journal of the Neurological Sciences*, vol. 426, Article ID 117487, Jul. 2021.
- [40] J. Ruzs, T. Tykalova, L. O. Ramig, and E. Tripoliti, “Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders,” *Movement Disorders*, vol. 36, no. 4, pp. 803–814, Apr. 2021.
- [41] S. Mian Qaisar, “Signal-piloted processing and machine learning based efficient power quality disturbances recognition,” *PLoS One*, vol. 16, no. 5, Article ID e0252104, 2021.
- [42] S. Mian Qaisar, “Event-driven coulomb counting for effective online approximation of Li-ion battery state of charge,” *Energies*, vol. 13, no. 21, p. 5600, 2020.
- [43] S. Mian Qaisar and A. Subasi, “Cloud-based ECG monitoring using event-driven ECG acquisition and machine learning techniques,” *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 623–634, 2020.
- [44] S. Mina Qaisar, D. Sidiya, M. Akbar, and A. Subasi, “An event-driven multiple objects surveillance system,” *International Journal of Electrical and Computer Engineering Systems*, vol. 9, no. 1, pp. 35–44, 2018.

Research Article

Kinematic Evaluation via Inertial Measurement Unit Associated with Upper Extremity Motor Function in Subacute Stroke: A Cross-Sectional Study

Ze-Jian Chen ^{1,2}, Chang He,³ Ming-Hui Gu,^{1,2} Jiang Xu ^{1,2} and Xiao-Lin Huang ^{1,2}

¹Department of Rehabilitation Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

²World Health Organization Cooperative Training and Research Center in Rehabilitation, Wuhan, Hubei, China

³Institute of Rehabilitation and Medical Robotics, State Key Lab of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China

Correspondence should be addressed to Jiang Xu; xujiang@hust.edu.cn and Xiao-Lin Huang; xiaolin2006@tjh.tjmu.edu.cn

Received 9 July 2021; Revised 4 August 2021; Accepted 11 August 2021; Published 20 August 2021

Academic Editor: G R Sinha

Copyright © 2021 Ze-Jian 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.

Kinematic evaluation via portable sensor system has been increasingly applied in neurological sciences and clinical practice. However, conventional kinematic evaluation rarely extends the context beyond the motor impairment level. In addition, kinematic tasks with numerous items could be complex and time consuming that pose a burden to test applications and data processing. The study aimed to explore the correlation of finger-to-nose task (FNT) kinematics via Inertial Measurement Unit with upper limb motor function in subacute stroke. In this study, six FNT kinematic variables were used to measure movement time, smoothness, and velocity in 37 participants with subacute stroke. Upper limb motor function was evaluated with the Fugl-Meyer Assessment for Upper Extremity (FMA-UE), Action Research Arm Test (ARAT), and modified Barthel Index (MBI). As a result, mean velocity, peak velocity, and the number of movement units were associated with the clinical assessments. The multivariable linear regression models could estimate 55%, 51%, and 32% of variance in FMA-UE, ARAT, and MBI, respectively. In addition, age, gender, type of stroke, and paretic side had no significant effects on these associations. Results show that FNT kinematic variables measured via Inertial Measurement Unit are associated with upper extremity motor function in individuals with subacute stroke. The objective kinematic evaluation may be suitable for predicting clinical measures of motor impairment and capacity to understand upper extremity motor recovery and clinical decision making after stroke. This trial is registered with ChiCTR1900026656.

1. Introduction

Upper extremity (UE) motor function is impaired in approximately 50–80% of individuals with acute stroke [1] and 40–50% with chronic stroke [2, 3]. Motor impairment results in poor movement control and has a major impact on functional capacity and activities of daily living (ADL) of stroke survivors [4]. To optimize UE recovery after stroke, it is crucial to select multilevel outcome measures for the interpretation of motor recovery and clinical decision making [5]. Although there have been extensive validated UE scales or tests to assess body structure, function, and

activity in clinical practice [6], these assessments often rely on subjectively rated ordinal scales with ceiling effects that may lead to examiner bias or lack sensitivity to detect potentially impactful changes of upper limb motor recovery [7].

Kinematic evaluation facilitates interpreting the mechanisms of motor restoration, which has been increasingly applied in neurological sciences and clinical practice [8–10]. Such technology is capable of providing detailed information regarding upper extremity function evaluation and delivering personalized interventions. According to the previous literature, kinematic assessment is usually

performed using arm-supported robots or optical-camera systems, based on fixed laboratory environments or expensive equipment that leads to several disadvantages [11–13]. From the technical perspective, robotic instruments are unable to capture the entire spectrum of UE motor impairment due to their mechanical structure [14]. Moreover, most robotic devices could not extend the value of kinematic scenarios beyond the impairment level according to the International Classification of Functioning, Disability and Health (ICF) framework [11, 15]. Optical camera systems raise privacy concerns inevitably and constrain participants into a laboratory environment with much setup time and cost [16].

Portable sensor systems, a novel approach of kinematic evaluation, can provide upper limb spatiotemporal measurements against gravity in a natural three-dimensional environment [17]. Inertial Measurement Units (IMU) are portable sensor devices combining the three-dimensional accelerometers, gyroscopes, and magnetometers to detect kinematic parameters. Kinematic analysis of motor impairment via Inertial Measurement Unit has been shown to be objective, sensitive, and quantitative. However, kinematic tasks with numerous items could be complex and time consuming that pose burden on test application, compliance issues, and data processing in previous studies [18, 19]. Moreover, its relationship with the multilevel UE clinical measures regarding ICF framework has not been fully investigated [14, 20].

In clinical practice, the finger-to-nose test (FNT) is commonly applied to evaluate upper limb coordination in patients with stroke and cerebellar ataxia [12, 21]. Compared with multi-item clinical scales that require trained personnel and as long as 30 minutes to complete, FNT could reduce task burden when estimating individual's UE performance [22]. Previous studies have shown that FNT could add value to measure UE coordination with construct, convergent, and discriminant validity [12, 23] as well as ADL-related dexterity [24, 25]. However, it remains unclear how FNT correlates with motor impairment, capacity, and ADL performance in individuals with subacute stroke from a kinematic perspective. Therefore, the purpose of this study was to explore the associations between FNT kinematic variables obtained via Inertial Measurement Unit and multilevel upper extremity motor function in subacute stroke survivors. Furthermore, we aimed to compare the amount of variance in clinical scales that could be explained by FNT kinematic variables. Hypothetically, kinematic metrics reflecting the UE movement strategy, smoothness, and velocity could be considered to measure more aspects of motor impairment (FMA-UE) than activity assessments (ARAT and MBI).

2. Materials and Methods

2.1. Study Design. This cross-sectional study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist. The study was performed in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the Clinical

Trials Ethics Committee of Huazhong University of Science and Technology on 24 October 2018. The study was registered in the Chinese Clinical Trial Registry (no. ChiCTR1900026656) on 17 October 2019.

2.2. Participants. Thirty-seven individuals with subacute stroke were recruited from the Department of Rehabilitation Medicine from December 2019 to January 2021 (Figure 1). The inclusion criteria were as follows: (a) clinical diagnosis of unilateral, first-ever subacute stroke verified by MRI or CT; (b) aged 18–80 years; (c) showing upper limb motor impairment (Fugl-Meyer Assessment of Upper Extremity <66); (d) able to complete the kinematic protocol; (e) no complicating medical history, such as visual, cardiac, or pulmonary disorders. Those who had other musculoskeletal or neurological conditions that affected arm function were excluded from the study [23]. All the participants were right handed [26] and have provided written informed consent prior to study entrance.

2.3. Clinical Assessments. Clinical assessments of the participants included the Fugl-Meyer Assessment of Upper Extremity (FMA-UE), Action Research Arm Test (ARAT), and modified Barthel Index (MBI). The FMA-UE is a validated and reliable assessment of poststroke upper limb motor impairment. FMA-UE is composed of 33 items that comprise four subscales (arm, wrist, hand, and coordination) regarding motor domains, and higher scores indicate less motor impairment of upper extremity [27]. The ARAT was used to evaluate UE functional capacity, including grasp, grip, pinch, and gross movement. ARAT consists of 19 four-point ordinal items, and higher scores indicate greater arm functional capacity [28]. The independence level in basic activities of living was assessed with the MBI, which consists of 10 items and higher scores indicate greater ADL independence [29].

2.4. Kinematic Assessment. Kinematic assessment was implemented with an Inertial Measurement Unit system (IMU, Noraxon USA Inc.). Each IMU sensor contains a coordinate system to assess accelerations and three-dimensional orientations at a 100 Hz sampling frequency. The IMU system had shown excellent reliability, accuracy, and precision in quantifying kinematic test [17]. Four sensors were placed on body segments, including head, upper arm, forearm, and hand, to detect UE kinematic information. Participants were required to sit in a height-adjustable chair with their hips and knees flexed to 90°. Upper extremity maintained in the neutral position, with elbow extension and palm downward initially. Standardized procedure for the finger-to-nose test was first presented by the same researcher, and then, it was imitated by the participants for three times before testing. The tests were recorded for five times, and a mean of the variable was used in statistical calculations [30].

Data were extracted through a semiautomated code in MATLAB software (The MathWorks, Natick,

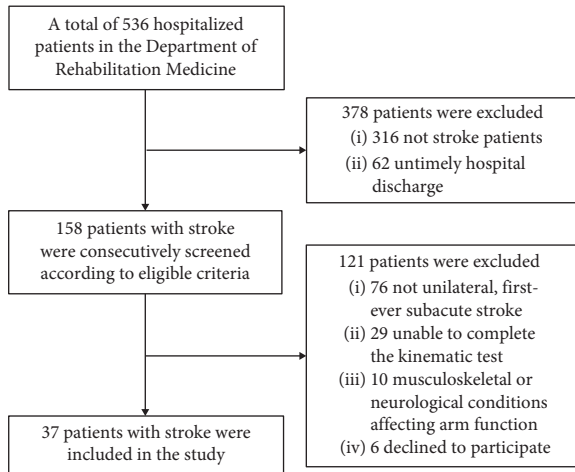


FIGURE 1: Flowchart of the study.

Massachusetts, USA) according to the anatomical coordinate system and joint rotation recommended by the International Society Biomechanical (ISB) [31]. Onsets and ends of FNT movements were defined with a velocity threshold of 50 mm/s [30]. In this cross-sectional study, six FNT kinematic variables were calculated: movement time (MT), mean velocity (VM), peak velocity (VP), percentage of time to peak velocity (TVP%), number of movement units (NMU), and normalized integrated jerk (NIJ) [30, 32]. MT was an objective quantitative variable defined as the time spent during the test to reflect movement performance. The maximum tangential velocity of index finger was calculated during each movement segment to get VP; VM was defined as the average tangential velocity. TVP% was the proportion of time taken from the onset of the movement to the peak velocity. The number of velocity peaks exceeding 10% of VP was characterized as NMU. NIJ was utilized to assess movement smoothness, which was calculated using jerk, M,T and length of the task according to the following formula:

$$NIJ = \sqrt{\frac{MT^5}{2 \times \text{length}^2} \times \sum \text{jerk}(t)^2}, \quad (1)$$

where jerk represented the third derivate of end point displacement, and length represented the shortest distance between initial and terminal positions of index finger.

2.5. Statistical Analysis. Statistical analysis was performed on IBM Statistical Package for Social Science (SPSS) version 22.0. Chi-squared test was used to examine categorical variables, and one-way ANOVA was used to examine quantitative variables. Shapiro–Wilk test or Q-Q plot was used to evaluate whether the quantitative data were normally distributed. Pearson’s correlation coefficients (r) were conducted between kinematic and clinical assessments. The limit for multicollinearity among independent variables was set at 0.7 for correlation coefficients. After controlling the influencing factors (including age, gender, type of stroke,

and paretic side), the kinematic metrics were included as independent variables into the multivariable linear regression to investigate the associations with clinical assessments. Probability for entry in backward regression was set at 0.05 and removal at 0.10. Adjusted R^2 values with P value, unstandardized coefficient (β), and unique partial correlation coefficients were used to estimate the contribution of each metrics to the models. A two-sided $P < 0.05$ was set as statistical significance.

3. Results

3.1. Demographics and Clinical Characteristics. Demographics and clinical characteristics of the participants are presented in Table 1. Thirty-seven individuals (28 male, aged 49.78 ± 10.26 years) with subacute stroke were recruited in this study from December 2019 to January 2021 (Figure 1). They had moderate-to-severe UE motor impairment (mean FMA-UE scores, 36.22 ± 17.69) and capacity (mean ARAT scores, 23.97 ± 17.38). Of the 37 participants, 26 (70.3%) had ischemic stroke and 11 (29.7%) had hemorrhagic stroke; 22 (59.5%) had left-sided hemiplegia and 15 (40.5%) had right-sided hemiplegia (Table 1).

3.2. Correlations between Clinical and Kinematic Measures. Correlations between clinical and kinematic measures are shown in Table 2. Mean velocity strongly correlated with the FMA-UE ($r = 0.85$, $P < 0.01$) and ARAT ($r = 0.80$, $P < 0.01$) and moderately correlated with MBI positively ($r = 0.58$, $P < 0.01$). Besides, all the clinical assessments correlated significantly with VP positively ($r = 0.55$ to 0.81 , $P < 0.01$) and NMU negatively ($r = -0.45$ to -0.65 , $P < 0.05$). However, MT, TVP%, and NIJ were not significantly associated with the clinical assessments (Table 2). As shown in Table 2, multicollinearities were observed between MT and NIJ, as well as among VM, VP, and NMU. For that reason, only the VM and NIJ/MT during the FNT task were inputted into the multivariable linear regression models to estimate variation in clinical assessments.

The results of multivariable regression analysis of the kinematic metrics against the clinical assessments are presented in Table 3. Backward multiple regression revealed that kinematic variables could explain the largest amount of variance in the assessment of UE motor impairment as measured by FMA-UE. The only significant predictor was the VM, which explained 55% of the FMA-UE variance ($F = 20.72$, $P < 0.01$). Moreover, the VM alone showed a significant contribution to the models, accounting for 51% of the ARAT variance ($F = 39.10$, $P < 0.01$) and 32% of the MBI variance ($F = 8.93$, $P < 0.01$) (Table 3). Moreover, demographics, including age, gender, type of stroke, and paretic side, showed no significant influence in any regression model.

4. Discussion

The cross-sectional study investigated the associations between FNT kinematic variables obtained via Inertial Measurement Unit and upper extremity motor function in

TABLE 1: Demographics and clinical characteristics ($n = 37$).

Characteristics ($n = 37$)	
Age (years)	49.78 ± 10.26
Gender (M/F)	28/9
Days between onset and enrollment	106.30 ± 65.46
Type of stroke (ischemic/hemorrhagic)	26/11
Paretic side (left/right)	22/15
FMA-UE (range 0–66)	36.22 ± 17.69
ARAT (range 0–57)	23.97 ± 17.38
MBI (range 0–100)	72.30 ± 22.20
Body mass index (kg/m ²)	24.43 ± 2.60
MT (s)	1.09 ± 0.31
VP (m/s)	1.61 ± 0.92
VM (m/s)	0.78 ± 0.44
TVP% (%)	42.23 ± 11.30
NMU	2.56 ± 1.25
NIJ	2.86 ± 1.98

FMA-UE, Fugl-Meyer assessment for upper extremity; ARAT, action research arm test; MBI, Modified Barthel index; MT: movement time; VP: peak velocity; VM: mean velocity; TVP%: percentage of time to peak velocity; NMU: Number of movement units; NIJ: Normalized integrated jerk.

TABLE 2: Correlations between clinical assessments and kinematic metrics ($n = 37$).

	FMA	ARAT	MBI	MT	VP	VM	TVP%	NMU
<i>Clinical assessment</i>								
FMA								
ARAT	0.94**							
MBI	0.62**	0.64**						
<i>Kinematic metrics</i>								
MT	0.11	0.07	0.15					
VP	0.81**	0.76**	0.55**	-0.03				
VM	0.85**	0.80**	0.58**	-0.11	0.96**			
TVP%	-0.11	-0.17	-0.14	-0.49*	-0.17	-0.11		
NMU	-0.65**	-0.59**	-0.45*	0.10	-0.70**	-0.74**	0.22	
NIJ	-0.24	-0.27	-0.10	0.71**	-0.27	-0.40**	-0.20	0.47*

FMA-UE, Fugl-Meyer Assessment for Upper extremity; ARAT, Action Research Arm Test; MBI, Modified Barthel Index; MT: movement time; VP: peak velocity; VM: mean velocity; TVP%: percentage of time to peak velocity; NMU: Number of movement units; NIJ: Normalized integrated jerk. *: $P < 0.05$. **: $P < 0.01$.

TABLE 3: Multivariable regression analysis of the kinematic metrics against the clinical assessments ($n = 37$).

Independent variables	Unstandardized coefficient β	Standard error	Partial unique correlations	P value of the variable	Adjusted R^2 (model P value)
<i>FMA-UE as dependent variable</i>					0.55 (<0.01*)
Constant	11.91	2.59	—	<0.01*	
VM	17.81	2.90	0.74	<0.01*	
NIJ	1.05	0.68	0.25	0.13	
<i>ARAT as dependent variable</i>					0.51 (<0.01*)
Constant	1.50	4.11	—	0.72	
VM	28.84	4.61	0.73	<0.01*	
<i>MBI as dependent variable</i>					0.32 (0.01*)
Constant	32.57	12.23	—	0.19	
VM	28.93	7.09	0.57	<0.01*	
MT	15.70	10.05	0.26	0.13	

FMA-UE, Fugl-Meyer Assessment for Upper Extremity; ARAT, action research arm test; MBI, Modified Barthel Index; MT: movement time; VM: mean velocity; NIJ: normalized integrated jerk. *: $P < 0.05$.

subacute stroke survivors according to the ICF framework. The results indicated that the mean velocity ($r = 0.58$ to 0.85), peak velocity ($r = 0.55$ to 0.81), and number of movement units ($r = -0.45$ to -0.65) were associated with all of the clinical assessments. Mean velocity was entered into the multivariable linear regression models and could estimate 55%, 51%, and 32% of variance in FMA-UE, ARAT, and MBI during the FNT task. Additionally, age, gender, type of stroke, and paretic side had no significant effects on these associations.

The previous kinematic literature mainly focused on predicting upper limb motor impairment. Our results extended the value of kinematic scenarios beyond the impairment level according to the ICF framework and suggested that FNT kinematics was more strongly associated with FMA-UE and ARAT than MBI. Lee et al. proposed an automated FMA system and showed high scoring accuracy in 79% of FMA test in nine stroke patients [33]. Their algorithms were shown appropriate for clinical use but lacked clinical interpretability in kinematic results because estimating clinical scale was not the only goal of portable sensors [14, 16, 34]. In addition, little was known about the associations between clinical activity-related scales and IMU sensors [35]. Our models established predictable correlations between FNT mean velocity via Inertial Measurement Unit and upper extremity motor function after stroke [36].

Due to multicollinearity among VM, VP, and NMU, only VM was entered into the multivariable models. Speed variables reflect how efficiently a person controls interaction torques of the agonist/antagonist muscles [32]. Analogous to our results, two studies using robotic device showed significant correlations between movement speed and FMA-UE in individuals with subacute [37] and chronic [38] stroke, respectively. Furthermore, movement smoothness is an important indicator of upper limb motor recovery after stroke [30]. Smoothness parameters evaluate the temporal organization or UE multijoint coordination [39]. In a study early after stroke, smoothness measured by NMU was able to predict upper limb motor recovery over time [40]. According to our results, NIJ was not significantly associated with the clinical assessments. However, smoothness should be interpreted with cautions because a single smoothness parameter may not completely reflect motor recovery of upper extremity [41].

Interestingly, the FNT kinematic metrics measured comparable aspects of motor impairment by FMA-UE ($R^2 = 0.55$) and functional capacity by ARAT ($R^2 = 0.51$). This was analogous with a prior work by Adans-Dester et al., which used eight motor tasks of Wolf Motor Function Test (WMFT) and found satisfactory results to estimate upper limb impairment and activity scales [42]. Although this was in line with our second hypothesis, the difference was small and needed to be further studied [19]. One possible explanation might be that participants had moderate-to-severe upper extremity motor impairment, leading to poor scores on manual dexterity of ARAT items. Future studies should therefore include much kinematic variables and comprehensive tasks at different UE segments to explore the correlations between the assessments. The low variance

explained by IMU kinematic variables in MBI could be that the FNT task did not measure distal dexterity of the upper extremity. As a result, variables in the models may not fully capture kinematic information in individuals with stroke [43]. Moreover, MBI is a questionnaire for ADL in a real environment and not an observational measure of UE motor function in an experimental setting. Thus, participants could use compensatory behaviors or actually the less affected UE to improve MBI scores, which may be difficult to explain with the current kinematic task.

Several limitations of this study should be acknowledged. First, the sample size restricted the number of kinematic variables entered into the multivariable linear regression models. Therefore, future studies could implement other statistical models, such as machine-learning approaches, to investigate the associations between FNT kinematic variables and upper extremity motor function in individuals with stroke [42, 44]. Second, this was a cross-sectional study and unable to investigate the longitudinal associations between kinematics and clinical measurements. Finally, the models did not include other kinematic tests and variables concerning trunk and interjoint movements, which may lead to task-related bias and loss of information [45, 46].

5. Conclusions

This study indicates that kinematic variables measured via Inertial Measurement Unit during the finger-to-nose task are associated with upper extremity motor function in individuals with subacute stroke according to the ICF framework. Furthermore, the objective kinematic evaluation may be suitable for predicting clinical measures of motor impairment and capacity to understand upper extremity motor recovery and clinical decision making after stroke.

Data Availability

The data files are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

ZJC and XLH contributed to the conception and design of the study. ZJC and CH contributed to data collection. ZJC carried out the data analysis and wrote the first draft of the manuscript. ZJC and MHG revised the manuscript. All the authors reviewed and approved the final version of the manuscript.

Acknowledgments

This work received financial support for the research and publication of this article from National Natural Science Foundation of China (U 1913601 and No. 91648203).

References

- [1] P. Langhorne, F. Coupar, and A. Pollock, "Motor recovery after stroke: a systematic review," *The Lancet Neurology*, vol. 8, no. 8, pp. 741–754, 2009.
- [2] G. Kwakkel, B. J. Kollen, J. van der Grond, and A. J. H. Prevo, "Probability of regaining dexterity in the flaccid upper limb," *Stroke*, vol. 34, no. 9, pp. 2181–2186, 2003.
- [3] D. Lloyd-Jones, R. J. Adams, T. M. Brown et al., "Heart disease and stroke statistics--2010 update: a report from the American Heart Association," *Circulation*, vol. 121, no. 7, pp. e46–e215, 2010.
- [4] C. Stinear, "Prediction of recovery of motor function after stroke," *The Lancet Neurology*, vol. 9, no. 12, pp. 1228–1232, 2010.
- [5] C. Villepinte, A. Verma, C. Dimeglio, X. De Boissezon, and D. Gasq, "Responsiveness of kinematic and clinical measures of upper-limb motor function after stroke: a systematic review and meta-analysis," *Annals of Physical and Rehabilitation Medicine*, vol. 64, 2020.
- [6] L. Santisteban, M. Térémetz, J.-P. Bleton, J.-C. Baron, M. A. Maier, and P. G. Lindberg, "Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review," *PLoS One*, vol. 11, no. 5, Article ID e0154792, 2016.
- [7] C. E. Lang, M. D. Bland, R. R. Bailey, S. Y. Schaefer, and R. L. Birkenmeier, "Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making," *Journal of Hand Therapy*, vol. 26, no. 2, pp. 104–115, 2013.
- [8] A. Schwarz, M. M. C. Bhagubai, G. Wolterink, J. P. O. Held, A. R. Luft, and P. H. Veltink, "Assessment of upper limb movement impairments after stroke using wearable inertial sensing," *Sensors*, vol. 20, no. 17, 2020.
- [9] F. Grimm, J. Kraugmann, G. Naros, and A. Gharabaghi, "Clinical validation of kinematic assessments of post-stroke upper limb movements with a multi-joint arm exoskeleton," *Journal of NeuroEngineering and Rehabilitation*, vol. 18, no. 1, p. 92, 2021.
- [10] Z. J. Chen, C. He, F. Guo, C. H. Xiong, and X. L. Huang, "Exoskeleton-assisted anthropomorphic movement training (eamt) for post-stroke upper limb rehabilitation: a pilot randomized controlled trial," *Archives of Physical Medicine and Rehabilitation*, vol. 6, 2021.
- [11] S. Balasubramanian, R. Colombo, I. Sterpi, V. Sanguineti, and E. Burdet, "Robotic assessment of upper limb motor function after stroke," *American Journal of Physical Medicine and Rehabilitation*, vol. 91, pp. S255–S269, 2012.
- [12] G. M. Johansson, H. Grip, M. F. Levin, and C. K. Häger, "The added value of kinematic evaluation of the timed finger-to-nose test in persons post-stroke," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 11, 2017.
- [13] C. M. L. Hughes, M. Baye, C. Gordon-Murer et al., "Quantitative assessment of upper limb motor function in Ethiopian acquired brain injured patients using a low-cost wearable sensor," *Frontiers in Neurology*, vol. 10, p. 1323, 2019.
- [14] P. Maceira-Elvira, T. Popa, A.-C. Schmid, and F. C. Hummel, "Wearable technology in stroke rehabilitation: towards improved diagnosis and treatment of upper-limb motor impairment," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, p. 142, 2019.
- [15] I. C. F. Who, *International Classification of Functioning, Disability and Health*, World Health Organization, Geneva, Switzerland, 2001.
- [16] B. Oubre, J.-F. Daneault, H.-T. Jung et al., "Estimating upper-limb impairment level in stroke survivors using wearable inertial sensors and a minimally-burdensome motor task," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 3, pp. 601–611, 2020.
- [17] F. Öhberg, T. Bäcklund, N. Sundström, and H. Grip, "Portable sensors add reliable kinematic measures to the assessment of upper extremity function," *Sensors*, vol. 19, no. 5, 2019.
- [18] S. Lee, Y.-S. Lee, and J. Kim, "Automated evaluation of upper-limb motor function impairment using fugl-meyer assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 125–134, 2018.
- [19] S. Del Din, S. Patel, C. Cobelli, and P. Bonato, "Estimating Fugl-Meyer clinical scores in stroke survivors using wearable sensors," *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2011, pp. 5839–5842, 2011.
- [20] A. Schwarz, C. M. Kanzler, O. Lambercy, A. R. Luft, and J. M. Veerbeek, "Systematic review on kinematic assessments of upper limb movements after stroke," *Stroke*, vol. 50, no. 3, pp. 718–727, 2019.
- [21] H. Tran, K. D. Nguyen, P. N. Pathirana, M. K. Horne, L. Power, and D. J. Szmulewicz, "A comprehensive scheme for the objective upper body assessments of subjects with cerebellar ataxia," *Journal of NeuroEngineering and Rehabilitation*, vol. 17, no. 1, p. 162, 2020.
- [22] P. Otten, J. Kim, and S. Son, "A framework to automate assessment of upper-limb motor function impairment: a feasibility study," *Sensors*, vol. 15, no. 8, Article ID 20097, 2015.
- [23] M. R. M. Rodrigues, M. Slimovitch, G. Chilingaryan, and M. F. Levin, "Does the Finger-to-Nose Test measure upper limb coordination in chronic stroke?" *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, p. 6, 2017.
- [24] W. Chen, C. Xiong, X. Huang, R. Sun, and Y. Xiong, "Kinematic analysis and dexterity evaluation of upper extremity in activities of daily living," *Gait & Posture*, vol. 32, no. 4, pp. 475–481, 2010.
- [25] K. Liu, C.-H. Xiong, L. He, W.-B. Chen, and X.-L. Huang, "Postural synergy based design of exoskeleton robot replicating human arm reaching movements," *Robotics and Autonomous Systems*, vol. 99, pp. 84–96, 2018.
- [26] M. Verdino and S. Dingman, "Two measures of laterality in handedness: the Edinburgh Handedness Inventory and the Purdue Pegboard test of manual dexterity," *Perceptual & Motor Skills*, vol. 86, no. 2, pp. 476–478, 1998.
- [27] J. See, L. Dodakian, C. Chou et al., "A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials," *Neurorehabilitation and Neural Repair*, vol. 27, no. 8, pp. 732–741, 2013.
- [28] J. H. Van der Lee, V. De Groot, H. Beckerman, R. C. Wagenaar, G. J. Lankhorst, and L. M. Bouter, "The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 1, pp. 14–19, 2001.
- [29] Y.-W. Hsieh, C.-H. Wang, S.-C. Wu, P.-C. Chen, C.-F. Sheu, and C.-L. Hsieh, "Establishing the minimal clinically important difference of the Barthel Index in stroke patients," *Neurorehabilitation and Neural Repair*, vol. 21, no. 3, pp. 233–238, 2007.
- [30] M. L. Schiefelbein, A. P. Salazar, R. R. Marchese et al., "Upper-limb movement smoothness after stroke and its relationship with measures of body function/structure and activity - a

- cross-sectional study,” *Journal of the Neurological Sciences*, vol. 401, pp. 75–78, 2019.
- [31] G. Wu, F. C. T. van der Helm, H. E. J. Veeger et al., “ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion-Part II: shoulder, elbow, wrist and hand,” *Journal of Biomechanics*, vol. 38, no. 5, pp. 981–992, 2005.
- [32] N. Nordin, S. Xie, and B. Wünsche, “Assessment of movement quality in robot- assisted upper limb rehabilitation after stroke: a review,” *Journal of NeuroEngineering and Rehabilitation*, vol. 11, no. 1, p. 137, 2014.
- [33] E. Park, K. Lee, T. Han, and H. S. Nam, “Automatic grading of stroke symptoms for rapid assessment using optimized machine learning and 4-limb kinematics: clinical validation study,” *Journal of Medical Internet Research*, vol. 22, no. 9, Article ID e20641, 2020.
- [34] Z. Nelson and E. Wade, “Relative efficacy of sensor modalities for estimating post-stroke motor impairment,” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2018, pp. 2503–2506, 2018.
- [35] J. Li, B. Pan, T. Jin et al., “A single task assessment system of upper-limb motor function after stroke,” *Technology and Health Care*, vol. 24, pp. S707–S715, 2016.
- [36] N. Hussain, K. S. Sunnerhagen, and M. Alt Murphy, “Recovery of arm function during acute to chronic stage of stroke quantified by kinematics,” *Journal of Rehabilitation Medicine*, vol. 53, 2021.
- [37] C. Duret, O. Courtial, and A. G. Grosmaire, “Kinematic measures for upper limb motor assessment during robot-mediated training in patients with severe sub-acute stroke,” *Restorative Neurology and Neuroscience*, vol. 34, no. 2, pp. 237–245, 2016.
- [38] L. Zollo, E. Gallotta, E. Guglielmelli, and S. Sterzi, “Robotic technologies and rehabilitation: new tools for upper-limb therapy and assessment in chronic stroke,” *European Journal of Physical and Rehabilitation Medicine*, vol. 47, no. 2, pp. 223–236, 2011.
- [39] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, “On the analysis of movement smoothness,” *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, p. 112, 2015.
- [40] L. van Dokkum, I. Hauret, D. Mottet, J. Froger, J. Métrot, and I. Laffont, “The contribution of kinematics in the assessment of upper limb motor recovery early after stroke,” *Neurorehabilitation and Neural Repair*, vol. 28, no. 1, pp. 4–12, 2014.
- [41] B. Rohrer, S. Fasoli, H. I. Krebs et al., “Movement smoothness changes during stroke recovery,” *Journal of Neuroscience*, vol. 22, no. 18, pp. 8297–8304, 2002.
- [42] C. Adans-Dester, N. Hankov, A. O’Brien et al., “Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery,” *Npj Digital Medicine*, vol. 3, no. 1, p. 121, 2020.
- [43] C. Bosecker, L. Dipietro, B. Volpe, and H. Igo Krebs, “Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke,” *Neurorehabilitation and Neural Repair*, vol. 24, no. 1, pp. 62–69, 2010.
- [44] Z.-J. Chen, C. He, N. Xia et al., “Association between finger-to-nose kinematics and upper extremity motor function in subacute stroke: a principal component analysis,” *Frontiers in Bioengineering and Biotechnology*, vol. 9, Article ID 660015, 2021.
- [45] A. Schwarz, J. M. Veerbeek, J. P. O. Held, J. H. Buurke, and A. R. Luft, “Measures of interjoint coordination post-stroke across different upper limb movement tasks,” *Frontiers in bioengineering and biotechnology*, vol. 8, Article ID 620805, 2020.
- [46] G. J. Kim, A. Parnandi, S. Eva, and H. Schambra, “The use of wearable sensors to assess and treat the upper extremity after stroke: a scoping review,” *Disability & Rehabilitation*, vol. 43, pp. 1–20, 2021.

Research Article

Initial Geometrical Templates with Parameter Sets for Active Contour on Skin Cancer Boundary Segmentation

Prachya Bumrungkun,¹ Kosin Chamnongthai ¹ and Wisarn Patchoo ²

¹Department of Electronic and Telecommunication Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi, 126 Pracha Uthit Road, Bangmod, Thung Khru, Bangkok 10140, Thailand

²School of Engineering, Bangkok University, Bangkok 10110, Thailand

Correspondence should be addressed to Kosin Chamnongthai; kosin.cha@kmutt.ac.th

Received 21 May 2021; Revised 4 July 2021; Accepted 8 July 2021; Published 4 August 2021

Academic Editor: G. R. Sinha

Copyright © 2021 Prachya Bumrungkun 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.

For active-contour-based surgery systems, the success of skin cancer boundary segmentation depends on the initialization point of the snake model, which is a task originally performed by skillful experts, and on the parameters set for the algorithms of active contour. This paper proposes initial geometrical templates and parameter sets for the active contour on skin cancer boundary segmentation. To establish initial geometrical templates and parameter sets for the active contour, first, template candidates, which are geometrically designed by users in advance, are simply calculated based on similarity with a skin cancer boundary, and the candidate with the least difference is selected as an initial template. Initially, all candidate templates are performed before the test with some selected skin cancer samples by randomly varying needed parameters to determine parameter sets for each template. The parameter set is therefore implicitly selected as the suitable set with the selected initial template. Experiments with 227 skin cancer samples were performed based on our proposed initial templates and parameter sets, and the results show 99.46% accuracy, 97.43% sensitivity, and 99.87% specificity approximately in which accuracy, sensitivity, and specificity were improved by 0.26%, 0.36%, and 0.26%, respectively, compared with the conventional method.

1. Introduction

According to a WHO's (World Health Organization) report in 2016 [1], cancer is the worst death cause for humans. Among cancer types including breast, cervical, and lung, skin cancer was ranked as number 19 of human death causes, and the number of patients dramatically increase by approximately 1.7 million in 2016 [1] due to stronger ultraviolet (UV) in recent years. Among several approaches to cure skin cancer including surgery, radiation, and photodynamic therapy, surgery comprising Mohs microsurgery, laser surgery, and electrodesiccation and curettage are currently widely accepted as effective methods with less pain. However, surgery basically depends on skillful medical doctors who are limited in number, and treatment is normally costly [2] such that the existence of an automatic skin cancer surgery system would be useful to assist medical doctors curing patients.

One of the important fundamental functions of the automatic skin cancer surgery system is skin cancer boundary segmentation [3]; so many researchers are trying to focus on the research problem of segmentation of the skin cancer boundary to ensure the success of automatic surgery. In fact, the skin cancer boundary complicatedly consists of many tiny curves and angles with low contrast in some parts. It truly becomes difficult to accurately segment the boundary for automatic skin cancer surgery. In case that the segmentation is not properly performed, the skin cancer is not completely removed due to reduced segmentation, and the cancer may subsequently spread throughout the entire body. Thus, some neighboring normal flesh is removed with the skin cancer as buffer because the medical doctor intends to ensure that all cancer is removed. Therefore, if automatic segmentation efficiently functions, it ensures that all of the skin cancer is removed,

and the pain caused by removing some neighboring flesh needlessly is simultaneously eased.

Previous research works of segmentation, especially those related to automatic skin cancer surgery, could be divided into a couple approaches that are supervised and unsupervised. The first group of supervised-based approaches [4–7] analyzed images of skin cancer and utilized existing image processing and machine-learning tools to segment the boundary of skin cancers. Some of these studies [4] focused upon cancer detection and discuss the benefits and costs in terms of automatic cancer detection and assistance systems. However, to implement an automatic skin cancer surgery system, segmentation fundamentally becomes a crucial basic function and vitally requires high accuracy. Moreover, the aforementioned supervised methods are evaluated to yield good results in skin cancer segmentation and detection.

In another unsupervised approach, Castillejos et al. [8] proposed wavelet-transformed fuzzy algorithms for dermoscopic image segmentation. This method used feature extraction in wavelet transform space before proceeding to the segmentation process, and three-color channels (RGB space) in wavelet-transformed space gather the color channels via the nearest neighbor interpolation (NNI). This type of preprocess using existing mathematic tools and some machine learning algorithms that were discussed in the supervised approach is highly evaluated as good mathematic segmentation methods, but the bad cases with negative faults are critically outstanding [9, 10]. These models were acceptable in applications of skin cancer detection that decide the cancer boundaries. Although methods used for active contours simultaneously provide poorer results [9, 10], negative faults were found less often compared with the mentioned methods. This was the critical point for the automatic surgery system that medically requires segmenting skin cancer boundaries accurately since the bad segmentation cases were seriously regarded as negative results for the surgery. In this case, the skin cancer was not completely removed and ultimately not cured. In the subgroup of active contour usage, Munir et al. [11] recently proposed adaptive active contours based on the variable kernel with constant initialization. This mathematically incorporated a force term that pushed the contour towards the object boundary by using a regularization term that has taken into account the smoothness of the level set function and an edge term that helped to stop the contour at required boundaries. This system achieved high accuracy, but initialization is not completely automatic to date. On the other hand, Mogali et al. [12] proposed template-based active contours using a generalized active contour formalism for image segmentation based on shape templates, and the shape template is subjected to a restricted affine transformation (RAT), which allows for translation, rotation, and scaling. The segmentation functions excellently for any shapes. Kirimasthong et al. [13] proposed a method of automatic initialization of GVF-type snakes in ultrasound images of breast cancer. The method was proved to deal well with ultrasound images of breast cancer. Rodtook et al. [14] proposed an automatic initialization of active contours and a

level set method in ultrasound images of breast abnormalities. The method successfully dealt with ultrasound images of malignant tumors, fibroadenomas, and cysts. Nevertheless, the complexity of automatic initialization remained as a problem in these mentioned methods for system implementation and to set parameters (weighting factors, iteration steps, etc.) for active contours depending on human skill [15] and it was not clearly reproducible.

The authors of this paper hence have focused on the research problems of automatic initialization and parameter setting for active contours. The initialization and parameter setting are considered in this paper to be improved in terms of human skill independence, simpleness, and reproducibility. Although deep learning tool as convolution neural network (CNN) is recently accepted as powerful for classification, it basically needs a huge number of samples for training, which may not be suitable for some medical problems. The authors first sought to perform experiments on skin cancer images using SVM and snake model using a semiautomated method [16]. It was confirmed to work well with some skin cancer samples but needed to be improved as a fully automatic method. The contribution of this paper hence is to create an algorithm to automatically initialize the active contour using a geometric template which is automatically selected from a group of geometric shape candidates based on some samples trained in advance. Since there exist many parameters in active contours which vitally influence convergence of the contours, parameters, which are matched with selected geometric template shapes, are automatically selected based on prior training of parameter sets. The active contours therefore can simply perform segmentation by the independence of any human skills.

This paper is constructed as follows: analysis of initialization and parameter setting for active contour-based segmentation is reported in Section 2. An overview of the imagined automatic surgery system and the proposed method are described in Section 3. Section 4 demonstrates the experimental results using proposed templates and parameter sets. The discussion of the selected initial templates and parameter sets is explained in Section 5. Finally, the conclusion is presented in Section 6.

2. Analysis of Initialization and Parameter Setting for Active-Contour-Based Segmentation

It is a fact that segmentation obtained by active contours is accepted as an excellent one, but it depends upon initialization and parameters based on human skill. If initialization is fit with the object shape, segmentation and active contours may be performed appropriately. As shown in Figure 1(a) where blue and red lines represent initialization manually performed by an expert and segmentation done by active contours, respectively, the segmentation result is observed to converge appropriately. In the opposite way, when the initialization is not fit with the object shape shown by a blue line, the red line representing segmentation is observed to converge inside the object, as shown in Figure 1(b). It is

obvious that initialization is one of the crucial keys for active contours, and originally it depends on human skill. Although some researchers have achieved good results to create algorithms for automatic initialization using seed explosion [13, 14], complexity remains as important issue that will be discussed in this paper. The approach of conventional methods is basically based on random distribution which is regarded as a good way to cover any kinds of shapes. Observing the shape of appropriate initialization, it looks similar to the segmentation but closely bigger. It is also very difficult to create such kind of the shape which is similar to any shapes of the skin cancer, because it is initially unknown. Our solution in this paper is to scope numbers of candidate shapes using similar geometric ones. As shown by an example of circle in Figure 1(c), although the number of convergence loops observed by thickness of the red lines is accumulated more than the case of good initialization, as shown in Figure 1(a), it finally can converge in the similar level with the initialization done by an expert.

On the other hand, it is also observed that parameters such as weight factors, iteration steps, alpha, beta, kappa, wline, wedge, and wterm influence the segmentation results. For instance, appropriate parameters are set on those cases, as shown in the first row of Figure 1, and all initializations converge well except for the bad initialization, as shown in Figure 1(b). If inappropriate parameters are applied to those cases, segmentation results represented by red lines reveal incorrect convergence with many loops. Obviously, this means parameters are another key for active contours.

As discussed earlier in Figure 1, initialization whose shape is similar to the skin cancer is the most preference for active contour segmentation, and an appropriate geometric shape is regarded as another choice used as a basic concept in this paper. Moreover, parameter setting is another crucial factor that controls active contour to converge appropriately. As shown in Figure 2, a skin cancer boundary is applied by different geometric templates which are rectangle, ellipse, and circle, as shown in row directions, respectively, and three parameter sets, which are assumed to be suitable for initialization, too less convergence, and too much convergence, are used in the column directions, respectively. These figures show the appropriate geometric shape with appropriate parameter set is preference, as shown in Figure 2(a), while others are unworkable based on inappropriate conditions of either initiation or parameter set, as shown in Figures 2(b)–2(i). This means both appropriate initiation and parameter set are really required in the implementation of fully automatic active contours, and the authors of this paper would find an algorithm to determine a geometric shape as initiation with a set of parameters which initially was fixed with the selected geometric shape.

Based on [17], the active contour or snake is defined as a deformable curve ($\mathbf{X}(s) = [\mathbf{x}(s), \mathbf{y}(s)]$), where $s \in [0, 1]$, and it is adjusted to minimize the energy (E) in the following equation:

$$E = \int_0^1 (\alpha |\mathbf{X}'(s)|^2 + \beta |\mathbf{X}''(s)|^2 + E_{\text{ext}}(\mathbf{X}(s))) ds, \quad (1)$$

where α and β are weighting parameters corresponding to elasticity and stiffness of the snake, respectively, and they are assumed to be uniform for all. $\mathbf{X}'(s)$ and $\mathbf{X}''(s)$ are 1st and 2nd order derivative of $\mathbf{X}(s)$ with respect to s .

In operation, the snake curve represented by the first and second terms in (1), which exactly implies snake initialization, may be adjusted through the image spatial domain to reach the external energy E_{ext} representing image features. At that time, the total energy (E) should reach minima or even none, and the image boundary is regarded to be obtained. Since, normally, a shape of a skin cancer boundary is arbitrary, the snake operation sometimes cannot be performed accurately to segment the boundary due to some local minimum even logically reaching no total energy.

As the research theme of this paper mentioned earlier, the authors of this paper try to find a simple way to automatically initialize for practical snake operation for practical applications. Mathematically, geometric shapes can be approximated as candidates of the initialization, and the criteria of geometric shape selection should be an issue to discuss.

If some geometrical shapes (T_i) are conceptually approximated in terms of template as initialization, a geometric shape (T_p), which is selected based on the most similarity to the skin cancer image (W) from a group of candidate geometric shapes including circle, ellipse, triangle, rectangle, pentagon, hexagon, possibly differed the least compared with the skin cancer image, as shown in Figure 3. The geometrical template can be selected as follows:

$$T_p = \min_{i=1,2,\dots,n} T_i - W, \quad (2)$$

where n is the number of template candidates.

The geometrical template candidates (T_i), in which their scale is determined based on the maximum skeleton (l_{max}) of the rough skin cancer boundary and their posture is varied by all angles (θ_j) starting from an initial angle, can be expressed as follows:

$$T_i = T_i \{ l_{\text{max}}, \theta_j \}_{j+t}^{2\pi+t}, \quad (3)$$

where t and j are initial angle and angles around a point ($j = 0, 1, 2, \dots, 2\pi$), respectively.

Therefore, the least difference between geometric initialization and skin cancer image is concluded as the condition for initialization determination in this paper.

3. Proposed Initialization and Parameter Setting Method for Active Contours

In the implementation of the proposed method, some samples of skin cancer images need to be selected and trained to obtain candidate templates and their suitable parameters in the training state. As shown in the left column in the flowchart of Figure 4, the training state starts from inputting some known skin cancer images, performs fitting and voting for selection of candidate templates from geometrical shapes, and experiments those selected candidate templates with active-contour parameters for selecting

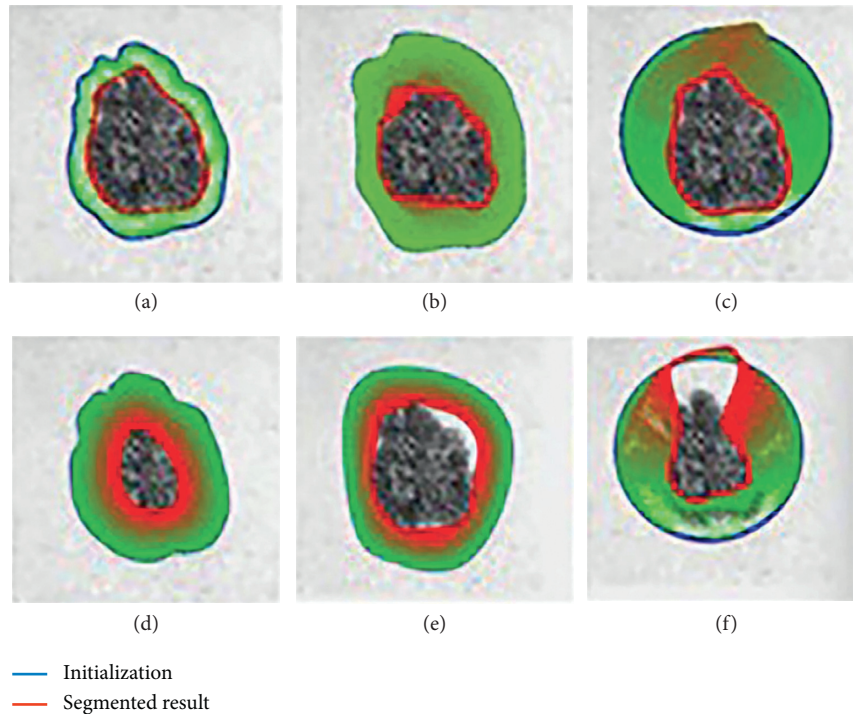


FIGURE 1: Comparison of different initialization and parameters. (a) Appropriate manual initialization with appropriate parameters. (b) Inappropriate manual initialization with appropriate parameters. (c) Geometrical initialization with appropriate parameters. (d) Appropriate manual initialization with inappropriate parameters. (e) Inappropriate manual initialization with inappropriate parameters. (f) Geometrical initialization with inappropriate parameters.

parameter set for each candidate template. Those mentioned candidate templates with their suitable active-contour parameter sets would be stored in a database for usage in the testing state. The template and parameter determination processes will be explained in Sections 3.1 and 3.2, respectively.

In the testing state, suppose an unknown skin cancer image is inputted for segmentation, a suitable template with its parameter set will then be selected by searching in the database of templates and parameter sets trained with geometrical shapes and active-contour parameters in advance, and an active contour will finally segment the skin cancer based on the selected geometrical shape template with parameter set, as shown in the right column of the flowchart in Figure 4. The processes of template determination and active contour segmentation are mentioned in Sections 3.1, and 3.3, respectively.

3.1. Geometrical Template Determination. Normally, the best initialization for active contour is considered to fix the shape closest to the workpiece. Since the workpiece boundary is originally unknown, a way to determine initialization automatically is to estimate a rough boundary of the workpiece and utilize a geometrical shape as a template, which is closest to the rough boundary, as initialization. In the estimation of the rough boundary of the workpiece, which is assumed to be a skin cancer boundary, some preprocessing such as binarization process can be used to simply extract a border of rough skin cancer boundary which is the foreground of

the image, and then offset should be added surrounding the extracted skin cancer boundary border to ensure the whole skin cancer boundary is picked up. In the utilization of geometrical shapes as candidate template for automatic initialization determination, all possible geometrical shapes, as shown in Figure 5, should be applied as a candidate to the rough skin cancer boundary with added offset, and all candidate geometrical shapes regarded as a template should be adjusted in scaling and rotation views to fit into the boundary. Conceptually, the centroid of a candidate geometrical template is first mapped in the same coordinates with the centroid of the rough skin cancer boundary, and differences between those mapped shapes are obtained in all scales and rotation angles of the template. In detail, the template first fixed its scale to be slightly bigger than the longest skeleton of the rough skin cancer boundary in the same centroid position, and then differences between the template and rough skin cancer boundary are obtained by rotating the candidate template in all angles, as shown in Figure 6. The algorithm of the mentioned steps is illustrated in Algorithm 1.

On the other hand, unknown skin cancer images performed segmentation in the testing state, as shown in the right column in Figure 6. In the testing state, all template candidates stored in the database of geometrical templates and parameter sets would be applied in a rough skin cancer boundary obtained by the binarization process. Like the processes in the training state, the candidate template with the least difference compared with the rough skin cancer boundary would be determined as the initial geometrical

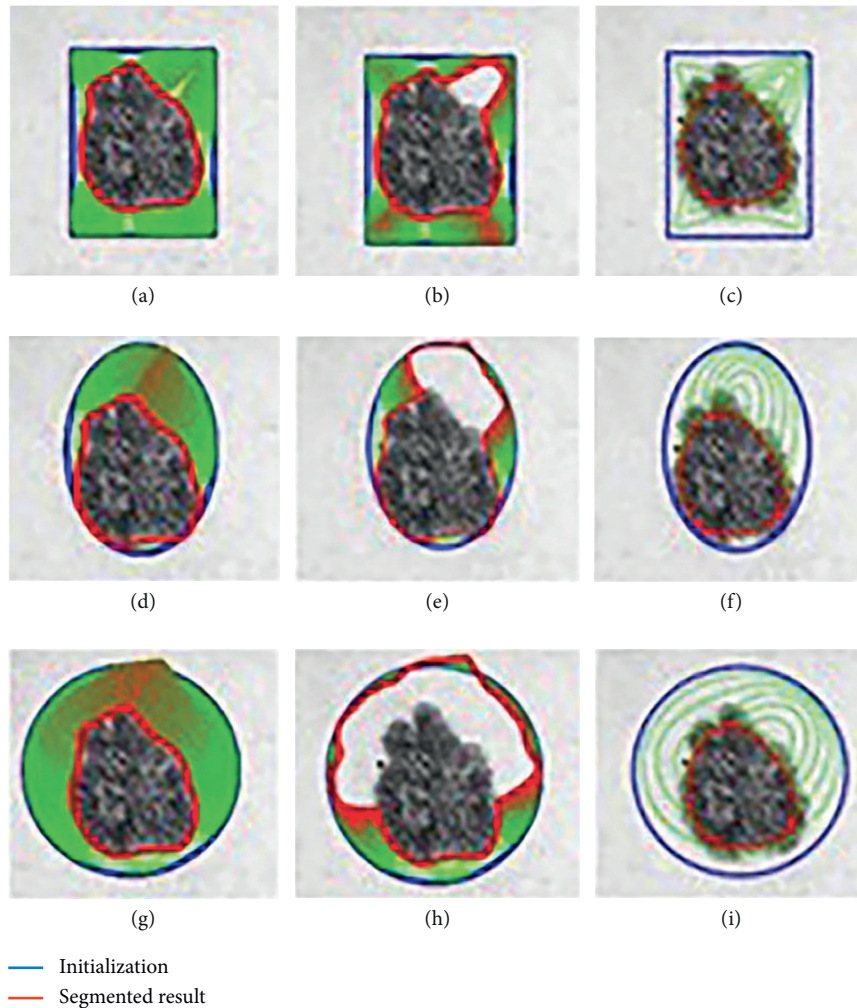


FIGURE 2: Comparison of snake initialized by geometric shapes with different parameters. (a), (d), and (g) Appropriate geometrical initialization with appropriate parameters. (b), (e), and (h) Appropriate geometrical initialization with too little convergence. (c), (f), and (i) Appropriate initialization with too much convergence.

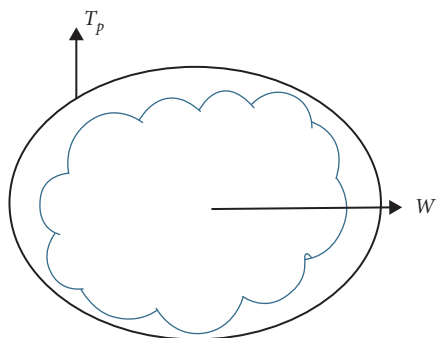


FIGURE 3: A geometric shape as initialization.

template for active contour operation, and the template and parameter set which has been trained and matched with the determined geometrical template are used in active contour segmentation in the next process.

In practice, it is almost impossible and redundant to find a suitable template from all existing geometrical shapes and find the difference between all possible templates with rough

skin cancer boundary by continuously varying the templates in all angles. As shown in Figure 7, circle, rectangle, and ellipse should be selected as candidate templates, while other kinds of geometrical shapes, which are rarely used, should be excluded from the template candidate group. These rarely used geometrical shapes should be considered to delete in the training state. Therefore, a practical way to reduce redundancy of template determination from all possible geometrical shapes is recommended to limit the number of geometrical template candidates by high possibility based on some samples of skin cancer image as training state in advance. To determine high possibility geometrical shapes as templates stored in the geometrical shape database, as shown in the left column in Figure 4, some samples of skin cancer boundary image are manually selected for finding a threshold value. Consequently, those selected skin cancer samples are performed before testing based on the steps illustrated in Algorithm 1 for preprocessing, template selection, and template rotation starting from the 3rd, 9th, and 15th lines, respectively. The number of determined geometrical templates would be simply counted, and a

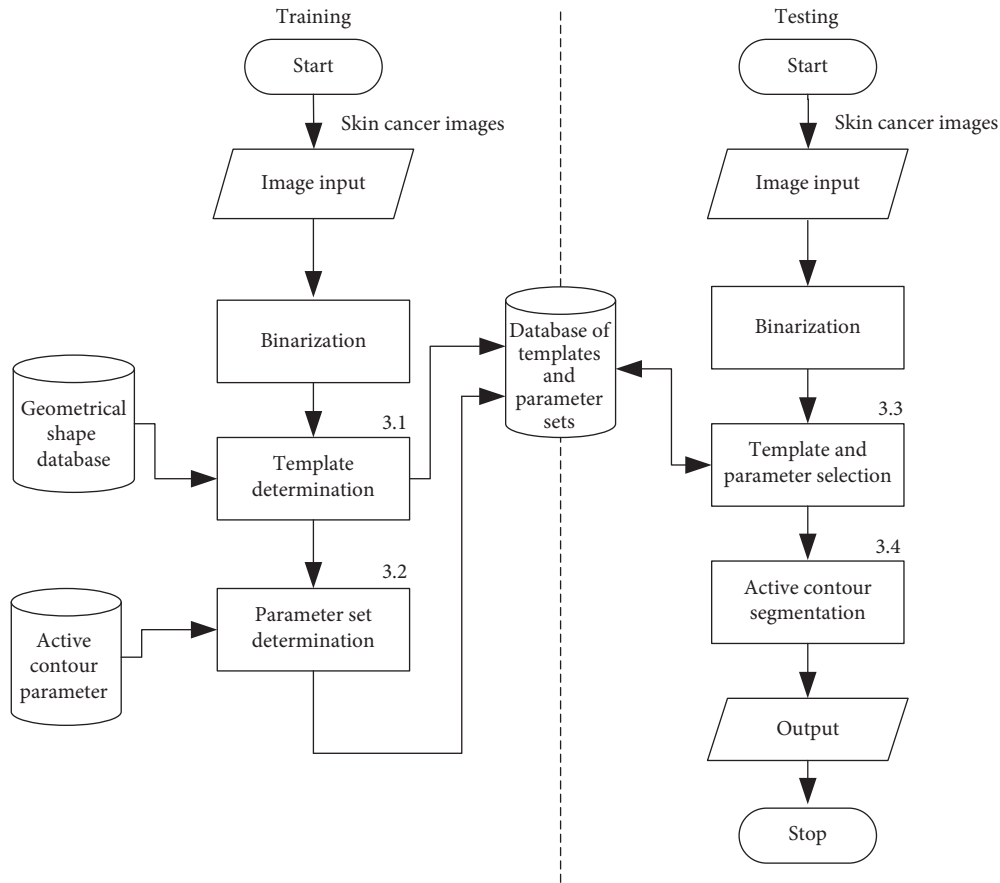


FIGURE 4: Flowchart of the proposed method. 3.1–3.4 represent subchapters explaining details of the processes.

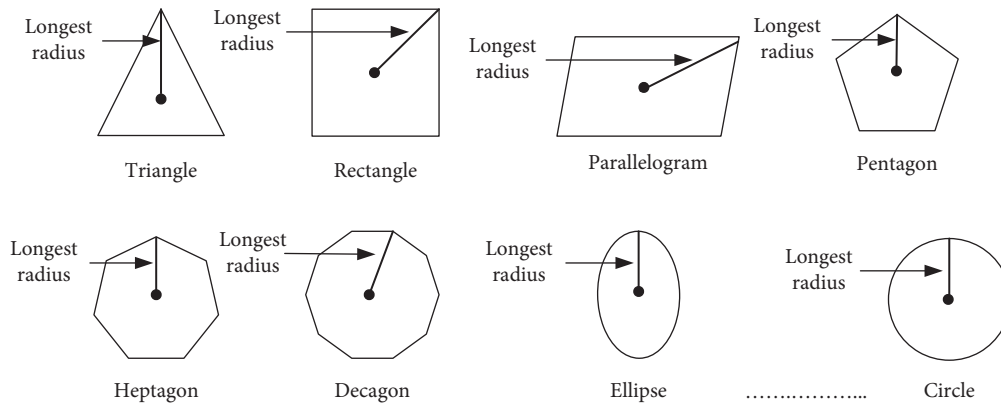


FIGURE 5: Geometrical shapes as template candidate.

thresholding value [18] should be statistically determined for dividing between highly selected and less selected geometrical shapes. This means some geometrical templates, which are frequently determined as initialization template, would be screened as template candidates for storing in the geometrical shape database.

3.2. Determination of Parameter Set for Candidate Templates. It has been proved that parameters which are matched with initialization will help active contour to segment

appropriately [19]. When geometrical shapes are selected as a template in the training state, parameters which are matched with those geometrical templates should be simultaneously performed before the test of active contour on some samples and determined in advance. These parameters which are the most matched in each geometrical template should be concluded as a parameter set of the geometrical template and would be utilized as initial parameters in the testing state.

In the training state, all parameters, which need to be set as initial ones, are listed up with varied ranges and steps, as

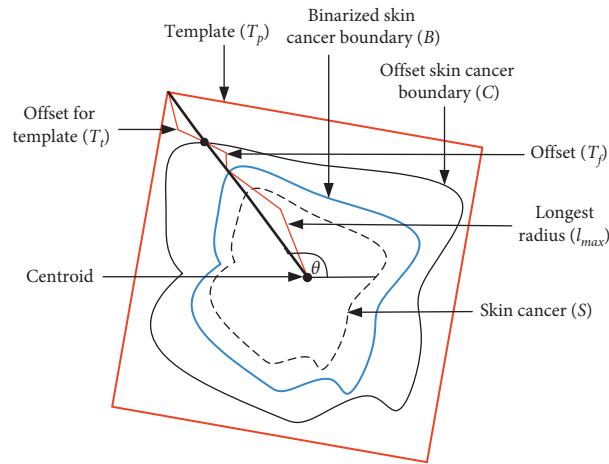


FIGURE 6: Mapping of geometrical shapes on a skin cancer boundary.

- (1) **START**
- (2) **SET** $T_p(i)$: candidate templates (e.g., circle, rectangle, ellipse, triangle, etc.), n : number of candidate templates, θ : template rotation angle, T_f : offset of rough skin cancer image
- (3) **REPEAT**
- (4) **INPUT** an image of skin cancer
- (5) **COMPUTE** binarization
- (6) **COMPUTE** rough contour of a skin cancer image
- (7) **COMPUTE** scaling the skin cancer image by offset (T_f)
- (8) **COMPUTE** centroid and the longest radius of skin cancer
- (9) **REPEAT**
- (10) **INPUT** candidate template ($T_p(i)$)
- (11) **COMPUTE** centroid and the longest radius of template
- (12) **COMPUTE** mapping between candidate template and skin cancer image based on centroid and the longest radius
- (13) **COMPUTE** scaling the template based on the longest radius of skin cancer image with offset
- (14) **SET** $\theta = 0$
- (15) **REPEAT**
- (16) **COMPUTE** difference between candidate template ($T_p(i)$) and skin cancer image
- (17) **CALCULATE** $\theta = \theta + 1$
- (18) **COMPUTE** rotation of candidate template on the centroid
- (19) **UNTIL** all rotation angles ($\theta \geq 360$)
- (20) **SELECT** angle with the smallest difference
- (21) **CALCULATE** $i = i + 1$
- (22) **UNTIL** all candidate templates
- (23) **SELECT** template and angle with the smallest difference with the skin cancer image
- (24) **UNTIL** all skin-cancer image samples
- (25) **END**

ALGORITHM 1: Template selection in training state.

shown by an example in Table 1. These ranges and steps of parameters would be used to vary in the pretest on some samples and concluded the best parameters by average as parameter set of the geometrical template. A computer program for finding parameter sets of all geometrical templates should be created following Algorithm 2. In Algorithm 2, a skin cancer image and geometrical templates are input in the 4th and 8th lines, respectively. Then, best parameter finding, parameter varying, and rotation are performed in 9th, 11th, and 13th, respectively.

In the training state, a geometrical template is performed mapping with some skin-cancer image samples while rotated

in all 360 degrees; the best parameter sets for all samples are selected based on the least differences with samples, and the average of those best parameters are determined as parameter set of the template as shown in Algorithm 2. The parameter sets of all templates are stored in the database as fixed-parameter sets of templates and ready to be retrieved and used for the operation of active contour in the testing state.

3.3. Template Selection and How to Apply a Template to a Skin Cancer Image. In the testing state, when a skin cancer image is input, a geometrical template and parameter set will be

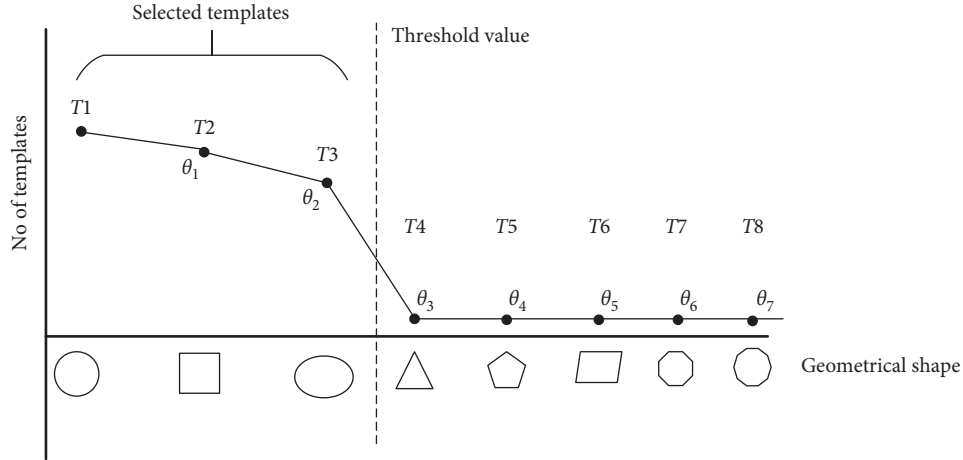


FIGURE 7: Determination of threshold value for screening geometrical shapes as template candidate.

TABLE 1: Parameters of active contour.

Parameter set	Range	Step
Interaction	100–400	100
nPoints	100–1000	100
Sigma1	1–10	1
Sigma2	1–20	5
Sigma3	0–1	1
Wline	0–0.01	0.01
Wedge	2–50	2
Wterm	0.01–100	10
Kappa	0–5	1
Alpha	0.02–7	0.2
Beta	0.01–2	0.01

selected, as shown on the right side of Figure 4. Like the training state, binarization is first performed to obtain a rough contour of the skin cancer image; then, an offset (T_f) is added to the rough contour of skin cancer for ensuring to cover the whole skin cancer, and centroid and the longest radius (l_{max}) of the workpiece are determined. Consequentially, mapping between the skin cancer image and candidate templates retrieved from the database is performed based on centroid and the longest radius and offset for template (T_i) is added to ensure the template will cover the workpiece. Differences between templates and workpiece are obtained while the candidate templates are rotated around the centroid in all angles, and the best candidate template and the best rotation angle with the smallest difference between the candidate template and offset skin cancer boundary (C) are selected as initialization and will be applied for active contour. Simultaneously, a parameter set, which was fixed in the training state and stored in the database, would be retrieved determined by template. The algorithm is illustrated in Algorithm 3. In Algorithm 3, the rough contour of a skin cancer image is obtained in the 4th line, and template mapping and selection are performed from the 7th line.

3.4. Operation of Active Contour. Currently, there exist many kinds of active contours such as normal gradient vector flow [20], convolution vector flow [21], dynamic directional

gradient vector flow [22], adaptive diffusion flow [23], and gradient vector flow (GVF) [24]. Users should consider selecting one of those active contours which is the most suitable for users' problems. The GVF has been selected in implementation and evaluation in this paper because it was designed and developed for the segmentation of complicate shapes and specially assumed to benefit varieties of medical applications.

It is well known that a couple of important keys for applying an active contour for image segmentation are initialization and parameters. In general, these are manually determined by experts based on the trial-and-error concept. The experts, who are going to tune up parameters to find the ones that are best matched with the images for segmentation, should well understand parameter characteristics for tuning suitably. These determined appropriate initialization and parameter set may contribute to energy minimization of active contour, especially for a local minimum.

This paper proposes initial geometrical templates for the automatic initialization of active contour. The geometrical template is first selected as the best matched with the skin cancer boundary image from a group of geometrical template candidates, and the selected template then performed mapping with the skin cancer boundary image in terms of scale and rotation. In addition, the parameter set, which was initially trained for each geometrical template, will be set with the initialization for active contour operation. To start running the operation of active contour, the mentioned geometrical template and parameter set are set in the program of active contour, as shown by an example in Table 1. When the active contour is executed, initialization will converge to the skin cancer image based on the energy minimization condition.

4. Experimental Results

This paper concentrates on the research problem of initial templates and parameter sets for initializing the active contour for skin cancer boundary segmentation and proposes a method for initial template and parameters setup. Since most of the medical image samples are sensitive and

```

(1) START
(2) SET  $T_p(j)$ : geometrical templates (e.g., circle, rectangle, ellipse, etc.), rng: range of parameter, rng_st [i]: starting range of parameter, stp [i]: step for range increment in a parameter, rng_end [i]: maximum range in parameters,  $n$ : number of templates,  $m$ : number of parameters, prm [j][i]: parameters for geometrical templates
(3) REPEAT
(4) INPUT a skin cancer image
(5) SET  $j = 1$ 
(6)   REPEAT
(7)     SET  $i = 1$ 
(8)     INPUT  $T_p(j)$ 
(9)     REPEAT
(10)      SET rng = rng_st [i]
(11)      REPEAT
(12)        CALCULATE rng = rng + stp [i]
(13)        REPEAT
(14)          COMPUTE mapping of template with a skin cancer image
(15)          COMPUTE difference between template and skin cancer image
(16)          UNTIL 360 degrees rotation
(17)          SELECT parameter for a skin-cancer-image sample by the smallest difference
(18)          CALCULATE (stp [i] = stp [i] × 2) and ( $i = i + 1$ )
(19)          UNTIL all ranges in a parameter (rng ≥ rng_end)
(20)          COMPUTE average of all selected parameters as a parameter in the set for a template (prm [j][i])
(21)          UNTIL all parameters ( $i = m$ )
(22)          CALCULATE  $j = j + 1$ 
(23)        UNTIL all templates
(24)     UNTIL all skin cancer images
(25) END

```

ALGORITHM 2: Parameter set determination for geometric templates.

```

(1) START
(2) SET  $T_p(i)$ : candidate templates (e.g., circle, rectangle, ellipse, triangle, etc.),  $n$ : number of candidate templates,  $\theta$ : template rotation angle,  $T_f$ : offset of rough skin cancer image
(3) INPUT an image of skin cancer
(4) COMPUTE binarization
(5) COMPUTE scaling the skin cancer image by offset ( $T_f$ )
(6) COMPUTE centroid and the longest radius of skin cancer
(7) REPEAT
(8)   INPUT candidate template ( $T_p(i)$ )
(9)   COMPUTE centroid and the longest radius of template
(10)  COMPUTE mapping between candidate template and skin cancer image based on centroid and the longest radius
(11)  COMPUTE scaling the template based on the longest radius of skin cancer image with offset
(12)  SET  $\theta = 0$ 
(13)  REPEAT
(14)    COMPUTE difference between candidate template ( $T_p(i)$ ) and skin cancer image
(15)    CALCULATE  $\theta = \theta + 1$ 
(16)    COMPUTE rotation of candidate template on the centroid
(17)    UNTIL all rotation angles ( $\theta \geq 360$ )
(18)    SELECT angle with the smallest difference
(19)    CALCULATE  $i = i + 1$ 
(20)  UNTIL all candidate templates
(21)  SELECT template and angle with the smallest difference with the skin cancer image
(22)  COMPUTE template mapping on the offset skin cancer image
(23)  INPUT parameter set of the selected template
(24)  COMPUTE active contour
(25) END

```

ALGORITHM 3: Template selection and applying the template to a skin cancer image.

confidential, it is normally difficult to find samples for experiments and evaluation. To verify the effectiveness and evaluate the performance of the proposed method, the GVF snake algorithm was selected as active contour, and 227 images of skin cancer, as shown in the 2nd column of Table 2, are used in experiments performed by the experiment set up, as shown in Table 3. Initial template candidates including circle, ellipse, and rectangle, which were automatically selected by our proposed method in terms of similarity with the skin cancer images, were used as candidate templates of initialization, and results of template selection and mapping are shown in the 3rd column of Table 2. Initially, parameters for those templates for the GVF snake algorithm have been trained with 48 skin cancer image samples, and parameter sets for selected geometrical shape templates were determined based on the proposed method and concluded in Table 4. Experimental results done based on our proposed method and experts are shown in the 4th and 5th columns of Table 2, respectively. Finally, errors which are differences between the results of our proposed method and experts are shown in the 6th column. Evaluation of our proposed method using geometrical shapes with parameter sets for automatic initialization can be concluded as 99.46% accuracy approximately with 13.61 sec per skin cancer image as computational time, while conventional methods [13, 14] achieved 96.41% and 99.20%, respectively, as shown in Table 5. On the other hand, sensitivity and specificity of conventional methods [13, 14] can be calculated [18] and concluded as 85.13% and 97.07%, and 99.83% and 99.16%, respectively, while the proposed method shows 97.43% and 99.87%, respectively, as shown in Table 5. Sensitivity (SEN), specificity (SPC), accuracy (ACC), and the Jaccard Index (JAC) can be calculated as follows [25].

Sensitivity:

$$SEN = \frac{TP}{TP + FN} \quad (4)$$

Specificity:

$$SPC = \frac{TN}{TN + FP} \quad (5)$$

Accuracy:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Jaccard index:

$$JAC = \frac{TP}{(TP + FP + FN)} \quad (7)$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively.

5. Discussion

This paper proposes initial geometrical templates and parameter sets as automatic initialization for the active contour on skin cancer boundary segmentation. The initial templates, which are geometrical shapes, such as circle,

rectangle, ellipse, pentagon, are initially trained with some skin cancer image samples for template candidate determination. The parameter sets are accordingly established by varying all parameters and selecting the best parameter groups for each template.

In testing, candidate geometric shapes are compared with the input skin cancer image. The initial template is then determined by the least difference with the candidate initial templates, and the parameter set is used to initialize and run the active contour for the skin cancer boundary segmentation. Since normally medical image samples are not open in public due to personal information, it is difficult to collect a lot of samples. Moreover, the samples for evaluation based on conventional methods are not open in public, it is impossible to directly compare performance on the same samples. Therefore, the authors of this paper attempted to collect skin cancer image samples in a number, in which evaluation results can be relied on, to compare with the conventional methods. Experiments were performed based on our proposed method, and the performance of the proposed method achieves as high as 99.46% accuracy approximately, which is similar and slightly better than the conventional method [14]. However, in the case of skin cancer surgery, segmentation would originally be performed based on expert skill with added offset, and specificity, which represents false positive, should be importantly concerned. Our proposed method improves 0.26% specificity compared with conventional methods so that it is regarded to be useful for skin cancer segmentation. On the other hand, although modern and robust deep learning tool as CNN is currently considered as an excellent and powerful tool for classification and should be effectively applied in the skin cancer segmentation problem, it has been reported that active contour can perform comparably with CNN with less computation time [26].

Furthermore, since our basic concept is to approximate a similar geometrical shape as the initial template applied in the initialization of active contour, our initialization is originally apart from the skin cancer boundary. Mathematically, any shaped template which is close to the skin cancer boundary is the most suitable, but the algorithm may be complicate and time-consuming. The idea of using geometrical shape as a template for initialization is considered as a solution for implementation. As seen in the experimental results in Table 2, there exist obvious errors in skin cancer images, T4, T13, T18, T19, T39, T51, T59, T67, T88, T168, T169, T179, T185, T186, and T197. If template candidates are increased, some errors are partially solved. In fact, if a template is appropriately selected during the initialization, the snake algorithm should perform well. Basically, the concept of template implicitly includes comparing some errors with an initialized shape performed by experts. These errors should be traded off with computational cost. To improve initialization for active contour, learning and observing initialization done by skillful experts is recommended as future work.

For skin cancer boundary segmentation, our proposed method achieves 99.46% accuracy which includes both concave and convex results. As observed, most of the results

TABLE 2: Experimental results of skin cancer segmentation.

No.	Input image	Selected template	Proposed method	Result by experts
T1				
T2				
T3				
T4				
T5				
T6				
T7				
T8				
T9				
T10				
T11				
T12				
T13				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T14				
T15				
T16				
T17				
T18				
T19				
T20				
T21				
T22				
T23				
T24				
T25				
T26				

TABLE 2: Continued.

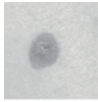
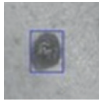
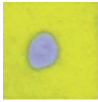
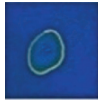
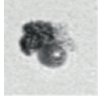
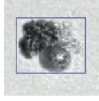

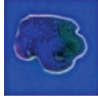
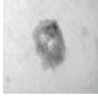

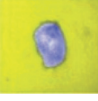

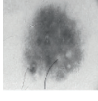
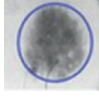
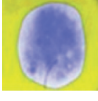
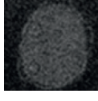
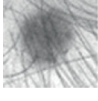

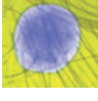
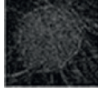



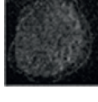
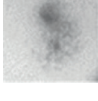

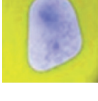
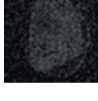


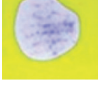
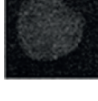


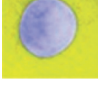
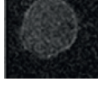



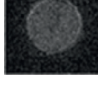
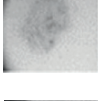


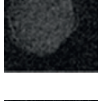
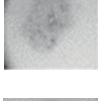

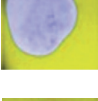

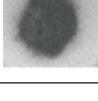

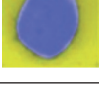
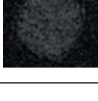
No.	Input image	Selected template	Proposed method	Result by experts
T27				
T28				
T29				
T30				
T31				
T32				
T33				
T34				
T35				
T36				
T37				
T38				
T39				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T40				
T41				
T42				
T43				
T44				
T45				
T46				
T47				
T48				
T49				
T50				
T51				
T52				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T53				
T54				
T55				
T56				
T57				
T58				
T59				
T60				
T61				
T62				
T63				
T64				
T65				

TABLE 2: Continued.

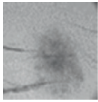
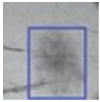
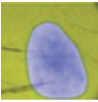
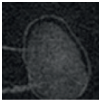
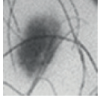
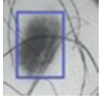
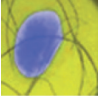
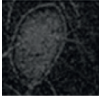
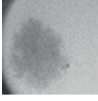
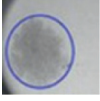
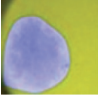
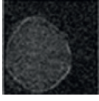
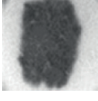
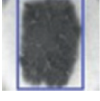


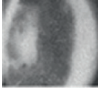
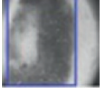





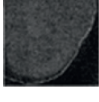
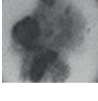

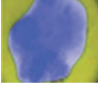
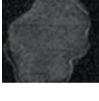
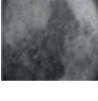
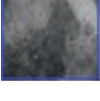


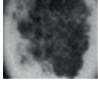

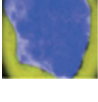
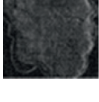
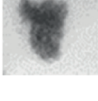







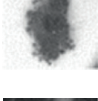

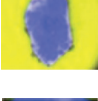
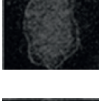


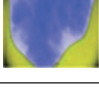
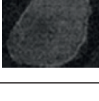
No.	Input image	Selected template	Proposed method	Result by experts
T66				
T67				
T68				
T69				
T70				
T71				
T72				
T73				
T74				
T75				
T76				
T77				
T78				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T79				
T80				
T81				
T82				
T83				
T84				
T85				
T86				
T87				
T88				
T89				
T90				
T91				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T92				
T93				
T94				
T95				
T96				
T97				
T98				
T99				
T100				
T101				
T102				
T103				
T104				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T105				
T106				
T107				
T108				
T109				
T110				
T111				
T112				
T113				
T114				
T115				
T116				
T117				

TABLE 2: Continued.

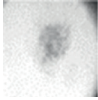
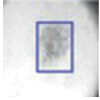
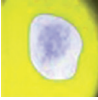
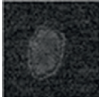

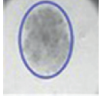



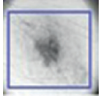







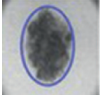


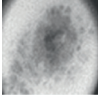
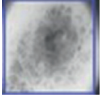
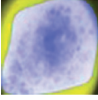
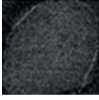

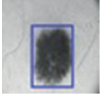



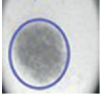


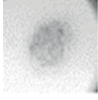
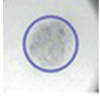
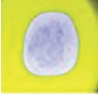
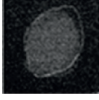

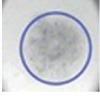









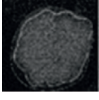




No.	Input image	Selected template	Proposed method	Result by experts
T118				
T119				
T120				
T121				
T122				
T123				
T124				
T125				
T126				
T127				
T128				
T129				
T130				

TABLE 2: Continued.

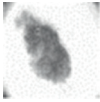


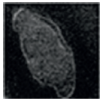
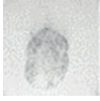


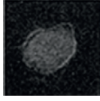
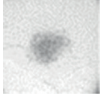
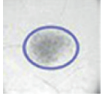

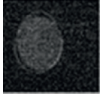
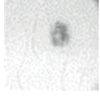
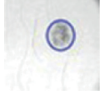

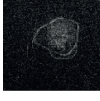
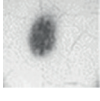
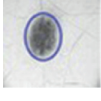

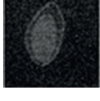
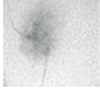
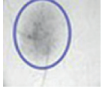

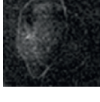
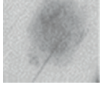
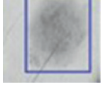
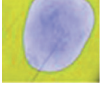
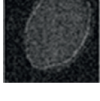
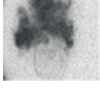
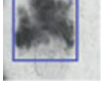

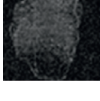
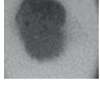

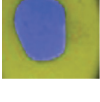
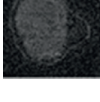
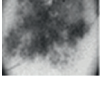


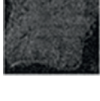
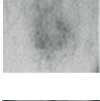



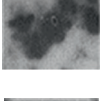
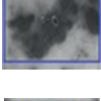


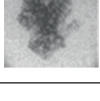
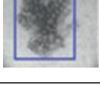
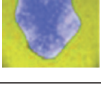

No.	Input image	Selected template	Proposed method	Result by experts
T131				
T132				
T133				
T134				
T135				
T136				
T137				
T138				
T139				
T140				
T141				
T142				
T143				

TABLE 2: Continued.

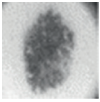
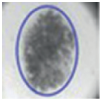
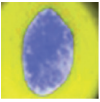
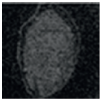
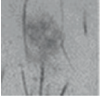
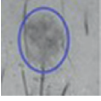


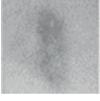
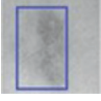

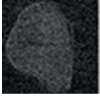
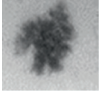

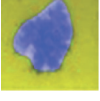
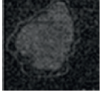
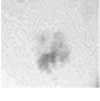


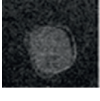



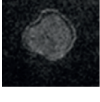



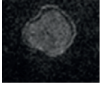


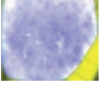


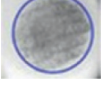
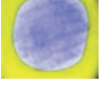
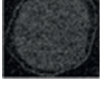
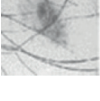
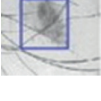
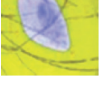
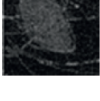
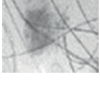
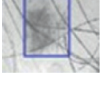
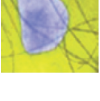
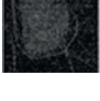
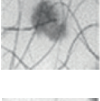
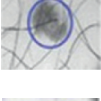
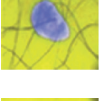
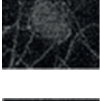


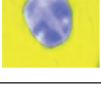
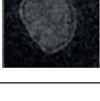
No.	Input image	Selected template	Proposed method	Result by experts
T144				
T145				
T146				
T147				
T148				
T149				
T150				
T151				
T152				
T153				
T154				
T155				
T156				

TABLE 2: Continued.

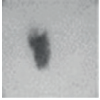
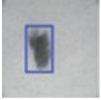

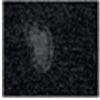
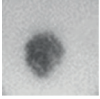
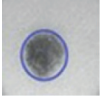
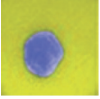
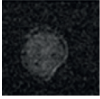
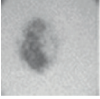
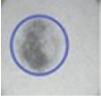
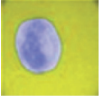
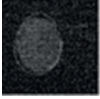
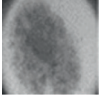
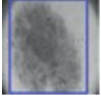
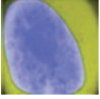
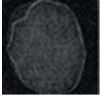
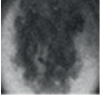
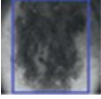
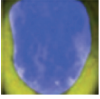
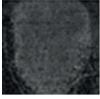
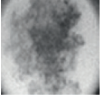


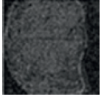
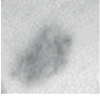


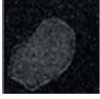
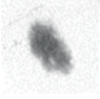


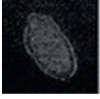
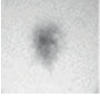
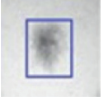

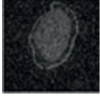
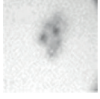
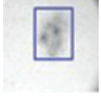

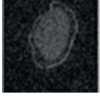
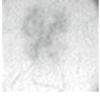
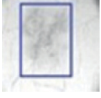

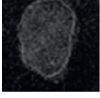



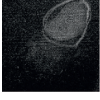
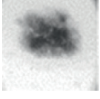

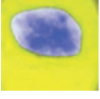
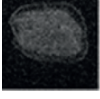
No.	Input image	Selected template	Proposed method	Result by experts
T157				
T158				
T159				
T160				
T161				
T162				
T163				
T164				
T165				
T166				
T167				
T168				
T169				

TABLE 2: Continued.

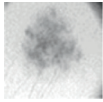


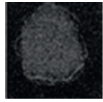
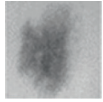
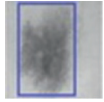

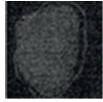
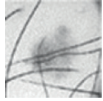
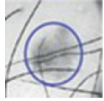
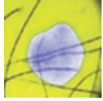
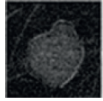
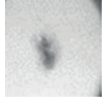
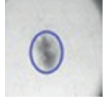
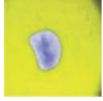
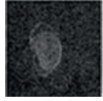



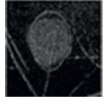
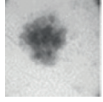
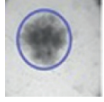
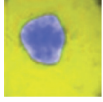
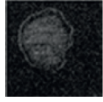
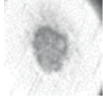
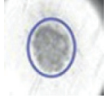

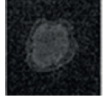
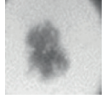

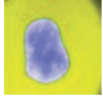
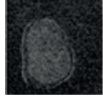
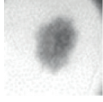


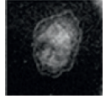
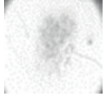


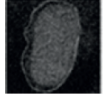
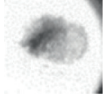

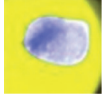
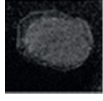
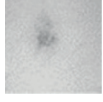


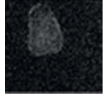
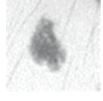
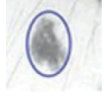
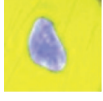
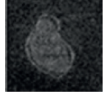
No.	Input image	Selected template	Proposed method	Result by experts
T170				
T171				
T172				
T173				
T174				
T175				
T176				
T177				
T178				
T179				
T180				
T181				
T182				

TABLE 2: Continued.




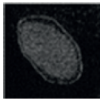
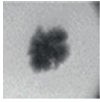
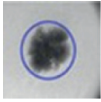
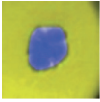
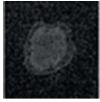
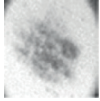
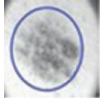
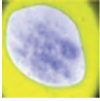
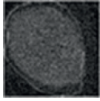

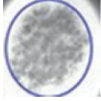
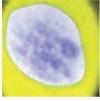
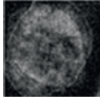



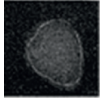



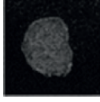
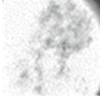
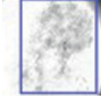

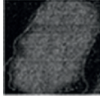
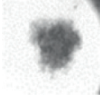


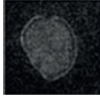
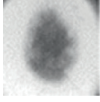
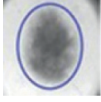
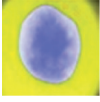
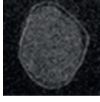
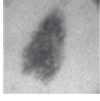
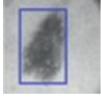
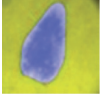
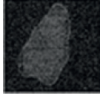
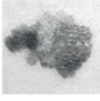
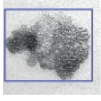
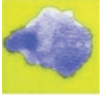
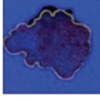
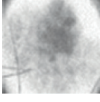
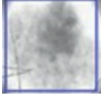

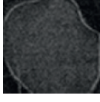
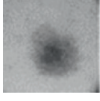
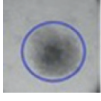
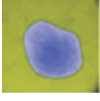
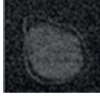
No.	Input image	Selected template	Proposed method	Result by experts
T183				
T184				
T185				
T186				
T187				
T188				
T189				
T190				
T191				
T192				
T193				
T194				
T195				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T196				
T197				
T198				
T199				
T200				
T201				
T202				
T203				
T204				
T205				
T206				
T207				
T208				

TABLE 2: Continued.

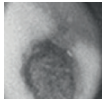
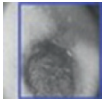

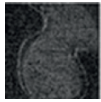
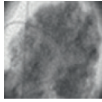
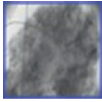

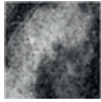
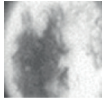
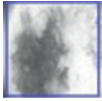

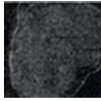
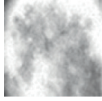


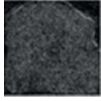

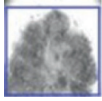

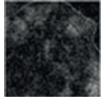
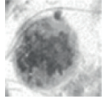
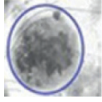
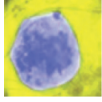
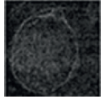
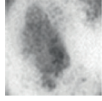
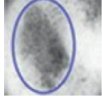

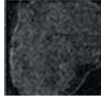
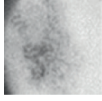
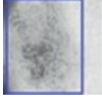
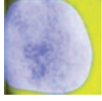
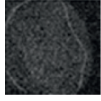
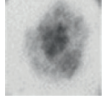
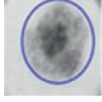
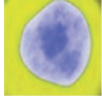
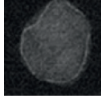
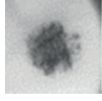
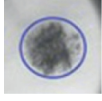
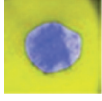
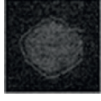
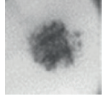
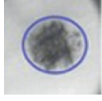
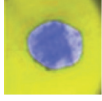
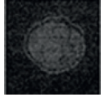
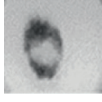

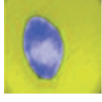

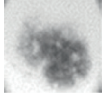
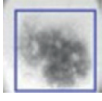

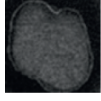
No.	Input image	Selected template	Proposed method	Result by experts
T209				
T210				
T211				
T212				
T213				
T214				
T215				
T216				
T217				
T218				
T219				
T220				
T221				

TABLE 2: Continued.

No.	Input image	Selected template	Proposed method	Result by experts
T222				
T223				
T224				
T225				
T226				
T227				

TABLE 3: Specification of experiment set up.

Image	Size	256 × 256 pixels
	Gray level	256
	No. of cancer boundaries per image	1
Program	Basic programming	MATLAB_R2018b
	Segmentation	Snake internal force matrix 2D, snake internal force matrix 3D, snake move iteration 2D, snake move iteration 3D, snake 2D, snake 3D
	Computer	Device name: LAPTOP-ENCF7NAD Processor: Intel(R) Core(TM) i5-1035G1 CPU@1.00 GHz 1.19 GHz Installed RAM: 8.00 GB (7.78 GB usable) Product ID: 00327-35165-87873-AAOEM System type: 64-bit operating system, x64-based processor

TABLE 4: A sample of parameter sets selected by the proposed method.

Parameter	Interaction	nPoints	Sigma1	Sigma2	Sigma3	Wline	Wedge	Wterm	Kappa	Alpha	Beta
<i>Parameter set for circular template</i>											
	1000	800	5	5	0	0	10	50	2	3	0.001
<i>Parameter set for the ellipse template</i>											
	1000	100	5	5	0	0	10	100	2	2	0.01
<i>Parameter set for the rectangular template</i>											
	1000	1000	5	5	0	0	10	150	2	3	0.01

TABLE 5: Comparison between conventional methods and the proposed method.

Model	Average accuracy (%)			
	Region-based measurement			
	JAC	SEN	SPE	ACC
Model ESM [13]	—	85.13	99.83	96.41
Model ESM/AC [14]	—	97.07	99.61	99.20
Proposed method	97.35	97.43	99.87	99.46

are convex. However, some are partially included in concave, and most of them are tiny. In medical practice, miss-segmentation in convex is better than concave because all skin cancer cells are removed. The concave segmented boundaries are recovered by offset as a final boundary.

6. Conclusion

Automatic skin cancer image segmentation using active contour, which originally depended on skillful initialization done by experts, required practical automatic initialization with appropriate parameter sets. This paper proposed a method of initial geometrical templates with parameter sets for active contour on skin cancer boundary segmentation. Some skin cancer images were initially trained to evaluate and select geometrical shapes as candidate templates by mapping the geometrical shapes with skin cancer image samples based on the centroid and the longest radius and finding shapes with the least differences as candidate templates. These candidate templates were then used to perform active contour by possible parameters and then find a parameter set which performed the best active contour segmentation as a parameter set for each candidate template. In testing, these candidate templates performed mapping with a skin cancer image and rotating around the centroid in the same manner with training, and a geometrical template with the least difference with the skin cancer image was determined as the initial template for active contour. The determined geometrical template with the trained parameter set then was initialized for the active contour segment. Finally, the effectiveness of the proposed method has been evaluated by experiments with 227 skin cancer images.

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

The authors are thankful to Ph.D. scholarship supported by Rajamangala University of Technology Isan (RMUTI), Surin Campus.

References

- [1] <http://www.who.int/>.
- [2] <https://www.cancer.org/cancer/basal-and-squamous-cell-skin-cancer.html>.
- [3] N. Banjan, P. Dalvi, and N. Athavale, "Melanoma skin cancer detection by segmentation and feature extraction using combination of OTSU and STOLZ algorithm technique," *SSRG International Journal of Electronics and Communication Engineering (SSRG-IJECE)*, vol. 4, no. 4, pp. 21–25, 2017.
- [4] P. G. Cavalcanti, J. Scharcanski, C. E. Martinez, and L. E. Di Persia, "Segmentation of pigmented skin lesions using non-negative matrix factorization," in *Proceedings of the 2014 IEEE Instrumentation and Measurement Technology Conference*, pp. 72–75, Montevideo, Uruguay, 2014.
- [5] O. S. Murumkar and P. Gumaste, "Feature extraction for skin cancer lesion detection," *International Journal of Scientific & Engineering and Technology Research*, vol. 4, no. 8, pp. 1604–1606, 2015.
- [6] A. Barrett, J. Dobbs, S. Morris, and T. Roques, *Practical Radiotherapy Planning*, Hodder Arnold, London, UK, 2009.
- [7] <https://www.cancer.gov/types/skin>.
- [8] H. Castillejos, V. Ponomaryov, L. Nion-de-Rivera, and V. Golikov, "Wavelet transform fuzzy algorithms for dermoscopic image segmentation," *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 578721, 11 pages, 2012.
- [9] J. Kim, M. Cetin, and A. S. Willsky, "Nonparametric shape priors for active contour-based image segmentation," *Signal Processing*, vol. 87, no. 12, pp. 3021–3044, 2007.
- [10] V. Jeya Ramya, J. Navarajan, R. Prathipa, and L. Ashok Kumar, "Detection of melanoma skin cancer using digital camera images," *Journal of Engineering and Applied Sciences*, vol. 10, no. 7, pp. 3082–3085, 2015.
- [11] A. Munir, S. Soomro, C. H. Lee, and K. N. Choi, "Adaptive active contours based on variable kernel with constant initialisation," *IET Image Processing*, vol. 12, no. 7, pp. 1117–1123, 2018.
- [12] J. K. Mogali, A. K. Pediredla, and C. S. Seelamantula, "Template-based active contours," in *Proceedings of the 2013 Computer Vision and Pattern Recognition*, vol. 1, pp. 1–14, Portland, OR, USA, 2013.
- [13] K. Kirimasthong, A. Rodtook, W. Lohitvisate, and S. S. Makhanov, "Automatic initialization of active contours in ultrasound images of breast cancer," *Pattern Analysis and Applications*, vol. 21, no. 2, pp. 491–500, 2018.
- [14] A. Rodtook, K. Kirimasthong, W. Lohitvisate, and S. S. Makhanov, "Automatic initialization of active contours and level set method in ultrasound images of breast abnormalities," *Pattern Recognition*, vol. 79, pp. 172–182, 2018.
- [15] P. Bumrungkun, W. Patchoo, and K. Chamnongthai, "Detection skin cancer using SVM and snake model," in *Proceedings of the 2018 International Workshop on Advanced Image Technology*, Chiang Mai, Thailand, 2018.

- [16] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Cengage Learning, Boston, MA, USA, 2007.
- [17] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [18] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Openstax Rice University, Houston, TX, USA, 2004.
- [19] J. Chopina, S. J. Miklavcic, and H. Lagaa, "Selection of parameters in active contours for the phenotypic analysis of plants," in *Proceedings of the 20th International Congress on Modelling and Simulation*, pp. 510–516, Adelaide, Australia, 2013.
- [20] J. Ning, C. Wu, S. Liu, and S. Yang, "NGVF: an improved external force field for active contour model," *Pattern Recognition Letters*, vol. 28, no. 1, pp. 58–63, 2007.
- [21] B. Li and S. T. Acton, "Active contour external force using vector field convolution for image segmentation," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2096–2106, 2007.
- [22] J. Cheng and S. W. Foo, "Dynamic directional gradient vector flow for snakes," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1563–1571, 2006.
- [23] Y. Wu, Y. Wang, and Y. Jia, "Adaptive diffusion flow active contours for image segmentation," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1421–1435, 2013.
- [24] C. Xu and J. Prince, "Gradient vector flow: a new external force for snakes," in *Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 66–71, San Juan, PR, USA, 1997.
- [25] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BioMed central Medical Imaging*, vol. 15, no. 29, pp. 1–28, 2015.
- [26] X. Zhu, Y. Wei, Y. Lu, M. Zhao et al., "Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging," *Computer Methods and Programs Biomedicine*, vol. 199, Article ID 105914, 2021.

Research Article

Multiple Sclerosis Lesion Segmentation in Brain MRI Using Inception Modules Embedded in a Convolutional Neural Network

Shahab U. Ansari ¹, Kamran Javed ^{1,2}, Saeed Mian Qaisar ^{3,4}, Rashad Jillani ¹
and Usman Haider ¹

¹Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan

²National Centre of Artificial Intelligence (NCAI), Saudi Data and Artificial Intelligence Authority (SDAIA), Riyadh, Saudi Arabia

³Electrical and Computer Engineering Department, Effat University, Jeddah 22332, Saudi Arabia

⁴Communication and Signal Processing Lab, Energy and Technology Research Center, Effat University, Jeddah 22332, Saudi Arabia

Correspondence should be addressed to Saeed Mian Qaisar; sqaisar@effatuniversity.edu.sa

Received 11 June 2021; Accepted 27 July 2021; Published 4 August 2021

Academic Editor: G R Sinha

Copyright © 2021 Shahab U. Ansari 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.

Multiple sclerosis (MS) is a chronic and autoimmune disease that forms lesions in the central nervous system. Quantitative analysis of these lesions has proved to be very useful in clinical trials for therapies and assessing disease prognosis. However, the efficacy of these quantitative analyses greatly depends on how accurately the MS lesions have been identified and segmented in brain MRI. This is usually carried out by radiologists who label 3D MR images slice by slice using commonly available segmentation tools. However, such manual practices are time consuming and error prone. To circumvent this problem, several automatic segmentation techniques have been investigated in recent years. In this paper, we propose a new framework for automatic brain lesion segmentation that employs a novel convolutional neural network (CNN) architecture. In order to segment lesions of different sizes, we have to pick a specific filter or size 3×3 or 5×5 . Sometimes, it is hard to decide which filter will work better to get the best results. Google Net has solved this problem by introducing an inception module. An inception module uses 3×3 , 5×5 , 1×1 and max pooling filters in parallel fashion. Results show that incorporating inception modules in a CNN has improved the performance of the network in the segmentation of MS lesions. We compared the results of the proposed CNN architecture for two loss functions: binary cross entropy (BCE) and structural similarity index measure (SSIM) using the publicly available ISBI-2015 challenge dataset. A score of 93.81 which is higher than the human rater with BCE loss function is achieved.

1. Introduction

Multiple sclerosis (MS) is a chronic disease that damages the nerves in the spinal cord, brain, and optic nerves. Axons in the brain are covered with a myelin sheath. Demyelination is a process in which the myelin sheaths start falling off and develops lesions in brain nerves. Millions of people are affected by MS worldwide which is mainly found in young people between 20 and 50 years of age. The symptoms caused by this disease are fatigue, memory problem, the problem in concentration, weakness, loss of balance, loss of vision, and many others. Diagnosing and treating this disease is very challenging because of its variability in its clinical expression.

These lesions can be traced in magnetic resonance imaging (MRI) using different sequences. Many features such as a volume and location are very important biomarkers for tracking the progression of the disease. Manually segmenting these lesions by expert radiologists is the most common practice in clinics, but this is tiresome, time consuming, and error prone. Figure 1 shows the manual segmentation of MS lesions by two raters in one slice of a brain MRI.

In recent years, automatic segmentation of MS lesions using convolutional neural networks (CNNs) have been investigated [1–5]. CNNs learn subtle features from the raw image data to facilitate 2D pixel (or 3D voxel) classification that ultimately leads to image segmentation. However, there

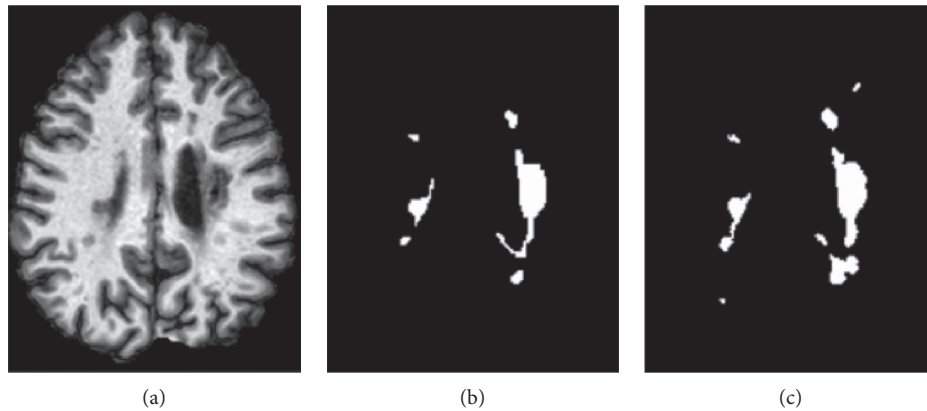


FIGURE 1: Manual segmentation of MS lesions: (a) T1w MRI, (b) manual segmentation by rater 1, and (c) manual segmentation by rater 2.

is no one-fit-for-all CNN model that could work for every classification problem or data. An expert knowledge has to be incorporated during the design phase of the CNN model based on the nature of the application and the data. Complex problems such as MS lesion segmentation require careful selection of the CNN architecture and training model for an optimum solution. In addition, automatic segmentation of an MS lesion in MRI may be challenging due to the following:

- (i) The lesion size and location are highly variable
- (ii) The edges between anatomical objects are not well defined in MR images due to low contrast
- (iii) The MR image of clinical quality may have imaging artifacts such as noise and inhomogeneity

In this work, we are proposing a novel CNN architecture for MS lesion segmentation. The MS lesions vary tremendously in size and shape, and sometime, it is difficult to detect in brain MR images. To address this particular challenge, inception modules, originally introduced by Google in GoogLeNet, are added in the CNN model [6]. The significance of the inception module lies in using multiple kernels of different sizes in parallel in an efficient way. This smart approach captures features of varying magnitude in the input data without overburdening the network with additional computations. The proposed model is trained for two loss functions, binary cross entropy (BCE) and structural similarity index measure (SSIM). The BCE loss function tries to maximize the difference of the probability distribution between two classes, in this case, lesion and nonlesion voxels [7]. SSIM, on the other hand, is a perception-based loss function that quantifies the similarity between two images [8].

The proposed solution for the MS lesion segmentation in brain MRI offers the following attributes:

- (i) Introduction of inception modules embedded in the CNN architecture for the segmentation of MS lesions with different shapes and sizes
- (ii) Comparison of MS lesion segmentation results using BCE and SSIM loss functions
- (iii) Improvement of performance of the proposed architecture in terms of the Dice coefficient, positive

predicted value, lesion-wise true positive rate, and volume difference of the segmented lesions compared to the gold standard

1.1. Literature Review. In past decade, deep neural networks have shown promising results in the segmentation of MS lesions in brain MR images. In [9], a novel architecture for segmenting MS lesions in magnetic resonance images by using a deep 3D convolutional encoder with the connections of shortcut in pathways was proposed. The method was evaluated on publicly available data from ISBI-2015 [10] and MICCAI-2008 [11] challenges. Authors compared their method with other five available approaches used for MS lesion segmentation. The final results show that their method outperformed the previous existing methods for MS lesion segmentation. In [12], the authors used a fully automatic multiview CNN approach for segmenting a multiple sclerosis lesion in longitudinal MRI data and tested on the ISBI-2015 dataset. Various deep learning techniques for the medical image analysis are presented in [13].

Valverde et al. have proposed a novel architecture for segmentation of a white matter (WM) lesion in multiple sclerosis (MS) using small number of imaging data [14]. This approach proposed a cascaded CNN model working on 3D MRI patches from FLAIR and T1w modalities. In this method, the output of the first network is retrained on the second network in series to reduce misclassification from the first network. The proposed model score is evaluated on the publicly available dataset of MICCAI-2008 and outperformed all the participant approaches. Roy et al. proposed a fully convolutional neural network (FCNN) to segment WM lesions in multicontrast MR images using multiple convolutional pathways [15]. The first pathway of the CNN contains dual convolutional filters for two image modalities. In the second pathway, the convolutional filters are applied to the output of the first pathway which are in parallel and concatenated. This method was evaluated on the ISBI-2015 dataset. A novel approach of using a fully 2D CNN to segment MS lesions in MR images is proposed in [16]. Maleki et al. have investigated the use of a CNN

model for the detection and segmentation of MS lesions [17].

In recent studies, a multimodal MRI dataset in tissue segmentation has shown promising results. In a recent work for brain tumor segmentation, a deep multitask learning framework that performs a performance test on multiple BraTS datasets was shown [18]. The authors claimed improvement over the traditional V-Net framework by using a structure of two parallel decoder branches. The original decoder performs segmentation, and the newly added decoder performs the auxiliary task of distance estimation to make more accurate segmentation boundary. A total loss function is introduced to combine the two tasks with a gamma factor to reduce the focus on the background area and set different weights for each type of label to alleviate the problem of category imbalance. Zhang et al. proposed the ME-Net model and obtained promising results using the BraTS 2020 dataset [19]. Four encoder structures for the four modal images of brain tumor MRI were employed with skip-connections. The combined feature map was given as input to the decoder. The authors also introduced a new loss function, that is, Categorical Dice, and set different weights for different masks. In another study, a 3D supervoxel-based learning method was proposed that demonstrated promising results in the segmentation of brain tumor [20]. The added features from multimodal MRI images greatly increased the segmentation accuracy. In another earlier study, Gabor texon feature, fractal analysis, curvature, and statistical intensity features from superpixels were used to segment tumors in multimodal brain MR images using extremely randomized trees (ERTs) [21]. The experimental results demonstrated the high detection and segmentation performance of the proposed method. Soltaninejad et al. proposed a method that used machine-learned features learned by fully convolutional networks (FCNs) and texon-based histograms as hand-crafted features [22]. The random forest (RF) classifier was then employed for the automated segmentation of brain tumor in the BraTS 2017 dataset.

Segmentation results can be greatly affected by the quality of the MRI images. Low resolution, intensity variations, and image acquisition noise hamper the accuracy of a segmentation task. Jin et al. proposed a deep framework for the segmentation of prostate cancer [23]. They had shown that the segmentation results were greatly improved by using bicubic interpolation and improved version of 3D V-Net. The bicubic interpolation of the input data helped in enhancing the relevant features required for prostate segmentation. Recently, attention-based methods have gained reputation in the segmentation of small but discrete objects in MRI images. In a study, for the enhancement of left atrium scars, a dilated attention network was used [24]. The proposed approach improved the accuracy of the scar segmentation to 87%. Liu et al. proposed a spatial attentive Bayesian deep learning network for the automatic segmentation of the peripheral zone and transition zone of the prostate with uncertainty estimation [25]. This method outperformed the state-of-the-art methods.

The heterogeneity of MS lesions poses a challenge for the detection and segmentation in MR images. An

attention-based fully CNN has also been used in the segmentation of prostate zones [26]. The authors in this work have proposed a novel feature pyramid attention mechanism to cope with heterogeneous prostate anatomy. Raschke et al. developed a statistical method to analyze heterogeneity of brain tumors in multimodal MRI [27]. The approach presented in the paper does not make any assumption on the probability distribution of the MRI data and prior knowledge of the location of tumors. This, according to the authors, gives an advantage for tumor segmentation of varying sizes and spatial locations. The proposed method consist of two deep subnetworks in which the first one was an encoding network that was responsible of extracting feature maps and the second was a decoding network and was responsible for upsampling feature maps. The proposed FCNN was evaluated on an ISBI-2015 dataset.

2. Proposed Methodology

As mentioned earlier, the shape and size of MS lesions vary dramatically. To detect these lesions using machine learning techniques is a challenging task. In the proposed methodology, a CNN model with inception modules is employed to automatically segment MS lesions in brain MRI. Filters of multiple sizes used in the inception modules capture features of MS lesions of different sizes. Prior to CNN model training, the images in the dataset are first preprocessed to remove image noise, intensity inhomogeneity, variability of intensity ranges, and the presence of nonbrain tissues. In this work, preprocessed ISBI-2015 image data have been used.

2.1. Dataset. The proposed algorithm uses the dataset of ISBI-2015 challenge [10] which is grouped in two categories, training and testing data. The training data are named ISBI-21 and are available publicly with 21 MRI images from 5 patients. In the training set, MR brain images of four patients with 4 time points and one with 5 time points with a gap of approximately a year are gathered. The test data are named as ISBI-61 which are not available publicly and have 14 subjects with 61 images. Each subject in the testing set has 4-5 time points, and each time point has a gap of approximately a year. These images contain longitudinal scans of all five patients, as shown in Figure 2. During training, we used 80 percent of the total patches of 100×100 size for training and the remaining 20 percent for validation.

2.2. Proposed Deep Network Architecture. In the CNN architecture, a kernel size and type of filters have to be selected carefully so that it can learn all the features which are useful in the classification of objects. Generally, filters of different sizes and pooling schemes are employed in different CNN layers in order to learn most present features in the data. The inception module, however, uses multiple kernels in each layer in parallel and then pools the features [28]. In the proposed framework, we have investigated the efficacy of inception modules embedded in the CNN model for the segmentation of MS lesions.

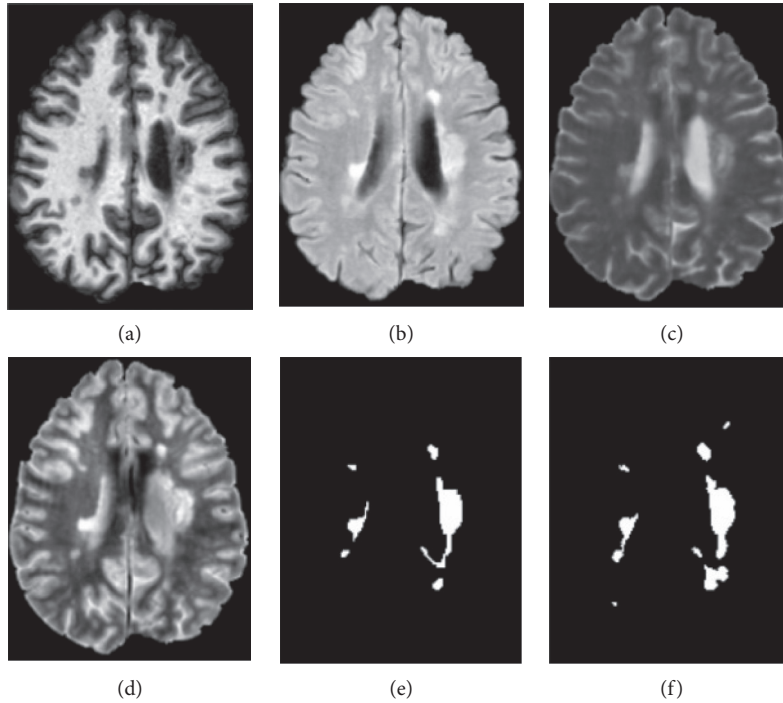


FIGURE 2: Sample of the ISBI dataset: (a) T1w, (b) FLAIR, (c) T2w, and (d) PDw, (e) manual delineation by rater 1, and (f) manual delineation by rater 2.

2.2.1. CNN Model. In the proposed method for segmentation of multiple sclerosis disease, we incorporated three inception modules in our CNN model. Each module consists of 1×1 , 3×3 , 5×5 , max pooling, and average pooling. The CNN model consists of two convolution layers with 64 feature maps followed by inception modules and then three convolutions layers. The final layer has one feature map for the prediction of lesion and nonlesion voxels. Figure 3 shows the complete architecture with inception modules embedded in the CNN layers. The model is trained with two different loss functions, i.e., binary cross entropy (BCE) and structural similarity index measure (SSIM). BCE is a measure of the difference between two probability distributions for a given random variable or a set of events and is used in binary classification tasks, whereas SSIM is a perceptual metric that quantifies image quality degradation caused by losses in data compression. For high similarity in images, the value of BCE is low and the value of SSIM is high.

2.2.2. Inception Module. The fundamental idea behind the GoogLeNet is the introduction of inception modules or inception blocks in the CNN architecture. In CNN, the feature maps learned from the previous layer are given as input to the next layer. The inception module takes the previous layer output and passes it to four different filter operations in parallel, as shown in Figure 4. The feature maps from all the filters are then concatenated to form the final output. The fundamental idea of using a 1×1 kernel in the inception module is just to shrink the depth of the feature maps [29]. The 1×1 convolutions preserve the parameters spatially that can be used when needed. This strategy in the

inception module can lower the dimensions of the feature maps which can eventually drop the computational cost.

2.3. Loss Functions. The proposed model is trained for two loss functions, binary cross entropy (BCE) and structural similarity index measure (SSIM). The BCE loss function tries to maximize the difference of the probability distribution between two classes, in this case, lesion and nonlesion voxels. It measures the performance of a classification model whose output is the probability between 0 and 1, i.e., the output of sigmoid activation. Mathematically, BCE loss for an output y with probability p can be computed as

$$\text{BCE} = -y \log p - (1 - y) \log(1 - p). \quad (1)$$

SSIM is a perception-based loss function that quantifies the similarity between two images. In SSIM, similarity between two images can be computed using a statistical model. Let μ_x and μ_y be the means, σ_x and σ_y be the variances, and σ_{xy} be the covariance of the two images x and y ; then,

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (2)$$

where C_1 and C_2 are regularization constants.

2.4. Model Implementation. The CNN model is implemented in Python using Keras [30] with TensorFlow library [31]. All the experiments were performed on the Nvidia GeForce RTX 2080 GPU. The deep network is trained end to end using patches. During the training phase of the CNN

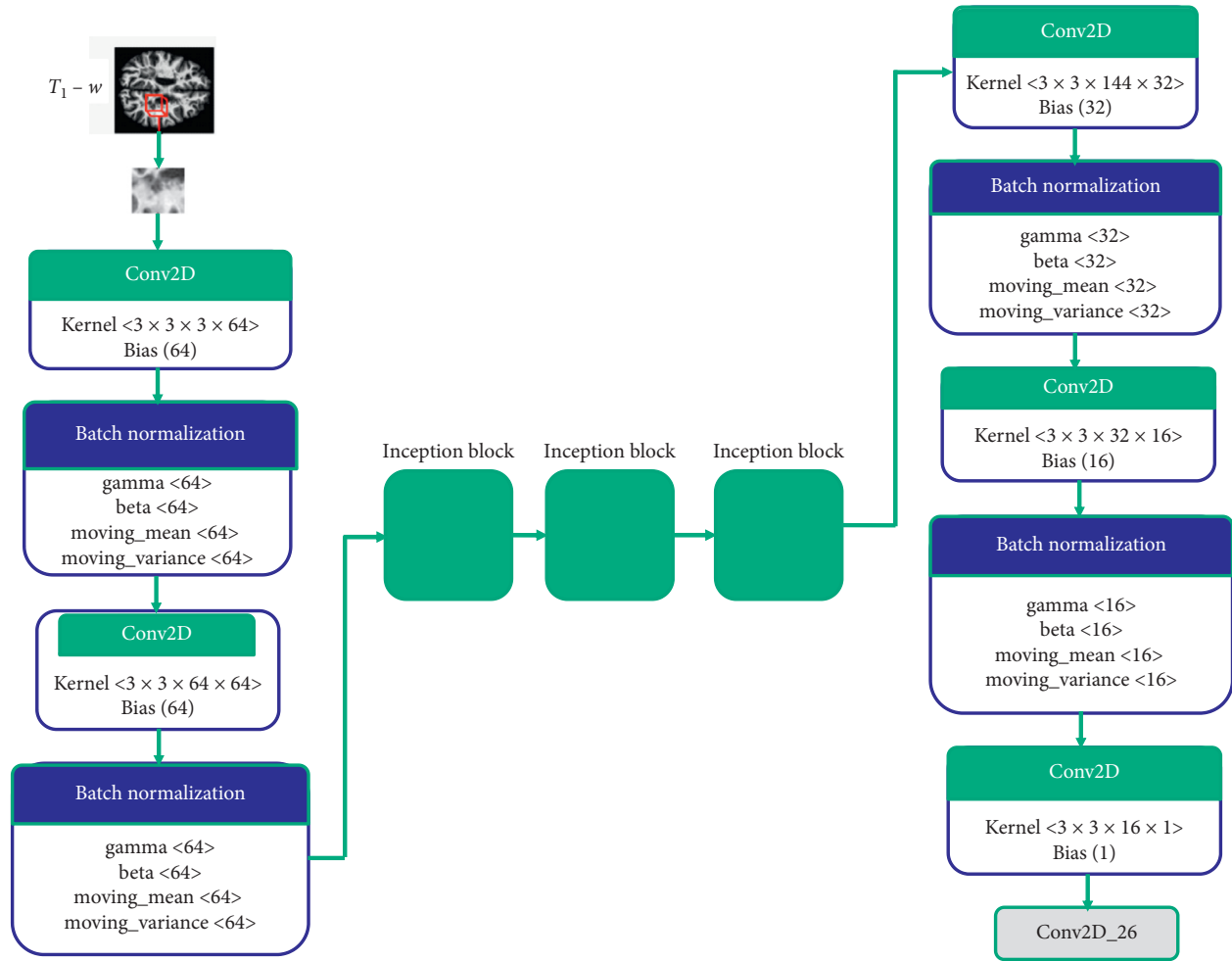


FIGURE 3: Proposed deep network architecture for MS lesion segmentation in brain MRI.

model, the patches are extracted from each slice in MR images. The training set is divided into two subsets, one for training the network and the other for validating the results. The optimization technique employed to update the parameters in the model is the Adam method [32]. In neural network parameter optimization, the Adam method shows better convergence. The hyperparameters used during network training include the fixed learning rate of 0.0001 for 50 epochs. These parameters' setting has produced sufficient convergence to optimal network parameters without overfitting the data. The size of the minibatch is set to 64, and each minibatch includes random number of patches. The best model from the validation set is selected at the 24th epoch which takes 48 hours on the GPUs.

3. Results and Discussion

3.1. Performance Metrics. Standard performance metrics for the assessment of the proposed CNN model have been employed. The Dice similarity coefficient measures reproducibility of segmentation as a statistical validation of manual annotation. Another similar metric is the Jaccard similarity index that gives the intersection between the machine segmentation and the ground truth. Positive

predicted value is the probability that people with a positive screening test result indeed have the condition of interest. The portion of positive voxels in ground truth that is also identified as positive in the automatic segmentation is captured by true positive rate. Lesion-wise true/false positive rate is the number of lesions that overlap/do not overlap in automatic segmentation and the ground truth. The difference is volume of automatic segmentation, and the ground is another important metric for the assessment of the performance of the CNN model. The Pearson correlation coefficient computes the correlation between the automatic segmentation and the ground truth. The overall score gives the average of the combined effect of all these performance metrics in a single number. Table 1 shows formulas for these performance metrics.

3.2. Feature Learning by Inception Modules. As suggested by the literature, the proposed CNN model is trained on T1w, T2w, and FLAIR sequences of the MRI data. Table 2 shows quantitative results for automatic MS lesion segmentation in MRI using the BCE loss function for test images at time points TP . Although, in the results, both Dice and Jaccard similarity indices are reported, they both convey the same

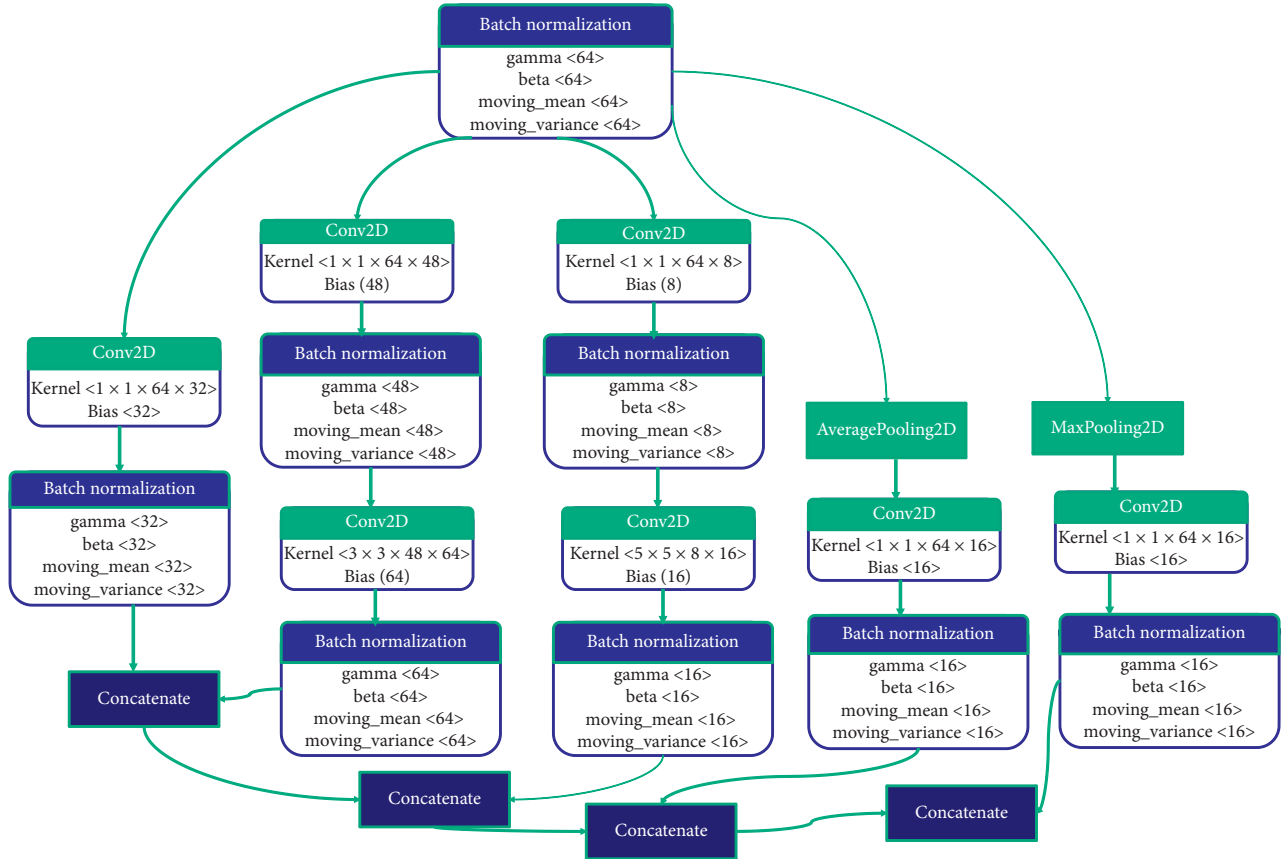


FIGURE 4: Inception block in the proposed CNN architecture.

TABLE 1: Performance metrics used in the proposed solution.

Metric	Formula
Dice similarity coefficient	$DSC = 2TP / (FN + FP + 2TP)$
Jaccard similarity coefficient	$JSC = TP / (TP + FP + FN)$
Positive predicted value	$PPV = TP / (TP + FP)$
True positive rate	$TPR = TP / (TP + FN)$
Lesion-wise true positive rate	$LTPR = LTP / RL$
Lesion-wise false positive rate	$LFPR = LFP / PL$
Volume difference	$VD = TP_s - TP_{gt} / TP_{gt}$
Pearson correlation coefficient	$Cor = cov(X, Y) / \sigma_X \sigma_Y$
Overall score	$SC = (1/ R + S) \sum_{R,S} ((DSC/8) + (PPV/8) + (1 - LFPR/4) + (LTPR/4) + (Cor/4))$

information. The performance metrics observed for the proposed CNN model have significantly outperformed when compared with the existing techniques, as shown in Table 3. Kernels of different sizes used in the inception modules help in extracting discriminative features for the automatic segmentation of MS lesions and background tissues in brain MRI. The most present features are ultimately pooled using max pooling and average pooling at various stages of the inception modules. The number of inception modules used in the CNN model is also very crucial in the architecture design. Using too many inception modules in MS lesion segmentation has degraded the results due to overfitting the model to the data. Also, poor results are obtained when the number of inception modules has been lowered. This may correspond to underfitting the CNN model for the

segmentation of MS lesions. Experiments have also confirmed that a mix of average pooling and max pooling works better by keeping the most present features in the high-level feature maps and averaging them in the low-level feature maps. The authors suggest that, for a specific application, the number and placement of inception modules, filter size, and pooling strategy have to be selected accordingly.

3.3. Comparison of BCE and SSIM Loss Functions. Two loss functions in training the proposed CNN model have been used, BCE and SSIM. Tables 2 and 4 report the quantitative results for the two loss functions. Table 5 gives the comparison of the two loss functions on the basis of the average values of the results. In the MS lesion classification, the BCE

TABLE 2: Quantification of MS lesion segmentation with the BCE loss function.

Subject	TP	Dice	Jaccard	PPV	TPR	LFPR	LTPR	VD
test01	1	0.6639	0.4969	0.8991	0.5263	0.1356	0.5068	0.4147
test01	2	0.6916	0.5286	0.9131	0.5566	0.0806	0.5128	0.3904
test01	3	0.682	0.5174	0.8845	0.5549	0.1452	0.5	0.3726
test01	4	0.6732	0.5074	0.9226	0.5299	0.1034	0.4667	0.4256
test02	1	0.6933	0.5306	0.7548	0.6411	0.1176	0.4653	0.1507
test02	2	0.6823	0.5178	0.8229	0.5828	0.087	0.4969	0.2918
test02	3	0.664	0.497	0.8241	0.5559	0.0638	0.4867	0.3254
test02	4	0.6409	0.4716	0.8529	0.5134	0.0631	0.5411	0.3981
test02	5	0.7099	0.5502	0.8444	0.6123	0.1277	0.4157	0.2748
test03	1	0.4949	0.3288	0.8944	0.3421	0.125	0.3056	0.6175
test03	2	0.5132	0.3451	0.9271	0.3548	0.1379	0.3333	0.6173
test03	3	0.4988	0.3322	0.9457	0.3387	0	0.4375	0.6419
test03	4	0.5838	0.4122	0.9242	0.4266	0.08	0.4667	0.5384
test04	1	0.8168	0.6903	0.8693	0.7702	0.1154	0.6944	0.114
test04	2	0.7928	0.6567	0.8205	0.7668	0.36	0.5172	0.0654
test04	3	0.8067	0.676	0.8099	0.8035	0.08	0.7586	0.0078
test04	4	0.7999	0.6665	0.8095	0.7905	0.2759	0.697	0.0234
Average		0.6711	0.5133	0.8658	0.5686	0.1234	0.5060	0.3335

TABLE 3: Comparison with the existing techniques.

Method	SC	DSC	PPV	LTPR	LFPR	VD
Birenbaum and Greenspan [12]	90.07	0.6271	0.7889	0.5678	0.4975	0.3522
Litjens et al. [13]	86.92	0.5009	0.5491	0.4288	0.5765	0.5707
Valverde et al. [14]	91.33	0.6294	0.7866	0.3669	0.1529	0.3384
Aslani et al. [16]	89.85	0.4856	0.7402	0.3034	0.1708	0.4768
Proposed	90.84	0.6306	0.7888	0.5736	0.2512	0.3444

TABLE 4: Quantification of MS lesion segmentation with the SSIM loss function.

Subject	TP	Dice	Jaccard	PPV	TPR	LFPR	LTPR	VD
test01	1	0.6061	0.4348	0.866	0.4662	0.0408	0.4247	0.4617
test01	2	0.6296	0.4595	0.8677	0.4941	0.0702	0.4487	0.4306
test01	3	0.6179	0.447	0.85	0.4853	0.0556	0.4024	0.429
test01	4	0.6194	0.4486	0.8814	0.4775	0.0909	0.4267	0.4582
test02	1	0.6608	0.4934	0.7923	0.5667	0.0864	0.375	0.2847
test02	2	0.631	0.4609	0.8246	0.5111	0.101	0.4025	0.3802
test02	3	0.5987	0.4273	0.8224	0.4707	0.0667	0.34	0.4277
test02	4	0.5909	0.4194	0.852	0.4523	0.0467	0.4452	0.4692
test02	5	0.6617	0.4944	0.8427	0.5446	0.0978	0.3614	0.3537
test03	1	0.4394	0.2815	0.7986	0.3031	0.0769	0.3333	0.6205
test03	2	0.4663	0.3041	0.8313	0.3241	0.08	0.4	0.6102
test03	3	0.4597	0.2985	0.852	0.3148	0.0909	0.375	0.6305
test03	4	0.5254	0.3563	0.8287	0.3847	0.2	0.3333	0.5358
test04	1	0.776	0.634	0.8535	0.7115	0.1304	0.5833	0.1664
test04	2	0.762	0.6156	0.8345	0.7012	0.2353	0.4138	0.1597
test04	3	0.7729	0.6299	0.8089	0.74	0.2273	0.5862	0.0852
test04	4	0.7792	0.6383	0.8187	0.7433	0.1667	0.6364	0.0921
test05	1	0.433	0.2763	0.3939	0.4806	0.45	0.2778	0.2201
test05	2	0.4622	0.3006	0.5652	0.391	0.1852	0.4082	0.3082
test05	3	0.5353	0.3655	0.6448	0.4576	0.1951	0.4923	0.2903
test05	4	0.5169	0.3485	0.6145	0.446	0.1923	0.375	0.2742
Average		0.5974	0.4350	0.7830	0.4984	0.1374	0.4210	0.3661

TABLE 5: Quantitative comparison of BCE and SSIM loss functions.

Loss function	SC	DSC	PPV	LTPR	LFPR	VD
BCE	90.84	0.6306	0.7888	0.5736	0.2512	0.3444
SSIM	89.01	0.5934	0.7288	0.4476	0.1935	0.3999

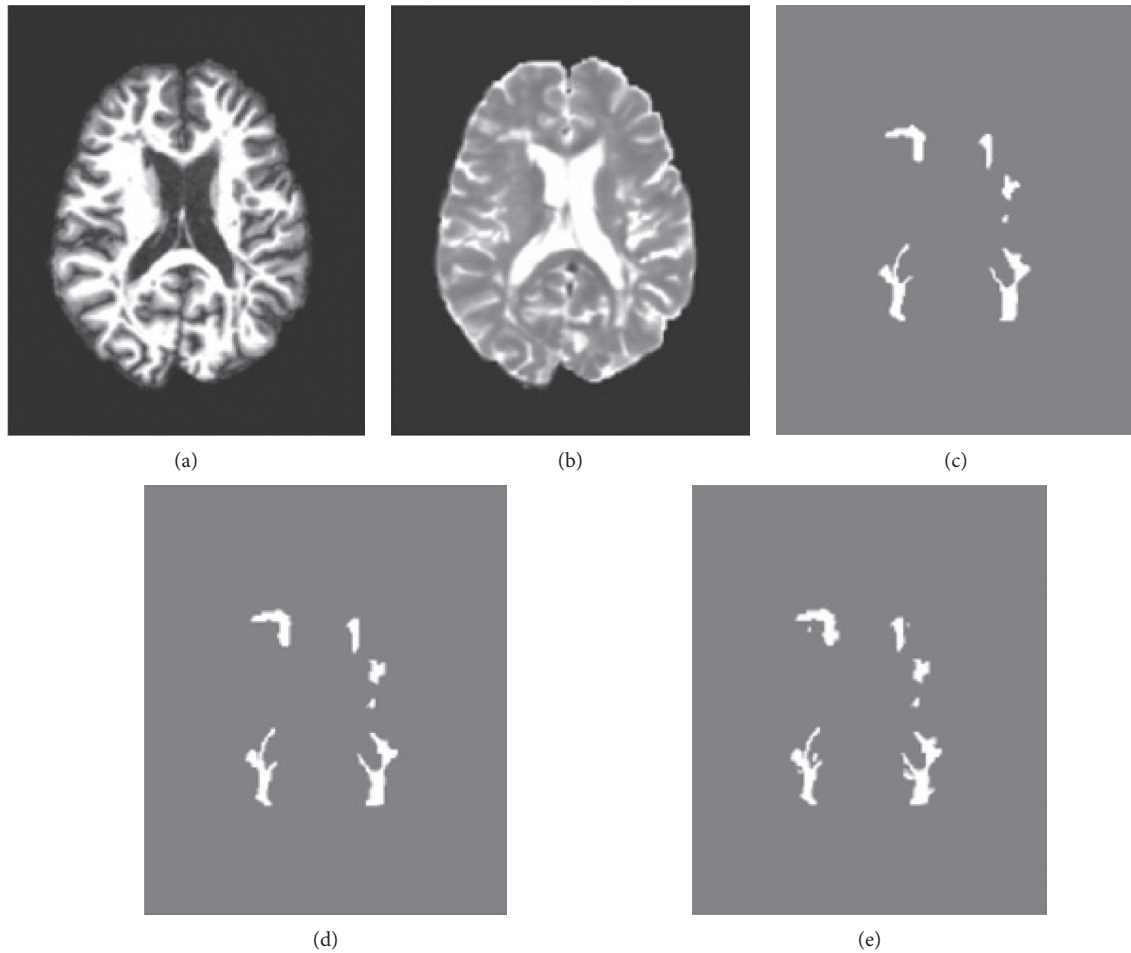


FIGURE 5: Comparison of the segmentation results when using BCE and SSIM loss functions. (a) T1w. (b) T2w. (c) Rater 1. (d) BCE. (e) SSIM.

loss function seems to work better than SSIM. This sounds very intuitive as BCE tries to evaluate the difference in the maximum likelihood between the predictions and ground truths. SSIM, on the other hand, quantifies the perceptual differences between the predictions and the ground truths. It uses luminance, contrast, and structure features to compute the similarity between two images. The reason why the BCE loss function works better than SSIM is because loss functions also depend on the activation functions used in the output layer. For sigmoid activation, the literature suggests that the BCE loss function is the natural choice due to its accuracy and efficiency. The automatic MS lesion segmentation using BCE and SSIM loss functions is illustrated in Figure 5.

3.4. Comparison with Existing Techniques. The proposed methodology is compared with different published techniques for MS lesion segmentation using the ISBI-2015 dataset. The comparison of the results is shown in Table 3. The Dice coefficient, PPV, LTPR, and VD obtained in the proposed methodology show that the model is generalized well for successfully handling new data. The performance of

Birenbaum and Greenspan's model, multiview CNN, includes a score of 90.07, DSC of 62.71%, PPV of 78.89%, LTPR of 56.78%, LFPR of 49.75%, and VD of 35.22%. This model produced the best LFPR result among the five techniques compared here. The performance of Litjens' CNN model used was the worst compared to the other techniques. The performance value of score was 86.92, the DSC was 50.09%, PPV was 54.91%, LTPR was 42.88%, LFPR was 57.95%, and VD was 57.07%. The second best performance was shown by the cascaded CNN architecture proposed by Ververde et al. It includes a score of 91.33, DSC of 62.94%, PPV of 78.66%, LTPR of 36.69%, LFPR of 15.29%, and VD of 33.84%. The results for the multibranch CNN model proposed by Aslani et al.'s model includes a score of 89.85, DSC of 48.56%, PPV of 74.01%, LTPR of 30.34%, LFPR of 17.08%, and VD of 47.68%. Finally, the performance of the proposed model was the best with a score of 93.81, DSC of 67.11%, PPV of 86.58%, LTPR of 50.60%, LFPR of 12.34%, and VD of 33.35%. The value of LTPR was the only metric that was worse than Valverde's and Aslani's models. The shortcoming in LFPR can further be investigated in the future model of the present work.

4. Limitations in Real Clinical Studies

The proposed work is an attempt to prove the efficacy of AI-based techniques in medical applications. In recent years, AI has gained reputation in automating tedious routine works in clinical settings. However, the diversity and inadequacy of the patient data for training a deep network have hampered practical use of AI-based techniques in clinics. As more and more data will become available and as deep neural networks will become more efficient, the practicability of these techniques will definitely improve.

5. Conclusions and Future Works

In this work, a CNN model with inception modules is investigated in automatic segmentation of MS lesions in MRI. The CNN model with inception modules seems to pick MS lesions of different sizes and shapes more successfully. The key advantage of inception modules is the use of different kernels such as 1×1 , 3×3 , and 5×5 that tend to extract salient features in the input of varying sizes. This improves the Dice coefficient, PPV, LTPR, and VD of the segmentation when compared to the existing techniques. These results have outperformed all the existing techniques. The success of Ververde's model can also be attributed to accurate learning of MS lesion features of various sizes and shapes. The performance of Birenbaum and Greenspan's model, multiview CNN, includes a score of 90.07, DSC of 62.71%, PPV of 78.89%, LTPR of 56.78%, LFPR of 49.75%, and VD of 35.22%. This model produced the best LFPR result among the five techniques compared here. The performance of Litjens' CNN model was the worst compared to the other techniques. The performance of the model used had a score of 86.92, DSC of 50.09%, PPV of 54.91%, LTPR of 42.88%, LFPR of 57.95%, and VD of 57.07%. The second best performance was shown by the cascaded CNN architecture proposed by Ververde et al. It includes a score of 91.33, DSC of 62.94%, PPV of 78.66%, LTPR of 36.69%, LFPR of 15.29%, and VD of 33.84%. The results for the model proposed by Aslani et al.'s model, multibranch CNN, includes a score of 89.85, DSC of 48.56%, PPV of 74.01%, LTPR of 30.34%, LFPR of 17.08%, and VD of 47.68%. Finally, the performance of the proposed model was the best with a score of 93.81, DSC of 67.11%, PPV of 86.58%, LTPR of 50.60%, LFPR of 12.34%, and VD of 33.35%. The value of LTPR was the only metric that was worse than Ververde's and Aslani's models. In the present study, we have also discovered that the BCE loss function works better than the SSIM loss function. The intuition behind this behavior of the model is that BCE tries to maximize the differences between the probability distributions predictions and ground truths. SSIM, on the other hand, seems to converge to local minima while quantifying the error loss. Another important reason is the sigmoid activation function used in the output layer for the binary classification. The authors believe this naturally supports the BCE loss function to produce more accurate and efficient results. In the future, this work can be further extended to integrate in different architectures such as the residual network (ResNet), UNet, parallel CNN, and

cascaded CNN on multiple datasets which are publicly available. The incorporation of event-driven processing can improve the performance of the suggested solution in terms of computational efficiency and compression [33–36]. Investigation based on this axis is another prospect.

Data Availability

The dataset, studied in this paper, is publically available at <https://biomedicalimaging.org/2015/program/isbi-challenges>.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was funded by Effat University, Jeddah, Saudi Arabia. The authors are thankful to Mr. Najam ur Rahman for his contribution in designing and implementing the CNN architecture. They wish to acknowledge the help and facility provided by the Artificial Intelligence in Medicine (AIM) Laboratory in Ghulam Ishaq Khan (GIK) Institute of Engineering Sciences and Technology and College of Engineering of Effat University.

References

- [1] A. Carass, S. Roy, A. Jog et al., "Longitudinal multiple sclerosis lesion segmentation: resource and challenge," *NeuroImage*, vol. 148, pp. 77–102, 2017.
- [2] C. Egger, R. Opfer, C. Wang et al., "MRI FLAIR lesion segmentation in multiple sclerosis: does automated segmentation hold up with manual annotation?" *NeuroImage: Clinica*, vol. 13, pp. 264–270, 2017.
- [3] P. Sati, J. Oh, J. Oh et al., "The central vein sign and its clinical evaluation for the diagnosis of multiple sclerosis: a consensus statement from the North American imaging in multiple sclerosis cooperative," *Nature Reviews Neurology*, vol. 12, no. 12, pp. 714–722, 2016.
- [4] D. García-Lorenzo, S. Francis, S. Narayanan, D. L. Arnold, and D. L. Collins, "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging," *Medical Image Analysis*, vol. 17, no. 1, pp. 1–18, 2013.
- [5] T. Kalincik, M. Vaneckova, M. Tyblova et al., "Volumetric MRI markers and predictors of disease activity in early multiple sclerosis: a longitudinal cohort study," *PLoS One*, vol. 7, no. 11, Article ID e50101, 2012.
- [6] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, Boston, MA, USA, June 2015.
- [7] K. Javed, N. Ud Din, S. Bae, and J. Yi, "Image unmosaicing without location information using stacked GAN," *IET Computer Vision*, vol. 13, no. 6, pp. 588–594, 2019.
- [8] K. Javed, N. U. Din, S. Bae, R. S. Maharajan, D. Seo, and J. Yi, "UMGAN: generative adversarial network for image unmosaicing using perceptual loss," in *Proceedings of the 16th International Conference on Machine Vision Applications (MVA)*, pp. 27–31, Tokyo, Japan, May 2019.

- [9] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Traboulsee, and R. Tam, "Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.
- [10] Isbi 2015, "International symposium on biomedical imaging," *IEEE Pulse*, vol. 6, no. 2, p. 61, 2015.
- [11] Miccai 2008, "Medical image computing and computer-assisted intervention," in *Proceedings of the 11th International Conference*, vol. 5241, Springer, New York, NY, USA, Lecture Notes in Computer Science, New York, NY, USA, September 2008.
- [12] A. Birenbaum and H. Greenspan, "Longitudinal multiple sclerosis lesion segmentation using multi-view convolutional neural networks," in *Deep Learning and Data Labeling for Medical Applications*, pp. 58–67, Springer, Berlin, Germany, 2016.
- [13] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [14] S. Valverde, M. Cabezas, E. Roura et al., "Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach," *NeuroImage*, vol. 155, pp. 159–168, 2017.
- [15] S. Roy, J. A. Butman, D. S. Reich, P. A. Calabresi, and D. L. Pham, "Multiple sclerosis lesion segmentation from brain MRI via fully convolutional neural networks," 2018, <https://arxiv.org/abs/1803.09172>.
- [16] S. Aslani, M. Dayan, L. Storelli et al., "Multi-branch convolutional neural network for multiple sclerosis lesion segmentation," *NeuroImage*, vol. 196, pp. 1–15, 2019.
- [17] M. Maleki, M. Teshnehlab, and M. Nabavi, "Diagnosis of multiple sclerosis (MS) using convolutional neural network (CNN) from MRIs," *Global Journal of Medicinal Plant Research*, vol. 1, no. 1, pp. 50–54, 2012.
- [18] H. Huang, G. Yang, W. Zhang et al., "A deep multi-task learning framework for brain tumor segmentation," *Frontiers in Oncology*, vol. 11, 2021.
- [19] W. Zhang, G. Yang, H. Huang et al., "ME-Net: multi-encoder net framework for brain tumor segmentation," *International Journal of Imaging Systems and Technology*, 2021.
- [20] M. Soltaninejad, G. Yang, T. Lambrou et al., "Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels," *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 69–84, 2018.
- [21] M. Soltaninejad, G. Yang, T. Lambrou et al., "Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 2, pp. 183–203, 2017.
- [22] M. Soltaninejad, L. Zhang, T. Lambrou, G. Yang, N. Allinson, and X. Ye, *MRI Brain Tumor Segmentation and Patient Survival Prediction Using Random Forests and Fully Convolutional Networks International MICCAI Brainlesion Workshop*, Springer, Cham, Berlin, Germany, 2017.
- [23] Y. Jin, G. Yang, Y. Fang et al., "3D PBV-Net: an automated prostate MRI data segmentation method," *Computers in Biology and Medicine*, vol. 128, Article ID 104160, 2021.
- [24] G. Yang, J. Chen, Z. Gao et al., "Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention," *Future Generation Computer Systems*, vol. 107, pp. 215–228, 2020.
- [25] Y. Liu, G. Yang, M. Hosseiny et al., "Exploring uncertainty measures in bayesian deep attentive neural networks for prostate zonal segmentation," *IEEE Access*, vol. 8, Article ID 151817, 2020.
- [26] Y. Liu, K. Sung, G. Yang et al., "Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention," *IEEE Access*, vol. 7, Article ID 163626, 2019.
- [27] F. Raschke, T. R. Barrick, T. L. Jones, G. Yang, X. Ye, and F. A. Howe, "Tissue-type mapping of gliomas," *NeuroImage: Clinica*, vol. 21, Article ID 101648, 2019.
- [28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Las Vegas, NV, USA, June 2016.
- [30] F. Chollet, "Keras," 2015, <https://keras.io/>.
- [31] M. Abadi, A. Agarwal, P. Barham et al., "Tensorflow: large-scale machine learning on heterogeneous systems," 2015, <https://arxiv.org/abs/1603.04467>.
- [32] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, <https://arxiv.org/abs/1412.6980>.
- [33] S. M. Qaisar, "Event driven filtering an intelligent technique for activity and power consumption reduction," *Int. J. Circuits Syst. Signal Process.* vol. 8, 2014.
- [34] S. Mina Qaisar, D. Sidiya, M. Akbar, and A. Subasi, "An event-driven multiple objects surveillance system," *International Journal of Electrical and Computer Engineering Systems*, vol. 9, no. 1, pp. 35–44, 2018.
- [35] S. M. Qaisar and A. Subasi, "Effective epileptic seizure detection based on the event-driven processing and machine learning for mobile healthcare," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2020.
- [36] S. M. Qaisar, "Signal-piloted processing and machine learning based efficient power quality disturbances recognition," *PloS One*, vol. 16, no. 5, 2021.

Research Article

ECG Signal Modeling Using Volatility Properties: Its Application in Sleep Apnea Syndrome

Maryam Faal  and Farshad Almasganj 

Department of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran

Correspondence should be addressed to Farshad Almasganj; almas@aut.ac.ir

Received 7 May 2021; Revised 21 May 2021; Accepted 14 June 2021; Published 8 July 2021

Academic Editor: G R Sinha

Copyright © 2021 Maryam Faal and Farshad Almasganj. 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 study presents and evaluates the mathematical model to estimate the mean and variance of single-lead ECG signals in sleep apnea syndrome. Our objective is to use the volatility property of the ECG signal for modeling. ECG signal is a stochastic signal whose mean and variance are time-varying. So, we propose to decompose this nonstationarity into two additive components; a homoscedastic Autoregressive Integrated Moving Average (ARIMA) and a heteroscedastic time series in terms of Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH), where the former captures the linearity property and the latter the nonlinear characteristics of the ECG signal. First, ECG signals are segmented into one-minute segments. The heteroskedasticity property is then examined through various tests such as the ARCH/GARCH test, kurtosis, skewness, and histograms. Next, the ARIMA model is applied to signals as a linear model and EGARCH as a nonlinear model. The appropriate orders of models are estimated by using the Bayesian Information Criterion (BIC). We assess the effectiveness of our model in terms of mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The data in this article is obtained from the Physionet Apnea-ECG database. Results show that the ARIMA-EGARCH model performs better than other models for modeling both apneic and normal ECG signals in sleep apnea syndrome.

1. Introduction

ECG signal has an essential role in medical diagnosis for the study of cardiac function and abnormalities. Considering the abnormal activity of heart or heart rate variation (HRV) could be an indicator of some diseases such as congestive heart failure (CHF) [1], sudden cardiac death (SCD) [2], and obstructive sleep apnea (OSA) [3]. OSA is a common respiratory disease characterized by a cessation in the airflow for at least 10 seconds [4]. The literature has stated that sleep apnea affects approximately 2% of women and 4% of men and that most of them are overweight [5]. Apnea increases accidents and mortality rates. Previous research considered apnea as a public health risk compared with smoking. Untreated OSA can also cause depression, high blood pressure, stroke, hypertension, death, and an increased risk of long-term and short-term disease. It increases the risk of myocardial infarction by up to 20% and heart attack by up to

40% [6]. Accurate and early diagnosis is an essential step in the control and prevention of sleep apnea. So, it attracts much attention in ECG research.

Polysomnography (PSG), a multimodality system, is the most accurate and precise method of sleep monitoring, which can be used to describe sleep stages and disorders. PSG measures electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), electrooculogram (EOG), and respiratory airflow and peripheral oxygen saturation (SpO₂). After collecting the PSG data, physicians rate the OSA events using statistical methods. The PSG system, on the other hand, has two major defects. First, manually scoring sleep stages according to the guideline requires the use of physicians and appropriate sensors is time-consuming and expensive. Therefore, PSG can only be performed in sleep laboratories, which delays detection and results in a long waiting list. Second, it is an obtrusive approach, which requires the attachment of several sensors

and wires. Sleep normality can be disrupted by the sensors and wires, making PSG inappropriate for long-term sleep researches [7]. So, developing methods that can accurately detect apnea with a few signals at home is critical. These approaches were focused on biosignals such as respiratory, snoring, SpO₂, and ECG signals, and several authors have achieved a high level of performance in terms of OSA detection [8–11].

Using wearable devices with some necessary biosensors for sleep disorder diagnosis is safer because these devices are designed with unobtrusiveness insight. Furthermore, they are simpler to use, easier to find, and less expensive than clinical measurements. The ECG is one of the most reliable physiological signals given by various wearable devices and used in many OSA studies [12]. Several researchers have suggested innovative methods for evaluating sleep quality and sleep apnea using only a single-lead ECG. The presence of irregular characteristics in the ECG signal is seen as a warning sign of sleep apnea. When sleep apnea occurs, the oxygen saturation decreases and the cardiovascular system is activated to maintain the oxygen intake constant. Furthermore, according to a clinical study in [7], patients' compliance is extremely low when wearing the pressure transducer sensor to achieve nasal and oral respiration. Patients usually pulled out the nasal cannula and nasal airflow data can be unreliable compared to the ECG signal due to lead loss. As a consequence, we chose ECG signals to model OSA and normal events in this study.

The majority of ECG approaches proposed in the literature for sleep apnea detection are based on feature extraction from single-lead ECG signals and using classifiers [8–11]. Zarei and Mohammadzadeh Asl [10] proposed a novel approach based on single-lead ECG autoregressive (AR) modeling and ECG feature extraction using the spectral autocorrelation function. Sequential forward feature selection (SFFS) is used to select the most appropriate features, which are then fed into a random forest to classify normal and apnea epochs. Singh et al. [8] extracted the mean and the standard deviation from the instantaneous amplitude (IA) and instantaneous frequency (IF) of each reconstructed component (RC) of heartbeat intervals and electrocardiogram-derived respiration (EDR) signals. Then, stacked autoencoder-based deep neural network (SAE-DNN) and support vector machine (SVM) are used to categorize apneic and normal segments. Zhang et al. [11] suggested a sleep monitoring model based on a single-channel electrocardiogram using a convolutional neural network (CNN). Rajesh et al. [9] extracted moments of power spectrum density, waveform complexity measures, and higher-order moments from the 1 min segmented ECG subbands obtained from discrete wavelet transform (DWT). The acquired feature set is fed to various classifiers such as SVM, linear discriminant analysis (LDA), random forest, and k nearest neighbors (kNN). All of these methods are purely data-driven.

Signal modeling and feature extraction is an essential step in the analysis of the ECG signals. Mathematical modeling of the ECG signal is widely used in many cardiovascular studies, such as ECG signals denoising, ECG

beats segmentation, arrhythmias detection, heart rate estimation, and synthetic ECG signal generation [13, 14]. Mathematically, modeling helps to understand how the model's factors influence the sensitivity and specificity in computer-aided diagnosis methods. Different models have been used for ECG signals such as autoregressive (AR) model [15], autoregressive moving average (ARMA) [16], generalized autoregressive moving average (GARMA) [17], data flow graph (DFG) model [18], generalized orthogonal forward regression (GOFR) [19], Gaussian mesa and bi-Gaussian functions [19], hidden Markov models (HMM) [20], morphological models [21], Hermite basis functions [22], Gaussian model [23], principal component analysis [24], Kalman filter [25], and time-varying autoregressive model (TVAM) [26]. These ECG models have fitted mathematical representations into HRV or the points of ECG signals and need ECG preprocessing to achieve essential components such as QRS complexes, P-wave, and T-wave. The main drawback is that using these components needs to determine the exact location of waves, which increases the computation time, and the system performance depends on the method used. Therefore, in this article, we used an unprocessed single-lead ECG signal, which is lower in cost. Only a few researchers used models to detect apnea [26–28]. Mendez et al. [26] used a time-varying autoregressive model (TVAM) to assess power spectral densities for the QRS complex area and the RR intervals. This study aims to use time series models to propose a new ECG signal model. This model can be used to detect normal and apneic ECG signals. Sharma and Sharma [27] used a linear combination of the lower order Hermite basis functions to estimate each QRS complex of the ECG signal. Hassan et al. [28] used a tunable-Q factor wavelet transformation (TQWT), and each subband was modeled using symmetric Normal Inverse Gaussian (NIG) pdf. One issue neglected in previous articles is that apnea is associated with fluctuations in the ECG process. This property can be used to model apnea and normal ECG signals.

Time-varying conditional standard deviation, usually called volatility, describes periods of high oscillations distributed with relative calm periods and plays a vital role in predicting time series fluctuations [29]. In statistics, heteroskedasticity indicates that a variable standard error is not constant over time. It has been proven that heteroskedasticity modeling through Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and their variants are helpful in the modeling varying volatility in nonstationary time series [29]. Real-world time series such as ECG signals have volatility. In the literature, ECG volatility and heteroskedasticity during apnea are underestimated. Huan Tsoukalas [30] showed that Integrated GARCH(1, 1) could model apneic ECG segments, and ARCH(1) can model the normal ECG recordings. Experimental observations showed that cardiovascular variations are complex, nonlinear, and nonstationary [31]. Linear models like AR, Moving Average (MA), and ARMA are coarse estimations of real-world systems and usually have poor performance in forecasting the evolution of nonstationary and nonlinear processes [31].

The ARIMA model is usually used to model these patterns [31]. Therefore, in this article, the ARIMA model is used to model linear features of ECG signals. The consistency of conditional variance is one of the essential assumptions used by conventional ARIMA models to forecast future values. If we assume that the ARIMA model fits an ECG signal, the conditional variance should be constant. It has been shown that during apnea, the homoskedasticity assumption is not correct [29]. Instead of using the ARIMA model, which focuses only on predicting the conditional mean of future values, clusters of abundant variance need to use models that can simultaneously predict both the conditional mean and the conditional heteroskedasticity of the system. Since ARIMA is a linear model, it cannot reflect nonlinear characteristics such as volatility. ARIMA is a linear model that reveals linear characteristics of the ECG signal, and nonlinear features such as heteroskedasticity of the ECG signal remain in residuals, which are modeled using a nonlinear ARCH or GARCH model. Therefore, the proposed model is based on the linear ARIMA model and a nonlinear GARCH model. First, the heteroskedasticity property of the ECG signal is examined through the ARCH/GARCH test, kurtosis, skewness, and histograms. Next, the linear characteristics of ECG signals are modeled using the ARIMA model. To model the nonlinear heteroskedasticity features of ECG signals, we use three different versions of the original GARCH model: GARCH, Glosten-Jagannathan-Runkle GARCH (GJR-GARCH), and Exponential GARCH (EGARCH). Finally, the model with the maximum likelihood value is selected as the best model among existing models. The best orders of the models are then selected using Bayesian Information Criteria (BIC) and the performance of the proposed model is assessed in terms of four criteria: mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

The article is organized as follows: Section 2 briefly addresses the fundamentals of ARIMA and GARCH models. Section 3 provides descriptions of the proposed ECG signal model, where data statistic measures and performance metrics are also provided. Associated numerical results are given in Section 4. Section 5 discusses the obtained results, and the article is concluded in this section, too.

2. Prediction Model

In this section, a brief review of the ARIMA and GARCH family models will be presented, respectively.

2.1. Autoregressive Integrated Moving Average (ARIMA). Autoregressive Integrated Moving Average (ARIMA) model is generalized as an ARMA model used in cases where the signal is nonstationary. The ARIMA(p, d, q) model consists of three parts: Autoregressive (AR), Integrated (I), and Moving Average (MA). For a given time series, an ARMA(p, q) model with p as the number of autoregressive terms and q as the sum of lagged forecast errors of the following type:

$$\left(1 - \sum_{k=1}^p a_k\right) X_t = \left(1 + \sum_{k=1}^q b_k\right) \epsilon_t, \quad (1)$$

where p is the number of autoregressive (AR) terms, a_k s are AR parameters, q is the number of terms in moving average (MA), b_k s are MA parameters, and ϵ_t is an independent error term. ARMA models assume that signals are stationary, and the performance of the ARMA model reduces whenever time trends and seasonality features exist. Methods such as ARIMA are used to remove or reduce these nonstationarity moments [32].

The ARIMA model of orders (p, d, q) is a process, X_t , whose differences $(1 - L)^d X_t$ satisfy an ARMA(p, q) model, which is stationary. d is a nonnegative integer (usually less than 2) and represents d^{th} difference of X_t to find a stationary time series. ARIMA models are always assuming the data variance is constant. The following equation can be used to describe the ARIMA(p, d, q) model.

$$\left(1 - \sum_{k=1}^p a_k L^k\right) (1 - L)^d X_t = \left(1 + \sum_{k=1}^q b_k L^k\right) \epsilon_t, \quad (2)$$

where $L\{X_t\} = X_t - X_{t-1}$ and d is the number of differences required to stationary time series, a_k s are AR parameters, p is the model's autoregression order (AR) and the number of differential series lags, b_k s are MA parameters, q is the order of the model's moving average (MA) and the number of prediction error lags, and ϵ_t is independent error terms. Modeling the ECG signal via ARIMA is essentially a three-stage iterative process that involves the following: identifying model order, model estimation, and checking the model.

2.2. Autoregressive Conditional Heteroskedasticity (ARCH) Model. Conditional volatility models are known as heteroscedastic models, meaning the variance is not constant. These models were widely used in finance because data appear to differ or be highly volatile in these areas. Volatility models were first introduced with the Autoregressive Conditional Heteroskedasticity (ARCH) model in economics by Engle [33]. In this model, the conditional variance varies during time as a function of previous errors. Suppose that $Z(t)$ is a strong white noise process, $Z(t) \sim N(0, 1)$; the process y_t can be an ARCH process, if a process is stationary and has the following properties:

$$y_t = \sigma_t Z_t, \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2, \quad (3)$$

where $Z(t)$ is a stochastic piece, σ_t is a standard deviation depending on time, q is the length of ARCH lags, $\alpha_0 \geq 0$, and $\alpha_i \geq 0, i = 1, 2, \dots, q$.

2.3. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model. Although the ARCH method has proven useful in modeling data instability, a relatively long lag is often needed. A simplified version of the ARCH

model, i.e., Generalized Autoregressive Conditional Heteroskedasticity (GARCH), was proposed by Bollerslev [34] to allow both longer memory and a more stable lag structure. GARCH modeling is a statistical method for time series modeling whose variances are a stochastic process widely used in modeling financial time series. The main idea of this model is that the conditional variance σ_t^2 has an AR structure and also depends on past values of σ_t . GARCH(p, q) is defined as follows:

$$\begin{aligned} y_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \end{aligned} \quad (4)$$

where p is the order of GARCH terms, q is the order of ARCH terms, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$. The GARCH model's application in various fields proves its ability to model data's uncertainty.

2.4. Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) Model. Some more complex GARCH parameterizations were suggested for modeling the conditional variance after the standardized GARCH model. These sophisticated models aim to capture better the empirically demonstrated stylized facts of the mechanism of conditional variance. The asymmetric effect of the negative return shocks, for example, is identified by the Exponential GARCH (EGARCH) model [35] and the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model [36]. To conclude, there is no consensus on which GARCH model offers the best for forecasting. Different studies prefer different GARCH parameters, with different study times, different asset groups, and different output assessment requirements. The asymmetric GARCH models are, however, usually favored over the symmetric GARCH model. The first model we used is Glosten-Jagannathan-Runkle GARCH (GJR-GARCH). GJR-GARCH is a nonlinear GARCH model that considers the asymmetries in response to the conditional variance in an innovation. The GJR-GARCH model's principle is that conditional variance dynamics admit that a regime switch depends on the sign of past innovations. It models the asymmetry in GARCH and defined by the following equations:

$$\begin{aligned} y_t &= \sigma_t Z_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{i=1}^q \gamma_i I_{t-i} y_{t-i}^2, \end{aligned} \quad (5)$$

where Z_t is i.i.d., $I_{t-i} = 0$ if $y_{t-i} \geq 0$, and $I_{t-i} = 1$ if $y_{t-i} < 0$

2.5. Exponential GARCH (EGARCH) Model. Nelson introduced the Exponential GARCH (EGARCH) (p, q) model [35] to catch the asymmetry:

$$y_t = \sigma_t Z_t,$$

$$\log(\sigma_t^2) = \alpha_0 + \sum_{k=1}^q \beta_k g(Z_{t-k}) + \sum_{k=1}^p \alpha_k \log(\sigma_{t-k}^2), \quad (6)$$

$$g(Z_t) = \theta Z_t + \lambda [|Z_t| - E(|Z_t|)],$$

$$Z_t = \frac{y_t}{\sigma_t},$$

where σ_t^2 is a conditional variance; α_0 , α , β , θ , and λ are coefficients. Z_t can be a regular normal variable, or it can come from a generalized distribution of errors. The structure of $g(Z_t)$ allows for the sign and magnitude of Z_t to have different impacts on the volatility. Since $\log(\sigma_t^2)$ can be negative, the parameters are not subject to sign restrictions.

3. ECG Signal Modeling Using ARIMA-EGARCH

ECG signals are segmented into one-minute segments. Each segment has 6000 samples, which contain either normal or apneic conditions. We calculated the order of the models from randomly selected 50% of the segments and then used a test and validated the model's output on the remaining segments. The overall proposed scheme is demonstrated in Algorithm 1.

3.1. Statistical Tests for ARCH/GARCH Effect. GARCH models can only be used when the data are volatile. We need to verify the volatility of data before using any GARCH models. We use various tests to explore ECG segments' statistical properties to examine whether GARCH family models provide an efficient ECG signals modeling. One of the approaches is by measuring histograms for verifying data distribution. Kurtosis is the indicator of peaks in the data distribution, and skewness is a symmetrical representation of a mean distribution. The series is volatile when the kurtosis value is greater than 3 and is skewed to either the left or the right. In simple terms, the heavy-tailed distribution indicates that the probability of encountering large deviations from the mean is higher than in the case of normal distribution. Kurtosis and skewness measurements are used as follows:

$$\begin{aligned} \text{Kurtosis} &= \frac{E(x - \mu)^4}{\sigma^4}, \\ \text{Skewness} &= \frac{E(x - \mu)^3}{\sigma^3}, \end{aligned} \quad (7)$$

where μ and σ are the mean and the standard deviation of x . Another method for testing ARCH/GARCH effects is the ARCH/GARCH test suggested by Engle [33]. This approach tests a null hypothesis that the ARCH/GARCH effect does

- (1) : decompose ECG signals into one-minute segments
- (2) randomly select 50% segments as a training set
- (3) select the best order of the ARIMA model from training segments using BIC
- (4) select the best order of GARCH, GJR-GARCH, and EGARCH models from training segments using BIC
- (5) choose a model with a maximum likelihood value
- (6) use maximum likelihood estimation (MLE) to identify model coefficients
- (7) evaluate the proposed model using MSE, RMSE, MAE, and MAPE on the remaining segments

ALGORITHM 1: The proposed procedure for modeling apneic and normal ECG signals.

not exist. Besides, this statistical test is asymptotically distributed as chi-square. The final test form is based on the Wilcoxon signed-rank test, which is a nonparametric statistical test to assess if two populations are similar without assuming that they obey the normal distribution. The null hypothesis of the test is that the output indicators are equivalent or comparable to populations for GARCH models versus ECG segments. The test statistics, W , is the total of the positive difference ranks (i.e., $x - y$) between the two samples. We set 0.05 as the level of significance for the test. If the “ P value” is less than 0.05, we can conclude that the significance level of the null hypothesis is violated. This test is given as follows:

$$z(x, y) = \frac{(W - n(n+1)/4)}{\sqrt{(n(n+1)(2n+1) - \text{tiedj})/24}}, \quad (8)$$

where n corresponds to the sample size of the $x - y$. Signrank uses $[\text{tie}, \text{ank}, \text{tiedj}] = \text{tiedrank}(\text{abs}(\text{diffxy}), 0, 0, \text{epsdiff})$ to get the tie adjustment value tiedj for the two-sample event.

3.2. Order Selection. An ARIMA(p, d, q) model can be constructed by visually inspecting the autocorrelation function (ACF) and partial autocorrelation function (PACF). However, using objectively defined parameters such as Akaike information criteria (AIC) and Bayesian Information Criterion (BIC) is a more objective approach to determine p, q , and d of an ARIMA(p, d, q) method. These information criteria are statistical model fit measures [37]. They provided a set of results and assessed the relative fitness of the model of a number of previously developed mathematical models. Each of these criteria defines a $c_n(k)$ formula, where k denotes the number of model parameters and n the number of observations. The model with the fewest parameters, k , is called the best fit, and the quantity $c_n(k)$ is the smallest. The AIC [38] is an information processing method focused on the principle of entropy. The AIC’s main concept is to look at the model’s difficulty and its fit to the sample data and come up with a score that combines the two. Its formula is [37].

$$\text{AIC}: c_n(k) = 2 \cdot \left(\frac{k}{n} \right) - 2 \ln \frac{(L)}{n}, \quad (9)$$

where L denotes the likelihood function, k denotes the number of model parameters, and n denotes the number of observations.

Schwarz [39] provides the BIC:

$$\text{BIC}: c_n(k) = k \cdot \ln \frac{(n)}{n} - 2 \ln \frac{(L)}{n}. \quad (10)$$

Both of these criteria have advantages and disadvantages. Shibata [40] studied the asymptotic properties of the AIC estimation, concluding that the AIC estimate is inconsistent and asymptotically overestimates k with a nonzero probability. BIC is known to underestimate k [41]. Therefore, in the present study, we used the BIC criterion. The model with the lowest amount of BIC value is chosen as the most suitable match.

3.3. Evaluation Methods of Model Sufficiency. There is usually no common criterion for evaluating a model’s forecast output and comparing it with other benchmark models [42]. Since there are no common parameters for measuring errors, various error metrics were used to verify the proposed model’s efficacy. The model performance evaluation is done in this analysis by comparing the expected values with their corresponding observed values using traditional performance metrics, such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) based on the following equations:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2, \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2}, \quad (12)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |X_i - \hat{X}_i|, \quad (13)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{X_i - \hat{X}_i}{X_i} \right| \times 100\%, \quad (14)$$

where N is the number of samples in one segment and X_i, \hat{X}_i are observed and predicted values in one segment. These errors are advisable for predicting time series with the same scale and the same data processing procedures. The model with the smaller value of MSE, RMSE, MAE, and MAPE is selected as the best model. MAE and RMSE calculated the average of forecast errors over a sample size n . MAE and

RMSE have the analyzed signal units. MAPE, which is a dimensionless quantity, assesses the predictive model's accuracy. In statistics, MAPE calculates the precision of the prediction of a forecasting system and is typically expressed as a percentage. The predictive potential of the proposed ARIMA-EGARCH model was assessed by using equations (11)–(14).

4. Numerical Results

4.1. Data. In this study, the Physionet Apnea-ECG dataset (<https://www.physionet.org/physiobank/database/apnea-ecg/>) was used [43, 44] because of its availability and widespread use in the literature. Recordings were obtained from 32 people (25 men and 7 women). A total of 35 recordings sampled at 100 Hz from normal subjects and subjects with OSA were used. All signals were segmented into one-minute segments, and each segment was labeled as apnea or normal by physicians. Recordings varied in length from slightly less than 7 hours to almost 10 hours and were divided into three groups:

Apnea group: with 100 minutes or more of apnea, the mean age is 50 years. The range of age is 29–63 years.

Borderline group: with 10–96 minutes of apnea, the mean age is 46 years. The range of age is 39–53 years.

Healthy group: with 5 minutes or fewer of apnea, the mean age is 33 years. The range of age is 27–47 years.

4.2. Results. The first step in the modeling method is to approximate the mean of the data. In the literature, numerous mean equation models have been studied. Among these latest processes, which have been suggested, the ARIMA-type model was one of the most commonly used approaches in the literature due to its simplicity of execution and its well-known ability to predict and forecast. Therefore, this article applies the ARIMA model as the mean equation. ECG signals are nonstationary. To implement the aforementioned time series models on ECG signals, we must ensure that the time series is stationary. If the data are nonstationary, then the first difference is used to transform it. Plotting the first difference data will show whether the data have been converted into a stationary sequence. The second difference is taken if it is still not stationary. Model fittings can be made once the time series is stationarity. In the ARIMA model, this mechanism determines the differentiating parameter “ d .” As “ d ” is typically less than 2, we created the new time series by first and second differentiating the ECG signal. To find p , q , and d order, several combinations of ARIMA(p , d , q) are tested by the BIC, and the model with the smallest amount of BIC is selected. Results show that ARIMA(5, 2, 4) has the minimum amount of BIC for most segments. So, we consider ARIMA(5, 2, 4) for all segments.

Since the ECG signal is nonstationary, both the mean and percentiles of the data are different at varying periods. This means that the residual series differs over time, and the constant variance concept of the standard time series models

is broken. This further proves that the volatility model is essential. The GARCH models can only be used on volatile data. That is why a histogram is plotted to analyze whether or not ECG segments follow a normal distribution. Figure 1 shows the histogram for one ECG segment.

Then, kurtosis is determined. The minimum and the maximum kurtosis for all ECG segments are 5.1479 and 281.4540, respectively. It is evident from the minimum value that all measured kurtosis is higher than the value of three predicted for Gaussian distribution and histogram skewed to the right. Kurtosis values indicate that ECG segments have heavy tail characteristics and are not normally distributed. We also performed a Wilcoxon signed-rank test. Table 1 gives the “ h ,” “ P value,” and “Stat” for the test. “ $h = 1$ ” indicates a rejection of the null hypothesis, and “ $h = 0$ ” indicates a failure to reject the null hypothesis at the 5% significance level. “Stat” has information about the test statistic. The “ P values” of the statistical test for the ECG segments versus GARCH are more than 0.05, as shown in Table 1. This indicates that the ECG segments are statistically similar to GARCH models. We also used the ARCH/GARCH test suggested with Engle. Results of Engle’s test are shown in Table 2,

“ H ” is the Boolean decision variable. “1” suggests a null hypothesis rejection that there is no ARCH/GARCH effect. “Stat” displays ARCH/GARCH test statistics, and the “critical value” calculates the critical value of the chi-square distribution. If the “Stat” is below the “critical value” point, at a meaning level equal to 5%, there is no GARCH effect. However, if “Stat” is more than “critical value,” it formally shows clear evidence for GARCH in this time sequence. We applied the ARCH/GARCH test to all of the ECG signals in the databases. Because of the limited space, we demonstrate in this section the results of some representative signals. We should also remember that the simulated results are identical for different ECG signals. As shown in Table 2, for all signals, “ H ” is 1, and “Stat” is more than “critical value,” which means the null hypothesis is rejected, and therefore, the ECG signals have an ARCH/GARCH effect. Finally, visual validation is also performed between the histogram of ECG signals and the GARCH family models. Figure 2 shows ECG segment histograms and Gaussian distribution function with the estimated mean and standard deviation from the data and a corresponding GARCH model histogram.

From Figure 2, it is evident that there is high accuracy between the histogram of ECG segments and the GARCH model. Compared to Gaussian, the studied distribution is sharper and has a zero peak with a heavier tail. We also plot the cumulative distribution function (CDF) of ECG segments and the corresponding GARCH model in Figure 3.

It is evident from Figure 3 that ECG segments and the GARCH data are from the same CDF. Considering the results in Tables 1 and 2 and Figures 1–3, we find an ARCH/GARCH effect in all examined ECG segments. It should be again reported that the results of modeling various ECG signals are identical. However, only a few results are shown here. Hence, we proposed GARCH family models and demonstrated that they were a suitable representation of ECG segments.

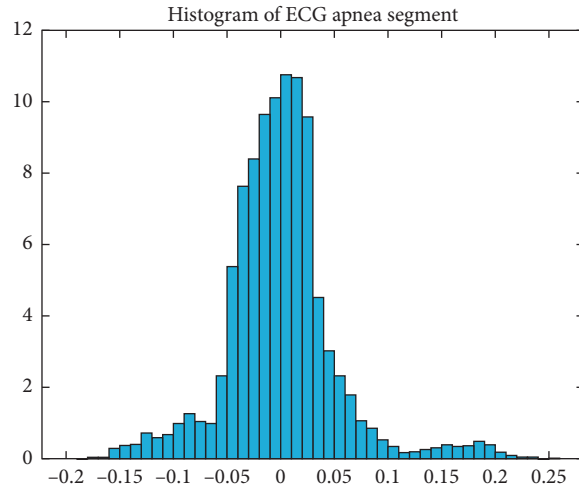


FIGURE 1: Histogram for an ECG segment.

TABLE 1: Results of the Wilcoxon signed-rank test.

Signal	H	Stat	P value
Signal 1	0	38	0.3223
Signal 2	0	41	0.1934
Signal 3	0	45	0.0840
Signal 4	0	38	0.3223
Signal 5	0	45	0.0840
Signal 6	0	37	0.3750
Signal 7	0	41	0.1934
Signal 8	0	41	0.1855

TABLE 2: Results of Engle's test for the existence of ARCH/GARCH effects.

Signal	H	Stat	Critical value
Signal 1	1	$2.6848e + 3$	3.8415
Signal 2	1	$3.1229e + 3$	3.8415
Signal 3	1	$3.1219e + 3$	3.8415
Signal 4	1	$3.2108e + 3$	3.8415
Signal 5	1	$3.1211e + 3$	3.8415
Signal 6	1	$2.8983e + 3$	3.8415
Signal 7	1	$3.2584e + 3$	3.8415
Signal 8	1	$3.2659e + 3$	3.8415

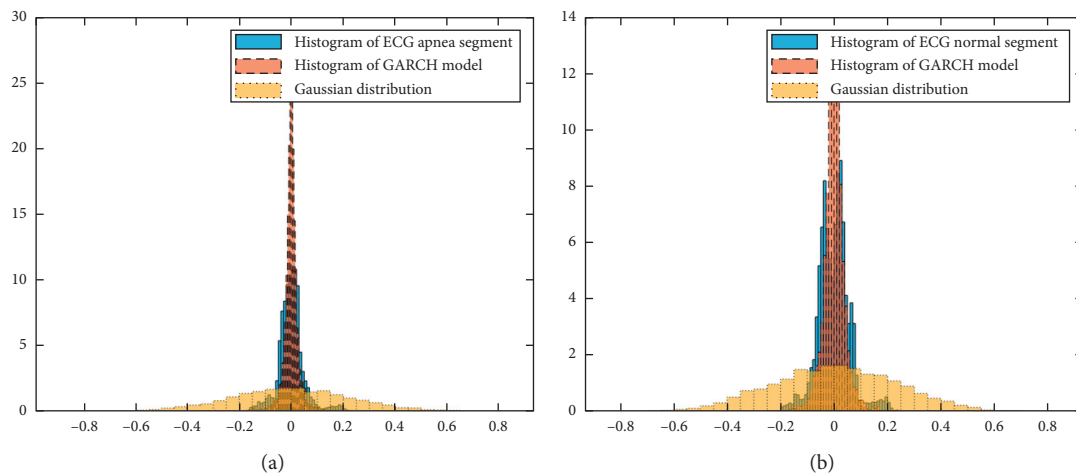


FIGURE 2: Histograms of ECG segment, corresponding GARCH model, and Gaussian distribution of (a) an apneic segment and (b) a normal segment.

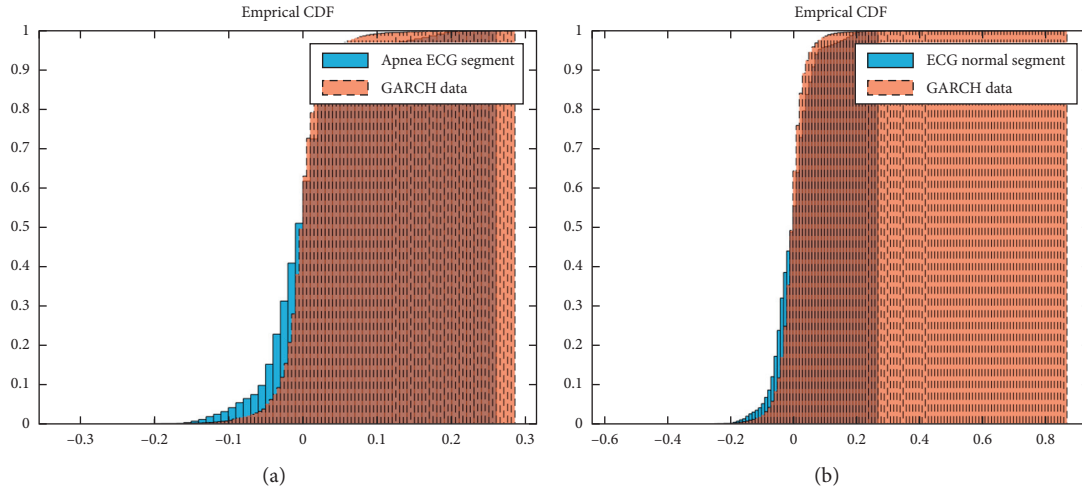


FIGURE 3: Comparison of cumulative distribution function (CDF) of ECG segments (solid line) and CDF of GARCH data (dash line) for (a) apneic and (b) normal segments.

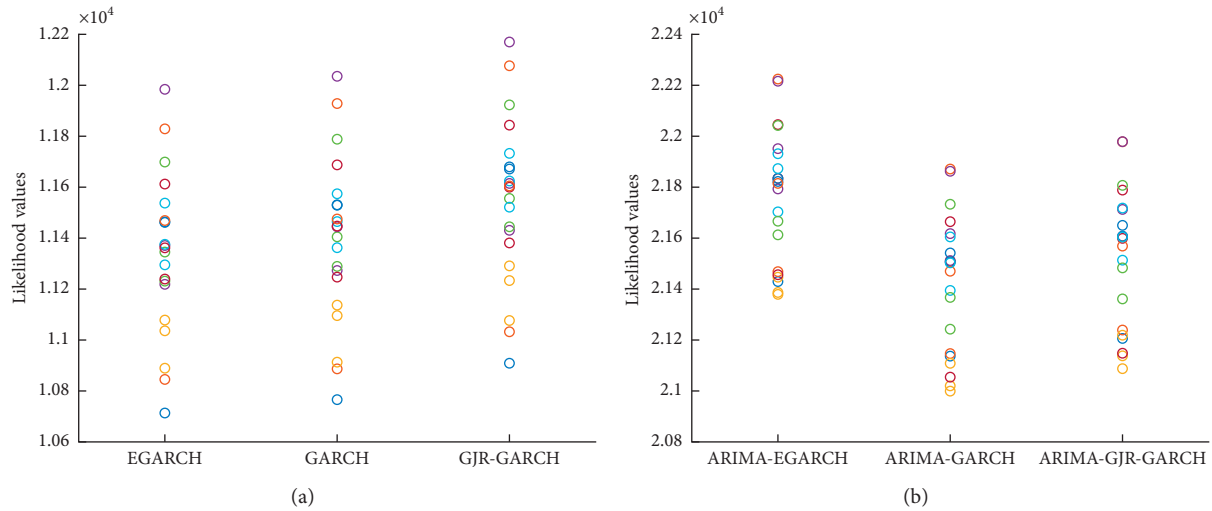


FIGURE 4: The likelihood values of (a) GARCH, EGARCH, and GJR-GARCH models; (b) ARIMA-GARCH, ARIMA-EGARCH, and ARIMA-GJR-GARCH models.

The literature is comprehensive on the GARCH family models. However, we limit our study to the three more common models for compactness, namely, GARCH, EGARCH, and GJR-GARCH. Estimating the volatility model requires order selection and parameter estimation, similar to the ARIMA model. Again, to find the proper order of GARCH, EGARCH, and GJR-GARCH models, several combinations were tested by the BIC, and the models with the smallest amount of BIC were selected.

Results show that GARCH(1,4), GJR-GARCH(1,5), and EGARCH(1,5) have the minimum BIC for most segments. So, we consider GARCH(1,4), GJR-GARCH(1,5), and EGARCH(1,5) for all segments. We calculated the log-likelihood amount of GARCH, GJR-GARCH, and EGARCH to find the best model among the others. We also computed log-likelihood for ARIMA-GARCH, ARIMA-GJR-GARCH, and ARIMA-EGARCH. The model with the maximum amount of log-likelihood was selected as a proper model.

The results are illustrated in Figure 4. As is evident from Figure 4, GJR-GARCH and ARIMA-EGARCH have a maximum amount of log-likelihood. Finally, in each segment, the maximum likelihood estimation (MLE) was used to identify model coefficients. MLE is applied to both ARIMA and GARCH models.

In the next step, we validated our proposed method on the remaining ECG segments, which were not used in the model estimation step. We run ARIMA and GJR-GARCH models on the Physionet Apnea-ECG database and compared the ARIMA-EGARCH model results with these models using MSE, RMSE, MAE, and MAPE. The reason for choosing these models is that linear models such as AR, MA, and ARMA are coarse estimations of real-world systems and usually have poor performance in forecasting the evolution of nonstationary and nonlinear processes; ARIMA model is usually used to model these patterns [31]. So, we considered the ARIMA model as the first model for comparison. On the

TABLE 3: Comparison of MSEs, RMSEs, MAEs, and MAPEs from different models estimated for apnea segments.

	ARIMA		GJR-GARCH		ARIMA-EGARCH	
	Ave	Std	Ave	Std	Ave	Std
MSE	0.0181	0.0133	0.2032	0.7924	0.0177	0.0129
RMSE	0.1232	0.0543	0.2990	0.3378	0.1216	0.0535
MAE	0.0666	0.0282	0.0736	0.0140	0.0613	0.0255
MAPE	13.6081%	42.5486%	42.3760%	678.0458%	4.0841%	15.9423%

Ave = average, Std = standard deviation.

TABLE 4: Comparison of MSEs, RMSEs, MAEs, and MAPEs from different models estimated for normal segments.

	ARIMA		GJR-GARCH		ARIMA-EGARCH	
	Ave	Std	Ave	Std	Ave	Std
MSE	0.0235	0.0131	0.6792	4.0802	0.0230	0.0128
RMSE	0.1454	0.0490	0.4477	0.6928	0.1437	0.0484
MAE	0.0810	0.0250	0.0962	0.0526	0.0754	0.0226
MAPE	16.1331%	97.8801%	18.5205%	161.6340%	4.7286%	29.5338%

Ave = average, Std = standard deviation.

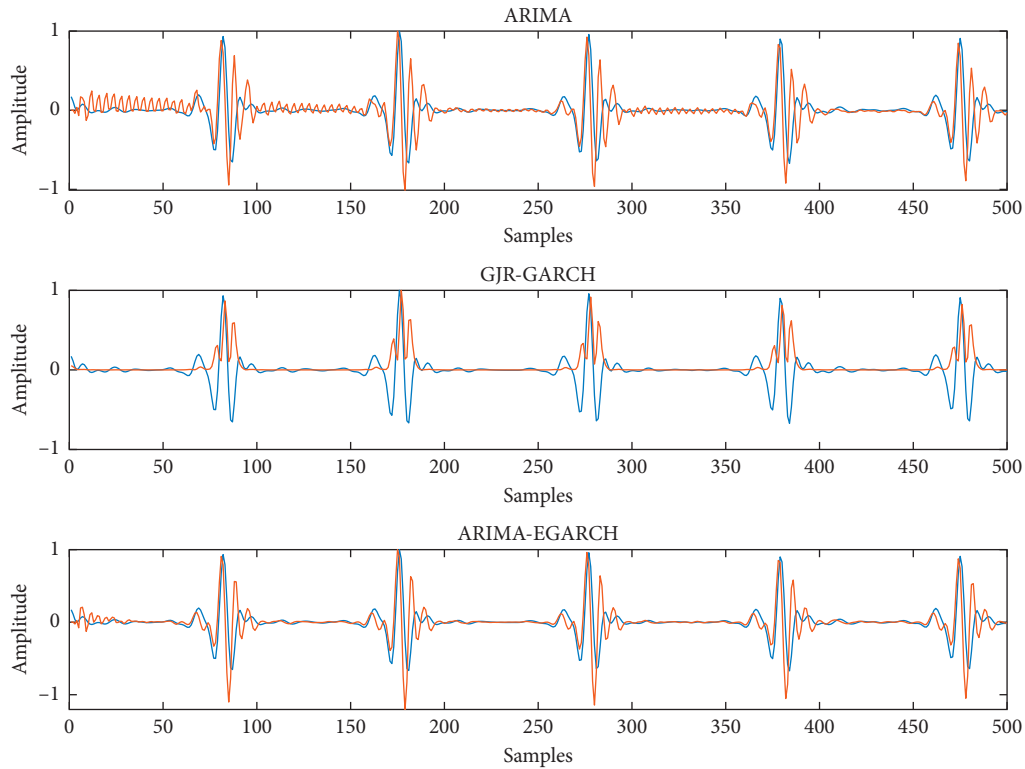


FIGURE 5: Fitting results using ARIMA, GJR-GARCH, and ARIMA-EGARCH models on a sample apneic ECG segment (blue line: ECG data; redline: estimated data).

other hand, we showed that ECG signals were heteroskedastic; GARCH family models can be used to model them. Since the log-likelihood of GJR-GARCH had maximum value, we selected GJR-GARCH as the second model. Average and standard deviation of MSE, RMSE, MAE, and MAPE based on the assessment criteria 11, 12, 13, and 14 for ARIMA, GJR-GARCH, and ARIMA-EGARCH models on apneic and normal ECG segments are presented in Tables 3 and 4, respectively.

Tables 3 and 4 demonstrate that in terms of averaged MSE, RMSE, MAE, and MAPE, the proposed ARIMA-EGARCH model outperforms all other models in modeling both apnea and normal ECG signals. The lowest prediction error value reflects the superiority of the proposed ARIMA-EGARCH model over the ARIMA model and the GJR-GARCH model.

The graphical validation of our model on the sample test data is illustrated in Figures 5 and 6 for sample apneic and normal ECG segments using three different models,

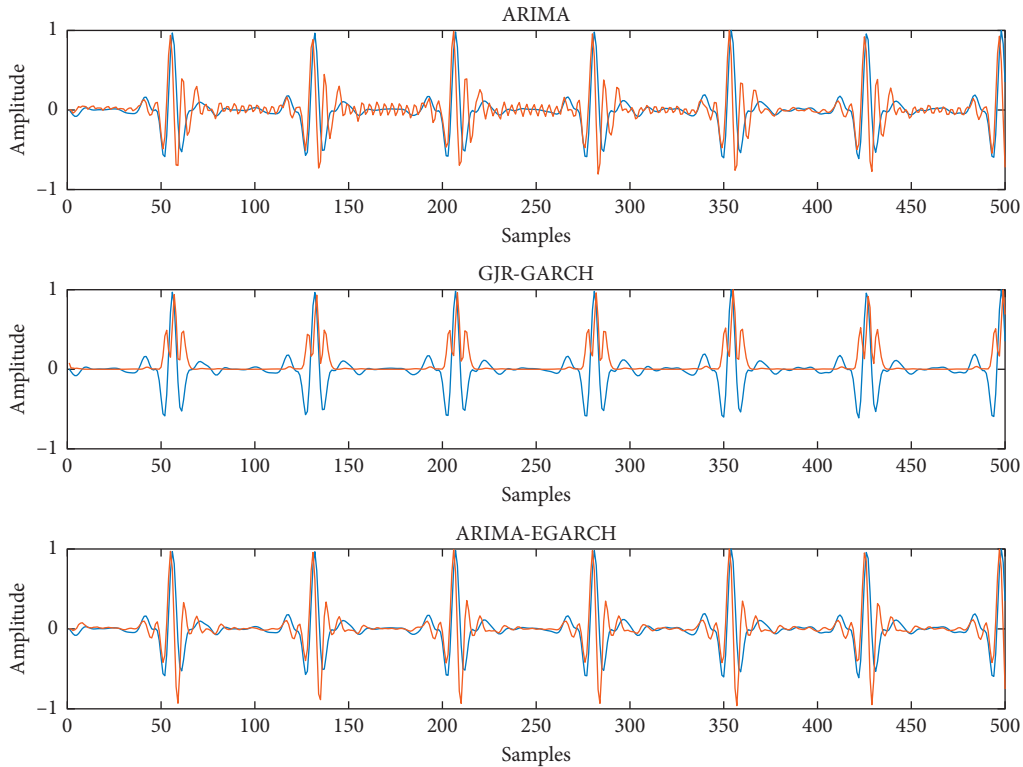


FIGURE 6: Fitting results using ARIMA, GJR-GARCH, and ARIMA-EGARCH models on normal ECG data (blue line: ECG data; redline: estimated data).

TABLE 5: Comparison of RMSEs and MAEs from ARIMA-EGARCH with other existing models for apnea segments.

Method	RMSE	MAE
ARCH(1) [30]	0.6614	0.5876
GARCH(1,1) [30]	0.6636	0.5808
Integrated GARCH(1,1) [30]	0.6325	0.5725
Student-t GARCH(1,1) [30]	0.6497	0.5625
Student-t Integrated GARCH(1,1) [30]	0.6551	0.5638
ARIMA(5,2,4)-EGARCH(1,5)	0.1216	0.0613

respectively. For better understanding, we only show the 500 first samples of each segment in figures.

Figures 5 and 6 show that the ARIMA-EGARCH model can best model both sudden and slow transients in apneic and normal ECG signals. Moreover, if we only consider slow changes in the ECG signal, we can see that a linear approach such as ARIMA cannot predict slow changes, but a nonlinear method like GJR-GARCH can predict almost slow changes (white noise). These findings confirm our claim that EGARCH models improve the estimation made by ARIMA and a combination of ARIMA and EGARCH models can complete each other in ECG signal modeling.

4.3. Comparison with Other Models. In this section, the performance of the selected model is compared with other models proposed in the literature. Since the results obtained from the same sample must be compared for meaningful comparison, we must compare our results with the results of models that used the Physionet Apnea-ECG. As mentioned

TABLE 6: Comparison of RMSEs and MAEs from ARIMA-EGARCH with other existing models for normal segments.

Method	RMSE	MAE
ARCH(1) [30]	0.6482	0.5652
GARCH(1,1) [30]	0.6508	0.5618
Integrated GARCH(1,1) [30]	0.7133	0.5843
Student-t GARCH(1,1) [30]	0.7347	0.5785
Student-t Integrated GARCH(1,1) [30]	0.7212	0.5841
ARIMA(5,2,4)-EGARCH(1,5)	0.1437	0.0754

in the Introduction, only a few researchers used models to detect apnea [26–28], and these articles fit models to QRS complexes and not the entire ECG. Therefore, we can not compare our results with them. On the other hand, these articles used the model’s parameters to detect sleep apnea, and they did not compare the estimated results with the actual ECG signal. Only Hu et al. [30] proposed a mathematical model for apneic and normal ECG signals. Therefore, to show our model’s capability, we compared our results with this article. Tables 5 and 6 compare RMSEs and MAEs from ARIMA-EGARCH with ARCH(1) [30], GARCH(1,1) [30], Student-t GARCH(1,1) [30], Integrated GARCH(1,1) [30], and Student-t Integrated GARCH(1,1) [30] for apneic and normal segments, respectively.

5. Discussion and Conclusion

This article describes a method for mathematical modeling of the ECG signal. Although the forecasting of time series is a

vast research area, it can be classified into short- and long-term predictions. Short-term forecasting can also be split into mathematical modeling and physiological modeling. The mathematical-based forecasting model uses mathematical representation and dynamic variations to predict the future status of the time series of the process that underlies it. It should be clear that no apnea detection method is provided in this article. We have shown that ECG signals are heteroskedastic, which means the conditional variance is not constant. In the literature, ECG volatility and heteroskedasticity during apnea are underestimated. Thus, we used this characteristic to model ECG signals.

The proposed ARIMA-EGARCH can model the mean and volatility of ECG signals in sleep apnea syndrome. This model can cover both linear and nonlinear characteristics of ECG signals. Using BIC, the best orders of the ARIMA and EGARCH models were estimated. The model parameters were approximated using the maximum likelihood estimation method. Finally, some metrics, including MSE, RMSE, MAE, and MAPE, between the actual and estimated ECG signals were calculated. The method is validated and compared to other methods, using recordings from the Physionet Apnea-ECG database containing ECG segments during sleep apnea and normal breathing. Visual quality and objective quality of the proposed approach were achieved in terms of MSE, RMSE, MAE, and MAPE. Since the results obtained from the same sample must be compared for meaningful comparison, we must compare our results with the results of models that used the Physionet Apnea-ECG. So, we compared our proposed model with models in [30] (see Tables 4 and 5). As it can be inferred from Tables 4 and 5, the proposed ARIMA-EGARCH model outperforms the other existing models for sleep apnea modeling. Experimental findings show that the ARIMA-EGARCH model can estimate both normal and apneic ECG signals. Our results are robust for selecting performance assessment criteria.

The proposed model has some advantages. The estimated model's parameters can be used as features for the automatic detection of sleep apnea [45]. Moreover, one of the current widespread therapies in sleep apnea is continuous positive airway pressure (CPAP) that blows constant air at a fixed pressure. ECG model can be used in CPAP machines in order to blow air only when apnea occurs.

Data Availability

In this study, the Physionet Apnea-ECG dataset is available at <https://www.physionet.org/physiobank/database/apnea-ecg/> was used.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Y. Isler, A. Narin, M. Ozer, and M. Perc, "Multi-stage classification of congestive heart failure based on short-term heart rate variability," *Chaos, Solitons & Fractals*, vol. 118, pp. 145–151, 2019.
- [2] F. Sessa, V. Anna, G. Messina et al., "Heart rate variability as predictive factor for sudden cardiac death," *Aging*, vol. 10, no. 2, pp. 166–177, 2018.
- [3] M. Ataei, S. Naghsh, M. Yazdchi, and M. Hashemi, "Chaos-based analysis of heart rate variability time series in obstructive sleep apnea subjects," *Journal of Medical Signals & Sensors*, vol. 10, no. 1, p. 53, 2020.
- [4] R. B. Berry, R. Budhiraja, D. J. Gottlieb et al., "Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. deliberations of the sleep apnea definitions task force of the American academy of sleep medicine," *Journal of clinical sleep medicine*, vol. 8, no. 5, pp. 597–619, 2012.
- [5] A. H. Yüzer, H. Sümbül, M. Nour, and K. Polat, "A different sleep apnea classification system with neural network based on the acceleration signals," *Applied Acoustics*, vol. 163, Article ID 107225, 2020.
- [6] E. Braunwald, A. S. Fauci, D. L. Kasper, S. L. Hauser, D. L. Longo, and J. L. Jameson, *Harrison's Principles of Internal Medicine*, McGraw-Hill, New York, NY, US, 2001.
- [7] P. Leelaarporn, P. Wachiraphan, T. Kaewlee et al., "Sensor-driven achieving of smart living: a review," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10369–10391, 2021.
- [8] H. Singh, R. K. Tripathy, and R. B. Pachori, "Detection of sleep apnea from heart beat interval and ecg derived respiration signals using sliding mode singular spectrum analysis," *Digital Signal Processing*, vol. 104, Article ID 102796, 2020.
- [9] K. N. V. P. S. Rajesh, R. Dhuli, and T. S. Kumar, "Obstructive sleep apnea detection using discrete wavelet transform-based statistical features," *Computers in Biology and Medicine*, vol. 130, Article ID 104199, 2021.
- [10] A. Zarei and B. Mohammadzadeh Asl, "Performance evaluation of the spectral autocorrelation function and autoregressive models for automated sleep apnea detection using single-lead ecg signal," *Computer Methods and Programs in Biomedicine*, vol. 195, Article ID 105626, 2020.
- [11] J. Zhang, Z. Tang, J. Gao et al., "Automatic detection of obstructive sleep apnea events using a deep cnn-lstm model," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 5594733, 2021.
- [12] N. Banluesombatkul, T. Rakthanmanon, and T. Wilaiprasitporn, "Single channel ecg for obstructive sleep apnea severity detection using a deep learning approach," in *Proceedings of the TENCON 2018-2018 IEEE Region 10 Conference*, pp. 2011–2016, Jeju, Korea, October 2018.
- [13] O. Sayadi and M. B. Shamsollahi, "Life-threatening arrhythmia verification in icu patients using the joint cardiovascular dynamical model and a bayesian filter," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2748–2757, 2011.
- [14] O. Sayadi, M. B. Shamsollahi, and G. D. Clifford, "Synthetic ecg generation and bayesian filtering using a Gaussian wave-based dynamical model," *Physiological Measurement*, vol. 31, no. 10, pp. 1309–1329, 2010.
- [15] K. Padmavathi and K. S. Ramakrishna, "Detection of atrial fibrillation using autoregressive modeling," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 5, no. 1, pp. 64–70, 2015.
- [16] J. Aguila, E. Arias, M. Artigao, and J. Miralles, "A prediction of electrocardiography signals by combining arma model with nonlinear analysis methods," *Recent Researches in Applied Computer and Applied Computational Science*, vol. 35, pp. 31–37, 2011.

- [17] O. Raach, T. R. Pillai, and A. Abdullah, "Gamma modeling of ecg and classification of arrhythmia," in *Proceedings of the 2018 8th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, pp. 26–31, Kuala Lumpur, Malaysia, May 2018.
- [18] Z. Zheyang Li and M. Minjie Ma, "Ecg modeling with dfg," in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 2691–2694, Shanghai, China, February 2005.
- [19] R. Dubois, P. Roussel, M. Vaglio et al., "Efficient modeling of ecg waves for morphology tracking," in *Proceedings of the 2009 36th Annual Computers in Cardiology Conference (CinC)*, pp. 313–316, September 2009.
- [20] R. V. Andreato, B. Dorizzi, and J. Boudy, "Ecg signal analysis through hidden markov models," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 8, pp. 1541–1549, 2006.
- [21] E. K. Roonizi and R. Sameni, "Morphological modeling of cardiac signals based on signal decomposition," *Computers in Biology and Medicine*, vol. 43, no. 10, pp. 1453–1461, 2013.
- [22] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 4, pp. 582–589, 2004.
- [23] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, 2003.
- [24] M. Chawla, H. Verma, and V. Kumar, "Ecg modeling and qrs detection using principal component analysisdoi," in *Proceedings of the Advances in Medical, Signal and Information Processing, 2006. IET 3rd International Conference On*, Dublin, Ireland, April 2006.
- [25] J. McNames and M. Aboy, "Statistical modeling of cardiovascular signals and parameter estimation based on the extended kalman filter," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 119–129, 2008.
- [26] M. O. Mendez, A. M. Bianchi, M. Matteucci, S. Cerutti, and T. Penzel, "Sleep apnea screening by autoregressive models from a single ecg lead," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 12, pp. 2838–2850, 2009.
- [27] H. Sharma and K. K. Sharma, "An algorithm for sleep apnea detection from single-lead ecg using hermite basis functions," *Computers in Biology and Medicine*, vol. 77, pp. 116–124, 2016.
- [28] A. R. Hassan, "Computer-aided obstructive sleep apnea detection using normal inverse Gaussian parameters and adaptive boosting," *Biomedical Signal Processing and Control*, vol. 29, pp. 22–30, 2016.
- [29] P. Sun and C. Zhou, "Diagnosing the distribution of garch innovations," *Journal of Empirical Finance*, vol. 29, pp. 287–303, 2014.
- [30] M. Y. Hu and C. Tsoukalas, "Conditional volatility properties of sleep-disordered breathing," *Computers in Biology and Medicine*, vol. 36, no. 3, pp. 303–312, 2006.
- [31] C. Cheng, A. Sa-Ngasoongsong, O. Beyca et al., "Time series forecasting for nonlinear and non-stationary processes: a review and comparative study," *IIE Transactions*, vol. 47, no. 10, pp. 1053–1071, 2015.
- [32] S. Singh and A. Mohapatra, "Repeated wavelet transform based arima model for very short-term wind speed forecasting," *Renewable Energy*, vol. 136, pp. 758–768, 2019.
- [33] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [34] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [35] D. B. Nelson, "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991.
- [36] L. R. Glosten, R. Jagannathan, and D. E. Runkle, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The Journal of Finance*, vol. 48, no. 5, pp. 1779–1801, 1993.
- [37] U. Kumar and V. K. Jain, "ARIMA forecasting of ambient air pollutants (O₃, NO, NO₂ and CO)," *Stochastic Environmental Research and Risk Assessment*, vol. 24, no. 5, pp. 751–760, 2010.
- [38] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike* Springer, Berlin, Germany, 1998.
- [39] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [40] R. Shibata, "Selection of the order of an autoregressive model by Akaike's information criterion," *Biometrika*, vol. 63, no. 1, pp. 117–126, 1976.
- [41] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society: Series B*, vol. 41, no. 2, pp. 190–195, 1979.
- [42] Y. Xu, W. Yang, and J. Wang, "Air quality early-warning system for cities in China," *Atmospheric Environment*, vol. 148, pp. 239–257, 2017.
- [43] T. Penzel, G. B. Moody, R. G. Mark, A. L. Goldberger, and J. H. Peter, "The apnea-ecg database," *Computers in Cardiology 2000*, vol. 27, pp. 255–258, 2000.
- [44] A. L. Goldberger, L. A. Amaral, L. Glass et al., "Physiobank, physiotookit, and physionet: components of a new research resource for complex physiological signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [45] M. Faal and F. Almasganj, "Obstructive sleep apnea screening from unprocessed ecg signals using statistical modelling," *Biomedical Signal Processing and Control*, vol. 68, Article ID 102685, 2021.

Research Article

EEG-Based Closed-Loop Neurofeedback for Attention Monitoring and Training in Young Adults

Bingbing Wang , Zeju Xu , Tong Luo , and Jiahui Pan 

School of Software, South China Normal University, Guangzhou 510631, China

Correspondence should be addressed to Jiahui Pan; panjh82@qq.com

Received 26 February 2021; Revised 2 June 2021; Accepted 7 June 2021; Published 15 June 2021

Academic Editor: Siti Anom Ahmad

Copyright © 2021 Bingbing 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.

Attention is an important mechanism for young adults, whose lives largely involve interacting with media and performing technology multitasking. Nevertheless, the existing studies related to attention are characterized by low accuracy and poor attention levels in terms of attention monitoring and inefficiency during attention training. In this paper, we propose an improved random forest- (IRF-) algorithm-based attention monitoring and training method with closed-loop neurofeedback. For attention monitoring, an IRF classifier that uses grid search optimization and multiple cross-validation to improve monitoring accuracy and performance is utilized, and five attention levels are proposed. For attention training, we develop three training modes with neurofeedback corresponding to sustained attention, selective attention, and focus attention and apply a self-control method with four indicators to validate the resulting training effect. An offline experiment based on the Personal EEG Concentration Tasks dataset and an online experiment involving 10 young adults are conducted. The results show that our proposed IRF-algorithm-based attention monitoring approach achieves an average accuracy of 79.34%, thereby outperforming the current state-of-the-art algorithms. Furthermore, when excluding familiarity with the game environment, statistically significant performance improvements ($p < 0.05$) are achieved by the 10 young adults after attention training, which demonstrates the effectiveness of the proposed serious games. Our work involving the proposed method of attention monitoring and training proves to be reliable and efficient.

1. Introduction

Attention can be characterized as a cognitive process in the brain that selectively focuses on some part of the available information [1]. Nevertheless, excessive media multitasking poses a serious issue with respect to the attention function of young adults, resulting in distraction and poor attention control. Therefore, there is an increasing demand for attention monitoring and training for young adults [2].

Numerous methods have been developed for attention training, meditation [3], and computer-based exercises [4], but these approaches may contribute to mental fatigue. In recent years, several researchers have explored methods with fewer side effects. For example, Putri et al. [5] proposed the method of regular high-intensity circuit training (HICT), which can improve attention function in young male adults.

In the same year, Luo and Zhang [2] conducted experiments to validate that noninvasive tactile training has an excellent effect on sustained attention in young adults. The main purpose of this paper is to investigate a method of attention monitoring and training based on closed-loop neurofeedback. We use brain-computer interface (BCI) technology, which utilizes recorded brain activity, primarily measured by electroencephalography (EEG), to execute communications between the brain and computers to manipulate the environment in a manner that is compatible with the intentions of humans [6]. Notably, EEG signals are the most frequently used. In contrast to the previously developed methods in [3, 4, 7], an EEG-based system can be used as a noninvasive neurofeedback platform to enhance individual attention and cognitive abilities [8]. Furthermore, a neurofeedback-based attention training system not only helps young adults but

also is suitable for children, especially those with attention deficit hyperactivity disorder (ADHD) [9].

BCI-based technology can be used to identify subtle shifts in individual attention [10]. Chiang et al. [11] developed an attention monitoring technique that integrates the minimum entropy principle approach (MEPA) and an associative Petri network (APN). Using a 14-electrode EEG device, a two-class classification accuracy of 90.4% was achieved based on 10 subjects in an online experiment. Hu et al. [12] compared the correlation-based feature selection (CFS) algorithm with other classification algorithms to evaluate attention at 3 levels (high, neutral, and low) and concluded that combining CFS and the K-nearest neighbors (KNN) data mining algorithm, which was used with a single valence resulted in the best performance. Six electrodes (C3, C4, Cz, P3, P4, and Pz) were used, and 10 subjects achieved an accuracy of 80.84% in their online experiment. In another study, Mohammadpour and Mozaffari [1] adopted an artificial neural network (ANN) to classify attention into four levels by using EEG signals from Fp1, Fp2, F3, F4, F7, F8, and Fz electrodes. An online experiment involving 5 subjects was conducted and an average accuracy of 79.75% was obtained. However, there have been two major issues to be further investigated. One issue is that the accuracy of attention monitoring needs to be improved, which may be related to the numbers of attention levels, detection algorithms, and types of feedback. The other issue is that more attention levels (i.e., ≥ 5) and fewer electrodes (i.e., ≤ 4) need to be explored in online attention monitoring, especially for EEG-based practical application.

Neurofeedback is an effective training technique based on brain waves and computer processing [13], and EEG-based neurofeedback training can provide real-time information to individuals regarding their brain function through BCI devices. Bettencourt et al. [14] used closed-loop neurofeedback from multivariate pattern analysis (MVPA) as a type of cognitive prosthetic to provide a neural error signal so that individuals could learn to properly evaluate the state of their attention. Mohammadi et al. [13] designed a computer game to train individual attention based on neurofeedback, and they summarized that the neurofeedback game not only helps individuals increase the possibility of success in controlling their attention but also decreases the time required for the training process. Although EEG-based neurofeedback training plays an important role in attention improvement, one should note that the above-mentioned training methods are not suitable for all the different mechanisms of attention, such as sustained attention, selective attention, and focus attention.

To address these above issues, we focus on EEG-based attention monitoring and training with closed-loop neurofeedback in this study. On one hand, we propose an improved random forest- (IRF-) algorithm-based monitoring method, which uses grid search optimization and multiple cross-validation to classify attention into five levels. The Personal EEG Concentration Tasks dataset involving 80 subjects was used to verify the effectiveness of the proposed attention monitoring method in the offline analysis. On the other hand, along with closed-loop neurofeedback, we

provide three serious game-type training modes based on sustained attention, selective attention, and focus attention, which might be promising in terms of self-regulated attention training. Four primary indicators, including the Schulte times, win times, game scores, and skill times, were evaluated in an online experiment.

The rest of this paper is organized as follows: Section 2 offers additional details on various methods, such as EEG data processing, classification algorithms, and game design. The succeeding section illustrates the process of the experiments and analyzes the results, followed by a discussion and our conclusions.

2. Materials and Methods

This section provides an overview of the utilized methods, as shown in Figure 1, which are separated into two modules: EEG-based attention monitoring and EEG-based training. The first module commences with attention monitoring workflows based on EEG signals. Then, EEG data processing and feature extraction are performed. Finally, the output classification obtained based on the IRF algorithm is presented. In the second module, we illustrate the principle of attention training and describe the implementation of serious games with closed-loop neurofeedback.

2.1. EEG-Based Attention Monitoring. In the attention monitoring module, an OpenBCI headset with 8 channels was used to collect EEG signals, and a wavelet transform algorithm was used to analyze and extract features for the preprocessed EEG data. Then, we utilized the IRF algorithm to classify attention.

2.1.1. EEG Data Preprocessing and Feature Extraction. Previous related studies have shown that the power spectral densities (PSDs) of delta, theta, alpha, beta, and gamma have certain correlations with human attention. To this end, we selected and extracted EEG features based on these findings. An OpenBCI headset was used to capture the EEG data. Additionally, in terms of EEG data processing and attention monitoring, there are two factors to be considered. First, the most active sites of attention need to be given priority because the response is not evenly distributed across the electrodes. Second, if the algorithm is time-consuming, the time delays will not conform to the real-time constraints, rendering the feedback meaningless. Given the complexity of all kinds of data processing and the activity of positions, this paper selected the relatively active and attention-relative channels located in the frontal and temporal lobes, that is, TP9, AF7, AF8, and TP10, following the research of Castillo et al. [15] and Taillez et al. [16]. The wavelet transform algorithm was then applied to extract the PSD features of the EEG signals.

Wavelet analysis involves a combination of the time domain and frequency domain and is suitable for multiscale time-frequency analyses. The wavelet basis is defined as follows:

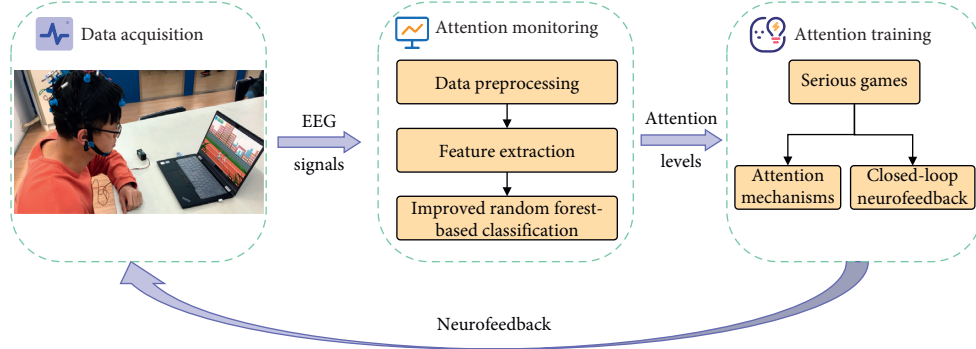


FIGURE 1: Architecture of the attention monitoring and training modules.

$$\psi_{s,a}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-s}{a}\right). \quad (1)$$

In (1), $\psi_{s,a}(t)$ represents the displacement and scale expansion of the basic wavelet, which can be used to decompose signals at different times. s denotes the translation factor, a indicates the scale parameter, and $1/\sqrt{a}$ is the normalization factor, which is proportional to $\psi_{s,a}(t)$:

$$\text{CWT}(s, a) = \int_{-\infty}^{\infty} f(t) \psi_{s,a}(t) dt, \quad (2)$$

$$f(t) = \frac{1}{C_\psi} \int \int_{-\infty+\infty}^{+\infty+\infty} \text{CWT}(s, a) \psi_{s,a}(t) \frac{dads}{da^2}. \quad (3)$$

In equation (2), a time-scale planar function $\text{CWT}(s, a)$ mapped by the signal $f(t)$ is shown. $\text{CWT}(s, a)$ represents a one-dimensional continuous wavelet transform. $C_\psi = \int_{-\infty}^{+\infty} (|\psi(u)|^2 / |u| du)$, and $\psi(u)$ is the Fourier transform (FT) of $\psi(t)$.

The continuous wavelet transform (CWT) is used for extracting the PSD features of EEG signals. In this study, we used Daubechies coefficients for the wavelet transform, as they are characterized by excellent time localization performance and a maximal number of vanishing moments for a given support set. Daubechies 4 enjoys compact support and an orthonormal wavelet with smoothness. Thus, an improved effect can be achieved through the analysis of nonstationary EEG signals. It is necessary to select a suitable number of decomposition levels to analyze an EEG signal; thus, we selected a decomposition level of 5 ($L=5$). In addition, the sampling frequency was 256 Hz, and the band-limited EEG was then subjected to a five-level decomposition coefficient of six subband signals through CWT. As shown in Figure 2, six subbands, including xD1, xD2, xD3, xD4, xD5, and xA1, represented the frequency range of the band-limited EEG signal [17], where xA is the decomposition approximation coefficient and xD is the decomposition detail coefficient.

Four wavelet thresholding methods were used in [18] to select an accurate threshold. We adopted the SURE threshold, which is an adaptive soft thresholding method. Once the threshold coefficients were extracted from each level, the effect of the noise on the EEG signals was removed.

We then used the inverse CWT to reconstruct the signals at each level.

The first reconstructed detail D1 was regarded as the noise component of the EEG signal, and the reconstruction details of the other four subband signals D2–D5 and the reconstruction approximation of the subband signal A5 yielded signal information relevant to each EEG frequency band. Furthermore, 5 PSD features were extracted for classification: the delta ($0 \text{ Hz} < f < 4 \text{ Hz}$), theta ($4 \text{ Hz} < f < 8 \text{ Hz}$), alpha ($8 \text{ Hz} < f < 16 \text{ Hz}$), beta ($16 \text{ Hz} < f < 32 \text{ Hz}$), and gamma ($32 \text{ Hz} < f < 64 \text{ Hz}$) bands. There were a total of 20 EEG features ($4 \times 5 = 20$).

2.1.2. Improved Random Forest Classifier. Compared with the correlation classification method, random forest classification removes noise more effectively and accurately, which contributes to the higher accuracy during the classification of noise-containing EEG signals. In addition, the random forest method offers stability, running efficiency, and reducing errors for imbalanced datasets. On this foundation, Belle et al. [19] compared the random forest and regression techniques for attention classification based on EEG signals, determining that random forest seems to work best for both modalities, which obtained an average accuracy of 85.7% for EEG. Thus, we choose the random forest algorithm to classify the attention level and propose the IRF method with higher accuracy.

The workflows of the IRF algorithm used for attention monitoring are shown in Figure 3. A random forest is a set of multiple decision tree classifiers $\{h(x, \phi_k), k = 1, \dots\}$, and the parameter set $\{\phi_k\}$ is an independent and identically distributed random vector. The input feature variables X are classified separately by each decision tree, and the results are relied on to make predictions. After that, the classification results with the most votes are attained as the output.

(1) *Decision Tree.* The random forest algorithm takes multiple samples from the original data through bootstrap resampling and generates multiple decision tree classification models. Three steps are involved in the establishment of a decision tree.

Step 1. Select a random bootstrap sample across N original training sets by using the sampling with the replacement method, and repeat k times.

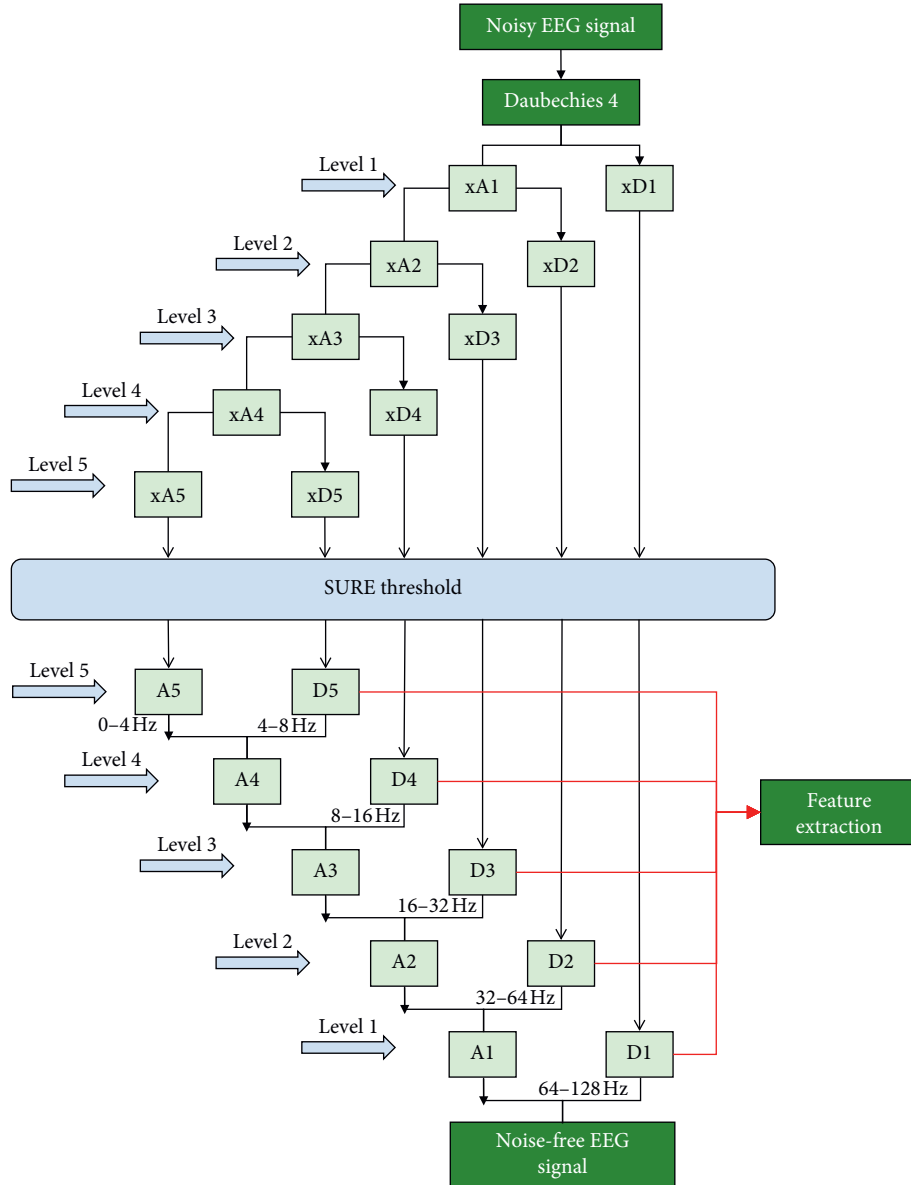


FIGURE 2: Wavelet multiresolution analysis.

Step 2. Train a decision tree with a training set including each bootstrap sample and recognize it as the root node of the sample. When each node is split, the feature variables $x (x \ll X)$ are extracted from the X total feature variables at random for calculation purposes, and the best feature obtained from the x feature variables is selected as a branch of the node to achieve minimal node impurity.

Step 3. Split each node as before without a pruning operation in the course of establishing the decision tree.

To achieve stable accuracy, two random factors are introduced during the establishment of the decision tree. One is the bootstrap samples drawn from the N original training sets. The other is the stochastic feature variable selected from the node of the decision tree.

(2) *Voting*. To enhance the mutual influence between the classification models and improve their prediction ability, diverse decision trees are constructed by using different samples. After k rounds of training are conducted, the optimal classification models $\{h_1(X), h_2(X), \dots, h_k(X)\}$ are obtained and combined in a sequence to acquire the ultimate classification results by using the simple majority voting method. Equation (4) presents the classification decision:

$$H(x) = \arg \max_Y \sum_{i=1}^k F(h_i(x) = Y), \quad (4)$$

where $H(x)$ is the classification model after the combination and $h_i(x)$ is one of the decision tree classification models. Y is the target variable, and the characteristic function is $F(h_i(x) = Y)$.

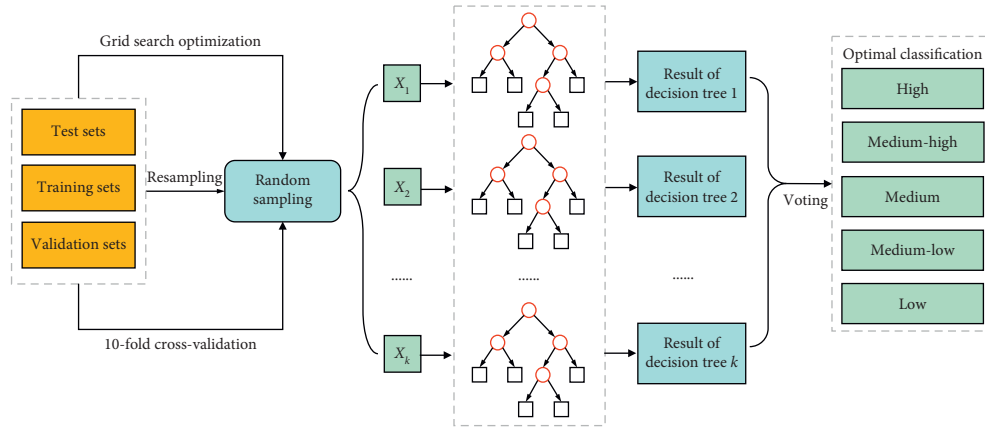


FIGURE 3: Workflows of the improved random forest algorithm used in attention monitoring.

(3) *Grid Search Optimization*. The random forest algorithm has high precision and runs fast. However, a large number of hyperparameters are generated in the course of operation. To obtain attention monitoring results with high accuracy, a grid search is used to optimize the parameters. The grid search algorithm involves meshing the variable regions, traversing all the grid points, solving for the objective function values that satisfy the constraint conditions, and selecting the optimal values [20]. Figure 4 shows the workflow of the optimization process.

To the best of our knowledge, it takes considerable time to traverse all the parameters, which decreases the training speed to some extent. In this paper, we use an improved grid search to increase the training speed. First, we use a large step size for a rough search over a wide range. The mesh built on the coordinate system consists of penalty parameters, the numbers of decision trees, and split features, $n_estimators$, $max_features$, and min_sample_leaf . When a set of parameters meets the set requirements, the optimal parameters and accuracy are output. In a case where more than one set of parameters meets the requirements, the set of parameters with the smallest penalty parameter is output as the best selection object. Then, the search range and step size are reduced to search the parameter set more accurately. The above steps are repeated with a step size of 2 to find the global hyperparameters.

(4) *Multiple Cross-Validation*. The accuracy of the proposed method is closely related to the ratio of training data to test data. To address this problem, S -fold cross-validation is conducted by randomly dividing the data into S subsets without repetition, of which $S - 1$ subsets are used for training and the remaining subset is used for testing:

$$\begin{aligned} &\{T_1, T_2, \dots, T_S\}, \\ &(T_i \cap T_j = \emptyset). \end{aligned} \quad (5)$$

This process is repeated S times, and S accuracies are obtained. After each round, $S - 1$ subsets are selected at random to be retrained. In our paper, we use 10-fold cross-validation, which mitigates the situation of overfitting and yields reliable results. The training set is split into 10 subsets, of which one subset is used for testing and the remaining subsets are used as the training set.

2.2. Neurofeedback-Based Attention Training. Three types of attention mechanisms and closed-loop neurofeedback were adopted to implement the attention training function. Closed-loop neurofeedback technology involves the self-regulation of an individual's brain activity by relying on real-time visual and auditory feedback regarding his/her brain patterns. This technique can maintain specific conditions in the brain states of young adults and improve their cognitive function through training.

We proposed an attention training method based on closed-loop neurofeedback technology that increases individual interest through the use of serious games and improves attention in a relaxing atmosphere without adverse reactions. At the same time, impaired concentration is associated not only with psychology but also with the three previously mentioned attention mechanisms: sustained attention, selective attention, and focus attention. According to the persistence, selectivity, and focus of attention, we designed three serious games (as shown in Figure 5), named Tug of War (Figure 5(a)), Adventures of Birds (Figure 5(b)), and Greedy Jelly (Figure 5(c)) in this study.

2.2.1. Serious Games with Closed-Loop Neurofeedback.

The young adults controlled each game by their attention levels; the specific implementation process is as follows: first, EEG signals were obtained from the young adults through OpenBCI. Then, their attention function was monitored by the IRF algorithm, and the results were quantified as young adults' attention levels during the games. The value of "high attention" was quantified as 1, the value of "medium-high attention" was quantified as 0.75, the value of "medium attention" was quantified as 0.5, the value of "medium-low attention" was quantified as 0.25, and the value of "low attention" was quantified as 0. In the sustained game, only when the young adults' attention levels were higher than a certain threshold would the strength of the character in the game be greater than that of the enemy. During the selective game, the bird's direction (upward, downward, or horizontal flight) was manipulated by the young adults' attention levels. For the focus game, the character, who was equipped with a special skill, would release his skill when the young adults'

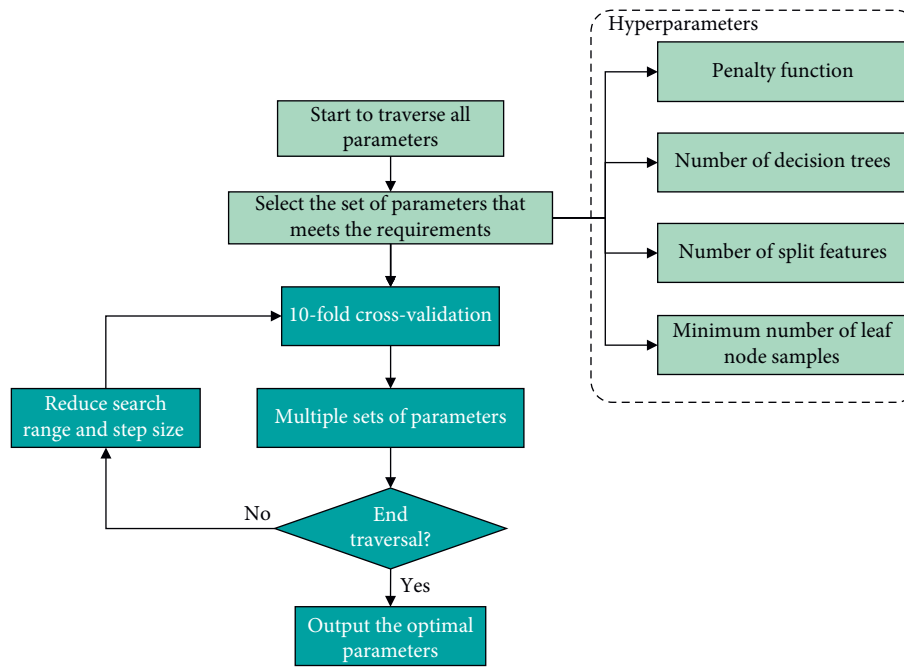


FIGURE 4: Workflow of the optimization process.

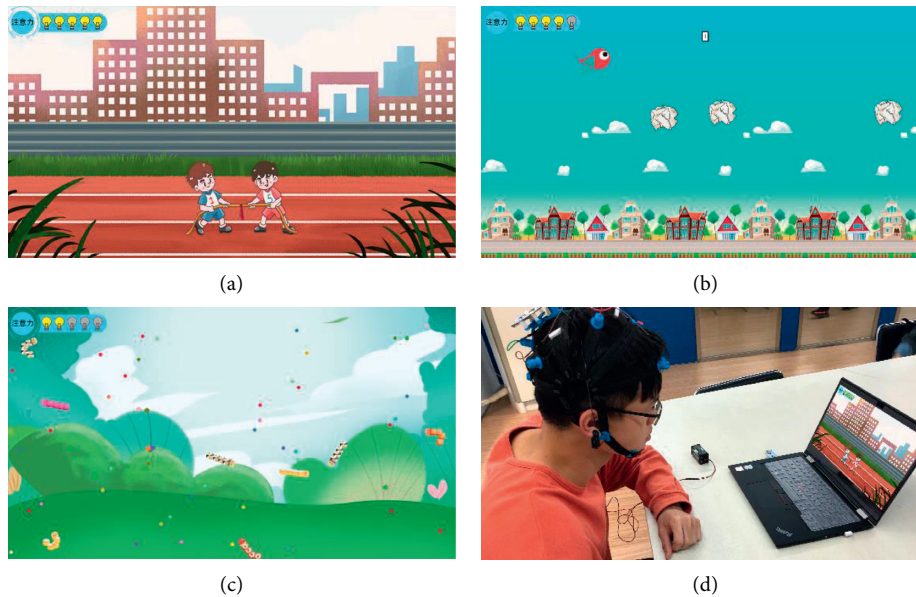


FIGURE 5: Three serious games were developed to improve attention. (a) Tug of War was designed in terms of the mechanism of sustained attention. (b) Adventures of Birds was designed in terms of the mechanism of selective attention. (c) Greedy Jelly was designed in terms of the mechanism of focus attention. (d) A subject playing a serious game with closed-loop neurofeedback.

attention was focused and reached the maximum level. In Figure 5(d), a subject is shown playing a serious game with closed-loop neurofeedback.

To increase the efficiency of attention training, we optimized the graphical user interface (GUI) of the games as follows:

- (1) We designed various game environments and characters by taking young adults as the basis to increase their interest in the GUI.
- (2) Feedback was provided to the young adults in real time in the form of a progress bar that showed their attention levels.

Closed-loop neurofeedback technology can be an additional option for enhancing attention, which means that young adults could control the characters in the games with their attention; at the same time, they could also receive attention feedback from the games. Furthermore, young adults could increase their focus by actively concentrating after becoming acquainted with their attention function.

3. Experiments and Results

3.1. Experiments for Attention Monitoring

3.1.1. Offline Analysis. In offline analysis, the Personal EEG Concentration Tasks dataset involving 80 subjects was used to verify the effectiveness of the proposed attention monitoring method. We selected 70% of the samples as the training set and 30% of the samples as the test set at random. Different algorithms were utilized to divide attention into the abovementioned five levels. The results are shown in Table 1.

Among the five algorithms, the IRF achieved an accuracy of 79.34%, with a loss rate of 21.76%, a recall rate of 76.18%, and a precision of 82.60%. The results show that the attention monitoring method based on the IRF algorithm obtained the highest accuracy rate.

3.1.2. Online Analysis. In the online experiment, we used an OpenBCI headset at 256 Hz to record the EEG data. Ten healthy subjects participated in the experiment, including 5 males and 5 females. The ages of the subjects ranged from 8 to 20 years old (mean = 15.95, std. = 4.63). The study was approved by the Ethics Committee of South China Normal University and complied with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Before the experiment started, the subjects sat on a comfortable chair without blinking or moving their bodies, and they completed the entire experiment according to the provided instructions. There were two sessions during the experiment: a calibration session and an evaluation session.

In the calibration session, each subject performed 20 trials, which took the subjects approximately one hour to complete. Furthermore, each subject was asked to enter their personal information on a computer at the start of the experiment for the purpose of extracting the EEG data and labeling them conveniently. At the beginning of each trial, the computer screen showed a 10-second countdown to help subjects adjust their attention. After that, a calculation task, a minesweeping game task, or an article task was presented on the computer screen, and subjects needed to select an option or read. We induced the attention of subjects through the assigned task and recorded their EEG data simultaneously. After clicking the finish button, the subjects were asked to fill in the valence of self-assessment manikins (SAMs) to report their attention states, that is, high attention, medium-high attention, medium attention, medium-low attention, and low attention. The overall process of this experiment is illustrated in Figure 6.

In the evaluation session, we used similar experimental trials to evaluate the model. In each trial, 6 algorithms (SVM,

TABLE 1: Accuracy (%) of five-level attention monitoring with different algorithms.

Algorithm	Accuracy	Loss	Recall	Precision
Support vector machine (SVM)	52.46	51.48	26.65	53.41
K-nearest neighbors (KNN)	54.75	43.60	50.81	50.12
AdaBoost	67.32	32.35	68.74	71.69
Extreme random tree (ET)	72.90	27.28	65.68	81.09
Random forest (RF)	73.22	26.45	64.02	79.54
Improved random forest (IRF)	79.34	21.76	76.18	82.60

KNN, AdaBoost, ET, RF, and IRF) were applied to detect attention. Ultimately, we calculated the accuracy of each approach by comparing the predicted attention levels and the actual labels.

3.1.3. Results. Figure 7 reveals the accuracy of the attention monitoring results for the 10 subjects, obtained by using the abovementioned five algorithms during the experiment. P values were calculated using a t -test to evaluate the accuracy differences between IRF and other algorithms with the SPSS tool, and the results are shown in Table 2, which were corrected for false discovery rate of $p < 0.01$.

The average accuracy rates of the various algorithms are displayed in Table 3, from which we can find that the accuracy of the IRF is significantly better than the accuracies of the other methods during the online attention monitoring experiment. There is a significant difference ($p = 0.016$) between the IRF algorithm and ET algorithm.

3.2. Experiments for Attention Training

3.2.1. Workflow of the Attention Training Experiment. Ten healthy subjects participated in this experiment, including 5 males and 5 females; these subjects were different from those who participated in the online experiment. The ages of the subjects ranged from 8 to 18 years (mean = 12.5, std. = 4.32). We performed a self-controlled study to validate the effectiveness of the training method, which offers good comparability and high reliability. During the course of the experiment, each subject sat quietly on a chair to avoid excessive movements that would affect the results.

Each subject performed 3 experiments. Before Experiment I and after each experiment, the subjects were required to complete a 5×5 grid of a Schulte table while recording the observed completion time and EEG signals.

There were 3 phases in each experiment: a preparation phase, a training phase, and a rest phase. The preparation phase, which lasted for 3 seconds, required subjects to actively refrain from noticing the game on the screen. During the training phase, the sustained game was presented, and the completion time was recorded. Afterwards, the selective game was carried out, and the scores were recorded when the subjects failed. Finally, the subjects performed the focus game and recorded the times at which they released the special skill. In addition, the resting phase provided a 5-

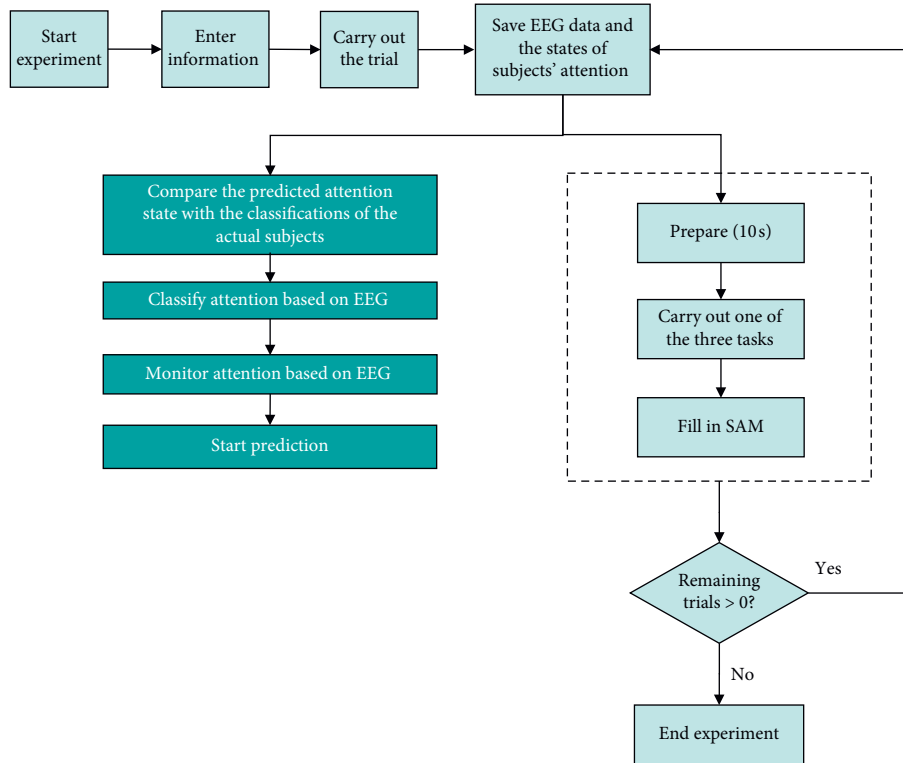


FIGURE 6: Workflow of the attention monitoring experiment.

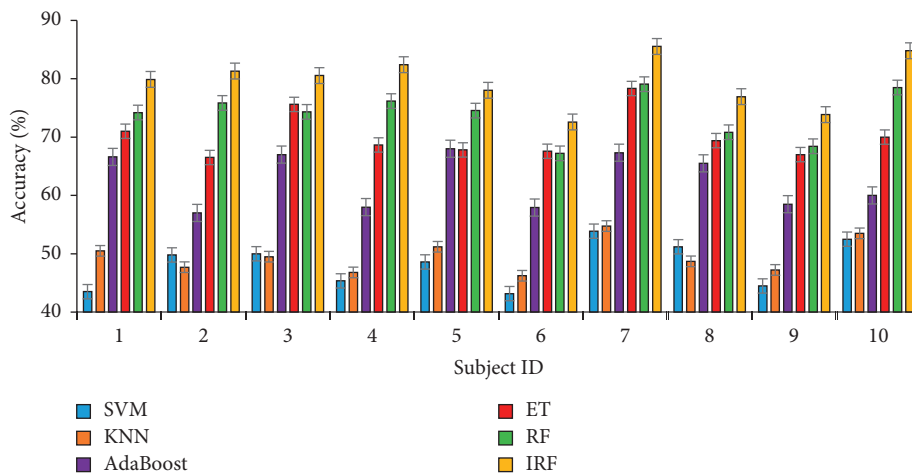


FIGURE 7: Accuracies (%) of the five-level attention results obtained via the SVM, KNN, AdaBoost, ET, RF, and IRF algorithms in online experiments. The x -axis of each subfigure corresponds to the subject IDs, and the y -axis corresponds to the accuracy rates (%).

TABLE 2: Results of a t -test with SPSS.

Group	SVM-IRF	KNN-IRF	AdaBoost-IRF	ET-IRF	RF-IRF
t -test statistics	-29.347	-30.623	-9.565	-7.753	-20.622
P value	<0.001	<0.001	<0.001	<0.001	<0.001

TABLE 3: Average accuracy rates (%) of various algorithms for five levels of attention in online experiments.

Algorithm	SVM	KNN	AdaBoost	ET	RF	IRF
Accuracy (%)	48.52 ± 5.35	50.49 ± 4.26	62.62 ± 4.70	72.95 ± 5.38	73.14 ± 5.93	79.06 ± 6.47

second relaxation time during which subjects could divert their attention from the screen.

3.2.2. Effectiveness. Four primary indicators, including the Schulte times, win times, game scores, and skill times, are proposed as follows:

- (1) Schulte times, which represent the times(s) required to complete the Schulte table.
- (2) Win times, which denote the times(s) required to win a sustained game.
- (3) Game scores, which indicate the scores obtained by each subject upon losing the selective game.
- (4) Skill times, which denote the times at which the special skill was released during the focus game.

The Schulte times of the 10 subjects before Experiment I and after each experiment are shown in Figure 8. Figures 9(a)–9(c) show the win times, game scores, and skill times, respectively.

In the above four figures, all indicators underwent remarkable changes ($p < 0.05$). The Schulte times and win times were obviously reduced. In contrast, the game scores and skill times increased significantly. These can be attributed to two reasons: (i) the familiarity of the subjects with the game environment after much practice and (ii) the effectiveness of neurofeedback. To explore the impact of the neurofeedback on the 3 experiments, we conducted an additional experiment without neurofeedback on the same subjects. We removed the neurofeedback elements from the three games in experiment III, such as the progress bar and the background sound effects used to display the subject's attention level in the GUI. The average results of the four indicators for each experiment are presented in Figure 10. For comparison purposes, we reduced the values of the win times by a factor of 10.

As the subjects became more familiar with the game, even without neurofeedback, their attention was improved to a certain extent. However, the rates of change of the four indicators without neurofeedback were much lower than those observed when using neurofeedback.

Moreover, to verify the accuracy of the control parameters, the Schulte times and the EEG signals input into the Schulte table before the experiment and after each experiment were analyzed. The Schulte times are often inversely proportional to the concentration and attention levels. Table 4 illustrates the comparison between the Schulte times and the results of attention monitoring among the subjects. The results showed that the shorter a given Schulte time was, the more concentrated the subjects were and the higher the attention monitoring results. In contrast, there was less enhancement of the monitoring results.

4. Discussion

The main work in our paper was to propose an IRF-algorithm-based attention monitoring and training method with closed-loop neurofeedback. For attention monitoring, we divided attention into five levels ranging from low to high

attention and applied the IRF algorithm to improve monitoring accuracy and performance. Furthermore, an offline experiment based on the Personal EEG Concentration Tasks dataset and an online experiment involving 10 young adults were carried out. The results yielded an average accuracy of 79.34% for the IRF algorithm. For attention training, we designed three training modes with neurofeedback, corresponding to sustained attention, selective attention, and focus attention. Furthermore, a self-control method with four indicators was used in the attention training experiment, and the results demonstrated a statistically significant performance improvement ($p < 0.05$) for the 10 tested young adults after attention training, thereby demonstrating the effectiveness of the proposed games.

From the perspective of the attention monitoring method, it is essential to achieve a promising level of accuracy and improved classification. At the same time, the attention mechanism and training method are vital for attention training. We show the differences between other studies and our work in Tables 5 and 6.

Most past studies explored attention training methods, such as the focused attention meditation (FAM) method proposed by Yoshida et al. [24]. In addition, Shereena et al. [25] used the EEG neurofeedback training method to design training tasks, aiming to enhance β waves for the purpose of suppressing θ waves. These studies related to attention training focused on EEG signals, with little emphasis on neurofeedback. Notably, several recent papers examined the method of combining neurofeedback and serious games, which is helpful for improving self-regulation skills in attention training with appropriate guidance; an example of this is the “ExerBrain” game [14], which assists individuals in improving their attention control due to the immersive and interactive feature of neurofeedback. Moreover, neurofeedback training has proven to be an efficient tool for sustained attention [26]. Nevertheless, attention, as a kind of higher-order cognition, comprises not only sustained attention but also selective attention and focus attention [25]. As such, this study provides a means by which to implement serious games and analyze three mechanisms of attention simultaneously to help young adults enhance their attention effectively.

The advantages of this paper primarily comprise the following points: (i) we utilized the IRF algorithm for five-level attention monitoring and obtained promising accuracy; (ii) we designed serious games in a multiangle and targeted manner with consideration of sustained, selective, and focus attention; (iii) we quantified the attention monitoring results and used them as the control parameters to manipulate the games with improved accuracy; and (iv) four indicators were proposed in the experiment to validate the effectiveness of the presented method.

In addition, the main finding is that the attention of young adults could be enhanced by using closed-loop neurofeedback in comparison with methods that ignore neurofeedback. This is in line with the results of previous attention monitoring studies [27]. On this basis, this paper used the IRF algorithm to classify 5-level attention and attained an accuracy of 79.34%, which is higher than those of

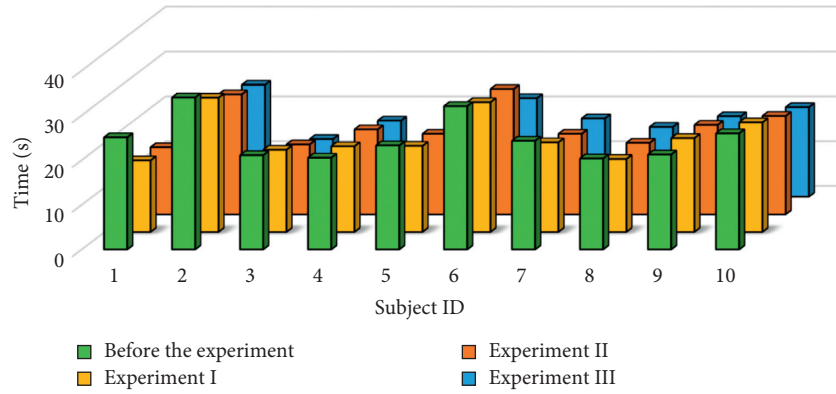
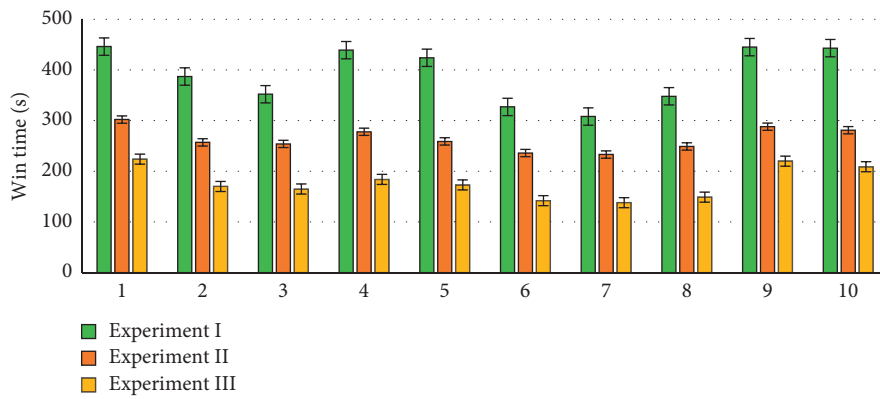
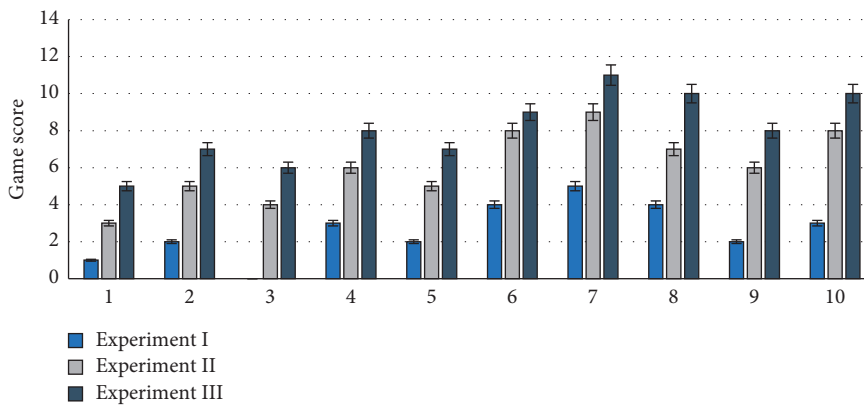


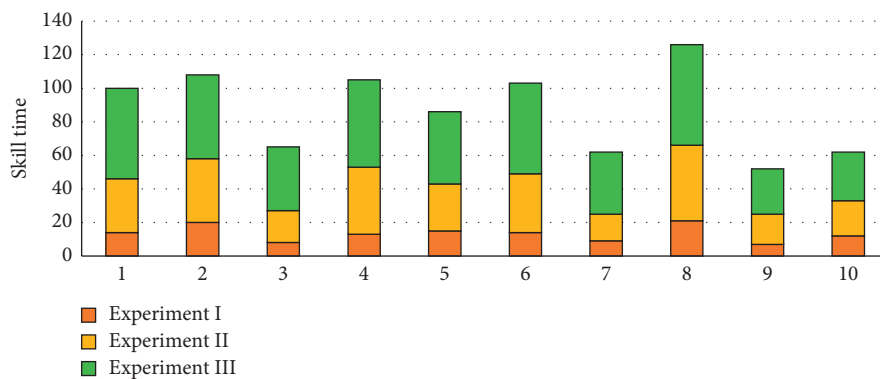
FIGURE 8: Schulte times before the experiment and during experiment (I), experiment II, and experiment III. The *x*-axis of each subfigure corresponds to the subject IDs, and the *y*-axis corresponds to the Schulte times.



(a)



(b)



(c)

FIGURE 9: The win times (a), game scores (b), and skill times (c) obtained during experiment (I), experiment II, and experiment III. The *x*-axis of each subfigure corresponds to the subject IDs, and the *y*-axis corresponds to the performances of the corresponding indicators.

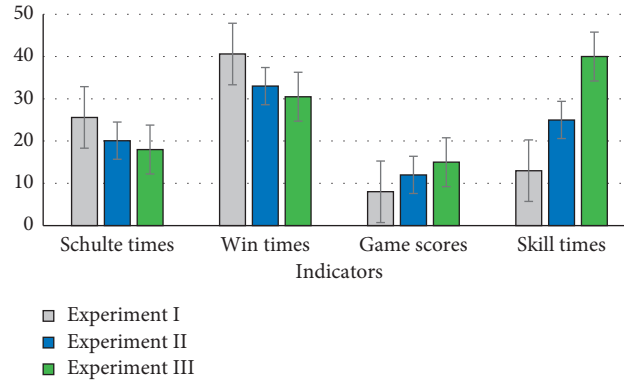


FIGURE 10: Average results of the four indicators in experiment (I), experiment II, and experiment III. The x -axis of each subfigure identifies each indicator, and the y -axis corresponds to the Schulte times, one-tenth of the win times, the game scores, and the skill times.

TABLE 4: Comparison of the Schulte times and the results of attention monitoring.

Subject	Before the experiment		Experiment I		Experiment II		Experiment III	
	T	D	T	D	T	D	T	D
1	25.12	Medium-low	16	High	15	High	14	High
2	34	Low	30	Low	26.8	Low	25	Medium-low
3	21.1	Medium	18.37	Medium-high	15.64	High	12.91	High
4	20.55	Medium-high	19.17	Medium-high	19	Medium-high	17	High
5	23.3	Medium-	19.25	Medium-high	18	Medium-high	16	High
6	32.07	Low	29	Low	28	Low	22	Medium
7	24.35	Medium-low	20.05	Medium-high	18	Medium-high	17.5	High
8	20.37	Medium-high	16.32	High	16	High	15.6	High
9	21.25	Medium	21	Medium	20	Medium-high	18	Medium-high
10	26	Medium-low	24.5	Medium-low	22	Medium	20	Medium-high

Note. T denotes the Schulte time of each subject, and D denotes the monitoring result of each subject.

TABLE 5: Comparison of different methods related to attention monitoring.

Reference	Algorithm	Classification	Accuracy (%)
[1]	ANN	4 levels	78
[21]	KNN	3 levels	67
[22]	Naive Bayes	3 levels	60
Our work	IRF	5 levels	79.34

TABLE 6: Comparison of different methods related to attention training.

References	Mechanism	Method
[2]	Sustained attention	A closed-loop tactile training process related to visual sustained attention.
[14]	Sustained attention	Closed-loop neurofeedback from MVPA as a type of cognitive prosthetic.
[23]	Sustained attention and selective attention	A 3D game with neurofeedback.
Our work	Sustained attention, selective attention, and focus attention	Three serious games related to the mechanism of attention with neurofeedback.

the other algorithms that were compared. Furthermore, consistent with past studies [11, 25], we showed that neurofeedback training can assist young adults in improving their attention. In the present study, we implemented three serious games for young adults with neurofeedback and incorporated three mechanisms of attention that were not fully considered in past studies. Furthermore, the attention

training method in this paper contributed to significant improvements in sustained attention, selective attention, and focus attention. Therefore, we can conclude that all mechanisms of attention can be improved with specific serious games.

The primary limitation of our study was the small number of subjects examined during the experiment. In

addition, there are a few datasets pertaining to the attention of young adults; these data are challenging to record due to the volatility of conductive media in BCIs. In addition, several external factors, such as the habits, motivations, and mental statuses of the subjects, were not considered.

5. Conclusions

We proposed an IRF-algorithm-based attention monitoring and training method with closed-loop neurofeedback, and we presented the mechanism of attention. In the future, we will collect attention data and attempt to fuse EEG and physiological signals (such as facial expressions and verbal speech) to improve the accuracy of attention monitoring. Furthermore, we will develop 3D serious games in the near future.

Data Availability

All the data included in this study are available upon request by contacting the corresponding author.

Conflicts of Interest

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

Acknowledgments

The authors express their gratitude to the 10 volunteers who participated in this experiment. They also sincerely appreciate the developers of Python, MNE, BrainFlow, and the other software packages used in this paper. This work was supported in part by the Guangzhou Science and Technology Plan Project Key Field R&D Project under Grant no. 202007030005, the National Natural Science Foundation of China under Grant no. 61876067, the Guangdong Natural Science Foundation of China under Grant no. 2019A1515011375, and the Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation under Grant no. pdjh2020a0145.

References

- [1] M. Mohammadpour and S. Mozaffari, "Classification of EEG-based attention for brain computer interface," in *Proceedings of the 2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS)*, pp. 34–37, IEEE, Shahrood, Iran, December 2017.
- [2] Y. Luo and J. Zhang, "The effect of tactile training on sustained attention in young adults," *Brain Sciences*, vol. 10, no. 10, p. 695, 2020.
- [3] R. Chen, Z. Yang, J. Li et al., "State and short-term effects of mindfulness meditation training on attention," *Journal of Vision*, vol. 20, no. 11, p. 1815, 2020.
- [4] Y.-Y. Tang and M. I. Posner, "Attention training and attention state training," *Trends in Cognitive Sciences*, vol. 13, no. 5, pp. 222–227, 2009.
- [5] T. A. Putri, M. Muniroh, Y. Purwoko et al., "Regular high intensity circuit training improves attention function and reaction time among male young adults," *Malaysian Journal of Medicine and Health Sciences*, vol. 16, no. 14, pp. 57–60, 2020.
- [6] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain-computer interface paradigms," *Journal of Neural Engineering*, vol. 16, no. 1, p. 011001, 2019.
- [7] M. Liu, J. Zhang, W. Jia et al., "Enhanced executive attention efficiency after adaptive force control training: behavioural and physiological results," *Behavioural Brain Research*, vol. 376, p. 111859, 2019.
- [8] A. N. Belkacem, N. Jamil, J. A. Palmer, S. Ouhbi, and C. Chen, "Brain computer interfaces for improving the quality of life of older adults and elderly patients," *Frontiers in Neuroscience*, vol. 14, p. 692, 2020.
- [9] X. Qian, B. R. Y. Loo, F. X. Castellanos et al., "Brain-computer-interface-based intervention re-normalizes brain functional network topology in children with attention deficit/hyperactivity disorder," *Translational Psychiatry*, vol. 8, no. 1, p. 149, 2018.
- [10] N. Kosmyna and P. Maes, "AttentivU: an EEG-based closed-loop biofeedback system for real-time monitoring and improvement of engagement for personalized learning," *Sensors*, vol. 19, no. 23, p. 5200, 2019.
- [11] H.-S. Chiang, K.-L. Hsiao, and L.-C. Liu, "EEG-based detection model for evaluating and improving learning attention," *Journal of Medical and Biological Engineering*, vol. 38, no. 6, pp. 847–856, 2018.
- [12] B. Hu, X. Li, S. Sun, and M. Ratcliffe, "Attention recognition in EEG-based affective learning research using CFS+KNN algorithm," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 1, pp. 38–45, 2018.
- [13] H. S. Mohammadi, E. Pirbabaei, M. J. Sisi, and Y. A. Sekhavat, "ExerBrain: a comparison of positive and negative reinforcement in attention training using BCI based computer games," in *Proceedings of the 2018 2nd National and 1st International Digital Games Research Conference: Trends, Technologies, and Applications (DGRC)*, pp. 167–171, IEEE, Tehran, Iran, November 2018.
- [14] M. T. Bettencourt, J. D. Cohen, R. F. Lee et al., "Closed-loop training of attention with real-time brain imaging," *Nature Neuroscience*, vol. 18, no. 3, pp. 470–475, 2015.
- [15] O. Castillo, S. Sotomayor, G. Kemper, and V. Clement, "Correspondence between TOVA test results and characteristics of EEG signals acquired through the muse sensor in positions AF7-AF8," in *Proceedings of the 5th Brazilian Technology Symposium*, pp. 111–120, Springer, Cham, Switzerland, December 2021.
- [16] T. Taillez, B. Kollmeier, and B. T. Meyer, "Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech," *European Journal of Neuroscience*, vol. 51, no. 5, pp. 1234–1241, 2020.
- [17] N. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. Ahmad, M. Islam, and J. Escudero, "Selection of mother wavelet functions for multi-channel EEG signal analysis during a working memory task," *Sensors*, vol. 15, no. 11, p. 29015, 2015.
- [18] N. K. Al-Qazzaz, S. Ali, S. A. Ahmad, M. S. Islam, and M. I. Ariff, "Selection of mother wavelets thresholding methods in denoising multi-channel EEG signals during working memory task," in *Proceedings of the 2014 IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 214–219, IEEE, Kuala Lumpur, Malaysia, December 2014.
- [19] A. Belle, R. H. Hargraves, and K. Najarian, "An automated optimal engagement and attention detection system using electrocardiogram," *Computational and Mathematical*

- Methods in Medicine*, vol. 2012, Article ID 528781, 12 pages, 2012.
- [20] X. Wang, G. Gong, N. Li, and S. Qiu, "Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization," *Frontiers in Human Neuroscience*, vol. 13, p. 52, 2019.
- [21] Y. Li, X. Li, M. Ratcliffe et al., "A real-time EEG-based BCI system for attention recognition in ubiquitous environment," in *Proceedings of 2011 International Workshop on Ubiquitous affective Awareness and Intelligent Interaction*, pp. 33–40, Beijing, China, September 2011.
- [22] R. Srinivasan, S. Thorpe, S. Deng, T. Lappas, and M. D'Zmura, "Decoding attentional orientation from EEG spectra," in *Proceedings of the International Conference on Human-Computer Interaction*, pp. 176–183, Springer, Berlin, Germany, July 2009.
- [23] K. P. Thomas, A. P. Vinod, and C. Cuntai Guan, "Design of an online EEG based neurofeedback game for enhancing attention and memory," in *Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 433–436, IEEE, Osaka, Japan, July 2013.
- [24] K. Yoshida, K. Takeda, T. Kasai et al., "Focused attention meditation training modifies neural activity and attention: longitudinal EEG data in non-meditators," *Social Cognitive and Affective Neuroscience*, vol. 15, no. 2, pp. 215–224, 2020.
- [25] E. A. Shereena, R. K. Gupta, C. N. Bennett, K. J. V. Sagar, and J. Rajeswaran, "EEG neurofeedback training in children with attention deficit/hyperactivity disorder: a cognitive and behavioral outcome study," *Clinical EEG and Neuroscience*, vol. 50, no. 4, pp. 242–255, 2019.
- [26] S. Jirayucharoensak, P. Israsena, S. Pan-Ngum, S. Hemrungronj, and M. Maes, "A game-based neurofeedback training system to enhance cognitive performance in healthy elderly subjects and in patients with amnesic mild cognitive impairment," *Clinical Interventions in Aging*, vol. 14, pp. 347–360, 2019.
- [27] N. Kosmyna, C. Morris, U. Sarawgi et al., "Attentivu: a biofeedback system for real-time monitoring and improvement of engagement," in *Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-2, Glasgow, UK, May 2019.

Research Article

Classification of Mental Stress Using CNN-LSTM Algorithms with Electrocardiogram Signals

Mingu Kang, Siho Shin, Jaehyo Jung , and Youn Tae Kim 

AI Healthcare Research Center, Department of IT Fusion Technology, Chosun University, 309 Pilmun-daero Dong-gu, Gwangju 61452, Republic of Korea

Correspondence should be addressed to Jaehyo Jung; jh.jung207@gmail.com and Youn Tae Kim; petruskim@chosun.ac.kr

Received 30 March 2021; Revised 30 April 2021; Accepted 21 May 2021; Published 7 June 2021

Academic Editor: G R Sinha

Copyright © 2021 Mingu Kang 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 mental stress faced by many people in modern society is a factor that causes various chronic diseases, such as depression, cancer, and cardiovascular disease, according to stress accumulation. Therefore, it is very important to regularly manage and monitor a person's stress. In this study, we propose an ensemble algorithm that can accurately determine mental stress states using a modified convolutional neural network (CNN)- long short-term memory (LSTM) architecture. When a person is exposed to stress, a displacement occurs in the electrocardiogram (ECG) signal. It is possible to classify stress signals by analyzing ECG signals and extracting specific parameters. To maximize the performance of the proposed stress classification algorithm, fast Fourier transform (FFT) and spectrograms were applied to preprocess ECG signals and produce signals in both the time and frequency domains to aid the training process. As the performance evaluation benchmarks of the stress classification model, confusion matrices, receiver operating characteristic (ROC) curves, and precision-recall (PR) curves were used, and the accuracy achieved by the proposed model was 98.3%, which is an improvement of 14.7% compared to previous research results. Therefore, our model can help manage the mental health of people exposed to stress. In addition, if combined with various biosignals such as electromyogram (EMG) and photoplethysmography (PPG), it may have the potential for development in various healthcare systems, such as home training, sleep state analysis, and cardiovascular monitoring.

1. Introduction

Stress is a mental and physical reaction that a person may feel when they find themselves in a difficult and/or unfamiliar environment or situation. Excessive stress accumulation can cause chronic diseases such as high blood pressure, heart disease, and cancer and, in severe cases, can lead to death [1, 2]. For this reason, stress observation is becoming increasingly important in modern society.

Studies measuring stress by using various biological signals such as electroencephalography (EEG), electromyogram (EMG), oxygen saturation, and pulse waves have been published [3–5]. However, these measurement methods require expensive and bulky systems to acquire data, are complicated and expensive to use, and require signal analysis by experts.

Existing studies using EEG signals have analyzed stress using support vector machines (SVMs), multilayer perceptrons (MLPs), and naïve Bayes (NB) and have obtained accuracies of 75%, 85.20%, and 64.29%, respectively [6–8]. However, because these studies used only 15 EEGs as training data, underfitting can occur. Furthermore, because an EEG produces a 7-channel signal, it involves a complex and time-consuming process to measure stress signals. Previous studies using EMG signals analyzed by SVM achieved an 85% accuracy. However, despite the same action being taken (the characteristic movement of the muscles), the magnitude of the signal amplitude varies from measurement to measurement, and noise in the signal makes it difficult to extract accurate feature points [9].

Studies that classify stress using an electrocardiogram (ECG) have been the most popular because the signal

acquisition method is simpler than other methods and a clear waveform is acquired. Two studies achieved 89.21% and 84.4% accuracy using SVM [10, 11], but extracting feature points can be difficult because of noise and the time required to measure multichannel ECG signals and because preprocessing is not always accurate. Two different studies achieved 75% and 89% accuracy by considering the standard deviation of the R-R interval of the heart rate variability (HRV) signal [12, 13]. Accurate stress classification is difficult because it takes more than 5 min to calculate the standard deviation of the R-R interval, and because the difference in parameter values is minimal. Furthermore, because the ECG waveform is not accurate in the frequency domain, it is difficult to extract feature points, making it difficult to directly evaluate or minimize the effect of noise generated by the human body.

In addition, certain research results have exhibited 63.97% and 82.7% accuracy using fuzzy c-means (FCM) clustering and convolutional neural network (CNN) [14, 15]. These studies have difficulty classifying stress signals because the distance between the center point and the data is slight, and the scale of the training data is small, making it easier for the occurrence of underfitting.

Certain earlier study results have exhibited 87.39% and 90.19% accuracy using CNNs and convolutional recurrent neural networks (CRNNs) [16, 17]. In these studies, the hierarchical structure of the stress classifier is complex, and there is a considerable amount of noise; therefore, it is difficult to achieve a high-stress classification accuracy by detecting an incorrect R peak value. Models based on long short-term memory (LSTM) achieved 88.13% accuracy [18]. However, owing to the high noise of the ECG signal, it is difficult to calculate the root mean square (RMS) of the R-R interval.

The aforementioned stress signal classification algorithm using the ECG signal has disadvantages such as underfitting, the calculation of a standard deviation for the R-R interval of a long-time HRV signal, and the detection of an incorrect R peak value. To overcome these problems, we propose an ensemble model that accurately classifies mental stress by combining CNN and LSTM. The proposed model extracts the $R - S_{\text{peak}}$ feature point using the threshold value, converts it into a spectrogram, and classifies the stress signal using ECG signal analysis.

To improve the stress classification accuracy, batch normalization (BN), flatten layers, and fully connected layers were added. Subsequently, the accuracy of the stress classification model was improved by separately classifying ECG signals in the time domain and frequency domain. Confusion matrices, receiver operating characteristic (ROC) curves, and precision-recall (PR) curves were used to evaluate the performance of the stress classification model. In this study, we proposed an ensemble method to classify the mental stress of the CNN-LSTM model using ECG signals. The data of the ST Change Database and WESAD Database were trained, and more than 98% classification performance was achieved.

2. Materials and Methods

2.1. Subject. Figure 1 shows the procedure for classifying stress signals. In this study, we used the ST Change Database and WESAD Database, which provide ECG signals that were acquired in different stress environments. The ST Change Database contains ECG data that records physical stress and consists of 28 ECG signals obtained from 15 male subjects [19]. The WESAD database contains 30 ECG signals measured at the wrist and chest obtained from 15 subjects (12 men and 3 women) [20].

2.2. Preprocessing and Feature Extraction. Electrocardiography is the most common way to check health status by noninvasively checking the electrical status of the heart. When taking an electrocardiogram, noise is generated by several factors, which greatly reduces ECG classification accuracy [21]. To solve this problem, we used a low-pass filter and confirmed that 90.89% of the noise was eliminated using a low-pass filter with a sampling frequency of 360 Hz and a cutoff frequency of 150 Hz.

Figure 2 shows the extracted $R - S_{\text{peak}}$ values from an ECG signal. By extracting these data under stress and without stress, the ECG can be accurately analyzed [22]. R_{peak} and S_{peak} were extracted from ECG signals after setting a threshold. R_{peak} extracted the pole when the threshold value was greater than 0.2 mV in one period of the signal and extracted the pole when the threshold value was less than -0.54 mV in one period.

In the under-stress state, the heart beats irregularly and quickly, the R-R interval of the ECG signal becomes narrow, and the $R - S_{\text{peak}}$ increases. On the other hand, in the unstressed state, the heart is relatively stable, the R-R interval widens, and the $R - S_{\text{peak}}$ decreases [23]. In each state, the average $R - S_{\text{peak}}$ without stress was found to be 1.47 mV, and under stress, it was 4.25 mV. Figure 3 shows the conversion of either signal (under stress or without stress) into a spectrogram.

2.3. CNN-LSTM Model Design. Figure 4 shows the architecture of the ensemble model proposed in this study. The classification layer consists of 14 levels.

Table 1 lists the structure of the layers comprising the ensemble model. First, $124 \times 124 \times 3$ image sequence data are input to the sequence input layer. Subsequently, the ECG image data are converted into an array form (vertical, horizontal, and channel) using a sequence folding layer and then transferred to the convolution layer.

The reason for using the sequence folding layer is so that the image sequence data can be converted into an array, arranged, and then transferred to the two-dimensional (2D) convolution layer. The first 2D convolution layer contains six filters of size 5×5 .

Because of calculating the convolution layer using equation (1), the size of the output value is $124 \times 124 \times 6$. Equation (1) represents the calculation process for the

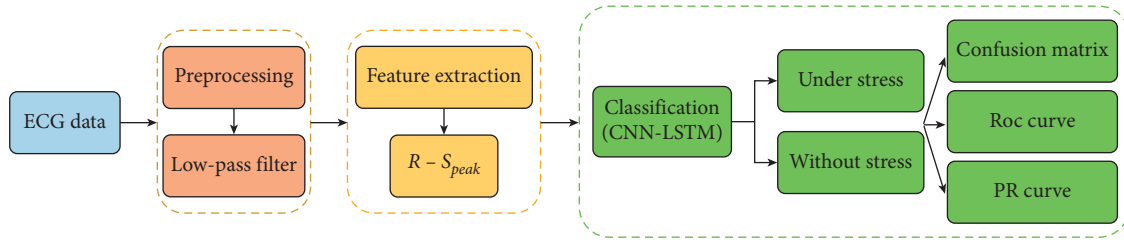


FIGURE 1: Procedure for classifying stress signals and validating the model.

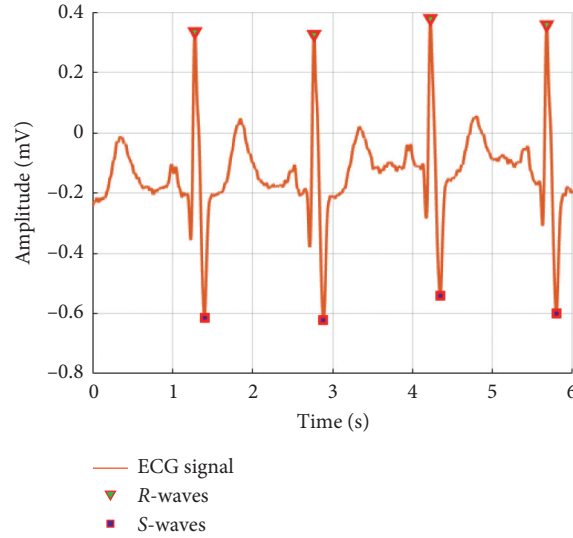


FIGURE 2: Feature point extraction by threshold.

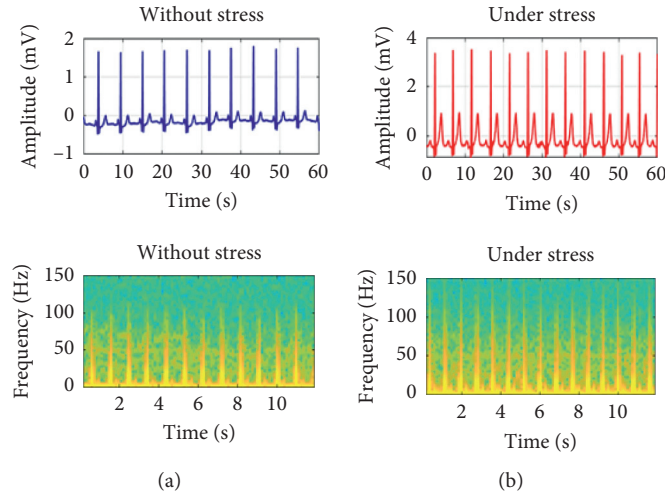


FIGURE 3: (a) Without stress ECG converted to a spectrogram. (b) Under stress ECG converted to a spectrogram.

convolution layer. When padding and stride are applied, and the size of the input data and filter is given, the output value can be calculated. H and W are the input data size, FH (filter height) and FW (filter weight) are filter size, S is the stride, P is padding, and OH (output height) and OW (output weight) are output value sizes.

$$(OH, OW) = \left(\frac{H + 2P - FH}{S} + 1, \frac{W + 2P - FW}{S} + 1 \right). \quad (1)$$

The output data are then connected to the batch normalization layer. After normalizing the size of the output

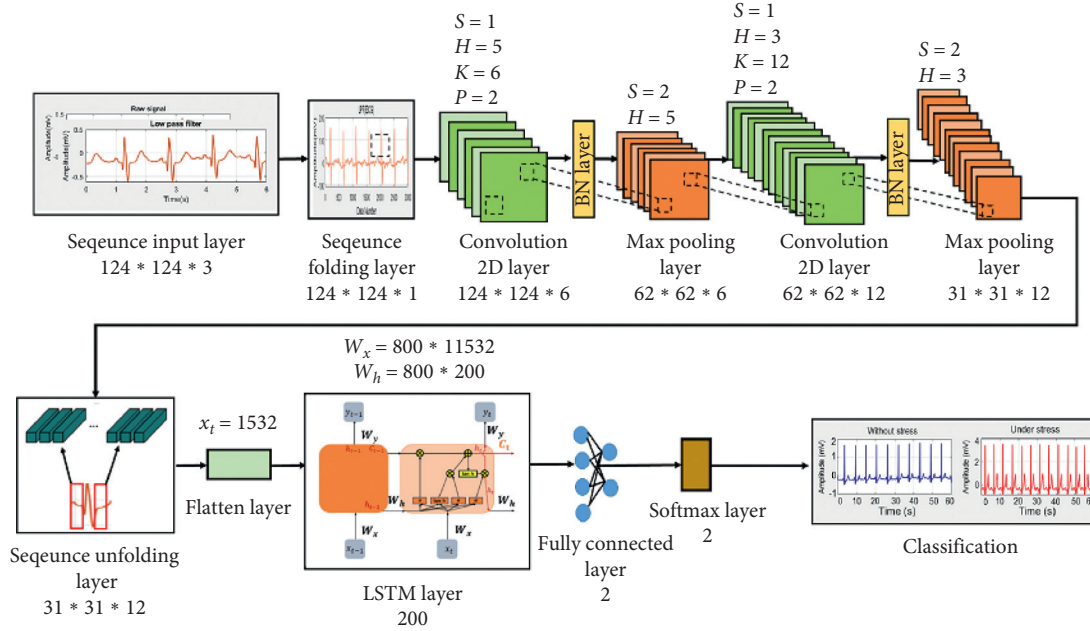


FIGURE 4: Classification architecture.

TABLE 1: Classification layers used to evaluate stress signals using CNN-LSTM.

Number	Layer	Activation	Weights	Bias
1	Sequence input layer	124 × 124 × 3	—	—
2	Sequence folding layer	124 × 124 × 1	—	—
3	Convolution 2D layer	124 × 124 × 6	5 × 5 × 3 × 6	1 × 1 × 6
4	Batch normalization layer	124 × 124 × 6	—	—
5	Max pooling layer	62 × 62 × 6	—	—
6	Convolution 2D layer	62 × 62 × 12	3 × 3 × 6 × 12	1 × 1 × 12
7	Batch normalization layer	62 × 62 × 12	—	—
8	Max pooling layer	31 × 31 × 12	—	—
9	Sequence unfolding layer	31 × 31 × 12	—	—
10	Flatten layer	11532	—	—
11	LSTM layer	200	Input: 800 × 11532 recurrent: 800 × 200	800 × 1
12	Fully connected layer	2	2 × 200	2 × 1
13	Softmax layer	2	—	—
14	Classification	—	—	—

data to $124 \times 124 \times 6$ in the batch normalization layer, it was connected to the max pooling layer. According to equation (2), the size of the output data is determined by dividing the row and column size by the pooling size.

$$(ORs, OCs) = \left(\frac{H}{P}, \frac{W}{P} \right). \quad (2)$$

Its output is fed to a batch normalization layer and then to a max pooling layer. The max pooling layer is a 2×2 filter with a stride of 2. As a result, the original data are reduced to a size of $62 \times 62 \times 6$. The second 2D convolution layer contains 12 filters of size 3×3 . As a result, the data are further reduced to a size of $31 \times 31 \times 12$. Normalization is then performed and the data are passed to the LSTM layer. To transfer the size of the output data to the LSTM layer, normalization was performed using a sequence unfolding

layer, and feature vectors were obtained using a flattening layer (or flattened layer).

The flattening layer has the advantage of not affecting the parameter by converting the output of the extracted feature map into a 1D array, which allows reconstructing the feature maps as the input to the LSTM [24]. At this time, the input is transmitted through the hidden layer of the LSTM.

A weight value of 800×11532 at the input layer is applied to equations (3)–(7), which represents the computational process of the LSTM layer, to extract the feature value. The LSTM layer consists of input gates (i_t , g_t), forget gates (f_t), and output gates (O_t). The LSTM layer is composed of an input gate (i_t , g_t), forgetting gate (f_t), and output gate (O_t). In each gate, a weight value is multiplied according to an input vector (x_t), a hidden state (h_{t-1}), and a cell state (C_t) using the sigmoid and Tanh functions, and then a feature value is extracted.

$$i_t = \sigma(W_x x_t + W_{hi} h_{t-1} + b_i), \quad (3)$$

$$g_t = \tanh(W_{xg} x_t + W_{hg} h_{t-1} + b_g), \quad (4)$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f), \quad (5)$$

$$O_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o), \quad (6)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ g_t. \quad (7)$$

Subsequently, it is applied to equation (8) to pass the feature value calculated at the output gate to the output layer. Equation (8) is the process of extracting a required feature value from several feature values calculated at the output gate. After extracting a feature value from -1 to 1 using the Tanh function, the feature value in the range calculated using the output gate is transferred to the output layer.

$$h_t = O_t \circ \tanh(C_t). \quad (8)$$

The feature value extracted from the LSTM layer classifies the image using a fully connected layer of size two and calculates a probability value for the image classified by the softmax layer. Subsequently, image classification is performed using the feature values extracted earlier using the fully connected layer, and the probability value of the classified image is calculated using the softmax layer. Finally, in the classification step, the signal is classified as either under stress or without stress.

Figure 5 shows the components of the convolution 2D layer and LSTM layer to which equations (1)–(8) are applied. Equations (1) and (2) show the calculation process of the convolution 2D layer among the CNN models, and equations (3)–(7) show the process of outputting feature values using the weight values of the input gate, forgetting the gate, and output gate in the LSTM layer. Equation (8) transfers the feature values in the range from the output gate to the output layer.

We used the confusion matrix, receiver operating characteristic (ROC) curve, and precision-recall (PR) curve to evaluate the stress signal classification performance of the proposed ensemble model [25]. The confusion matrix is a matrix that allows one to evaluate how accurately the predicted value is compared to the actual observed value. We used ECG data from the ST Change Database (DB) and the WESAD DB. The total number of data points was 58. However, with such a small amount of data, it is difficult to accurately evaluate the stress signal classification model. Therefore, to improve the accuracy of the classification model and better analyze its performance, the data were doubled by transforming the time domain data to frequency domain data using the fast Fourier transform (FFT), as indicated in Figure 6. After preprocessing, the performance of the ensemble model was evaluated using 58 time domain data and 58 frequency domain data.

3. Experimental Results

Table 2 shows the accuracy, sensitivity, specificity, precision, and negative predictive values obtained to evaluate the classification model's performance using formulas (9)–(13) [26–28]. Formula (1) defines accuracy and indicates the probability of accurately classifying all under stress and without stress conditions. In the formula, TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively. For the time and frequency domains, the accuracies of the stress classifier were 94.8% and 98.3%, respectively.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (9)$$

Sensitivity refers to the proportion of data correctly classified as without stress to all without stress data (actual observed data). In the time and frequency domains, the sensitivities of the stress classifier were 96.4% and 100%, respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{FN} + \text{TP}}. \quad (10)$$

Specificity is the proportion of data correctly classified as under stress among all under stress data (actual observed data). In the time and frequency domains, the sensitivities of the stress classifier were 96.4% and 100%, respectively.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (11)$$

Precision is the ratio of the data correctly classified by the stress classification algorithm as without stress to the value of all data classified as without stress. In the time and frequency domains, the precision of the stress classifier was 93.1% and 96.6%, respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (12)$$

The negative predictive value is the ratio of data classified correctly as under stress to the actual value without stress data. In the time and frequency domains, the negative predictive values of the stress classifier were 96.6% and 100%, respectively.

$$\text{Negative Predictive Value} = \frac{\text{TN}}{\text{TN} + \text{FN}}. \quad (13)$$

Figure 6 shows the results of the classification model's performance using a confusion matrix. The matrix on the left of Figure 7 uses the data converted to the time domain, and the matrix on the right is the result of using the data in the frequency domain. The highest classification accuracy of the proposed ensemble model was 98.3% for the frequency domain. In previous studies, the accuracy of the model was 83.6% [29]. These results indicate that accuracy was improved by 14.7% using the proposed ensemble compared to previous results.

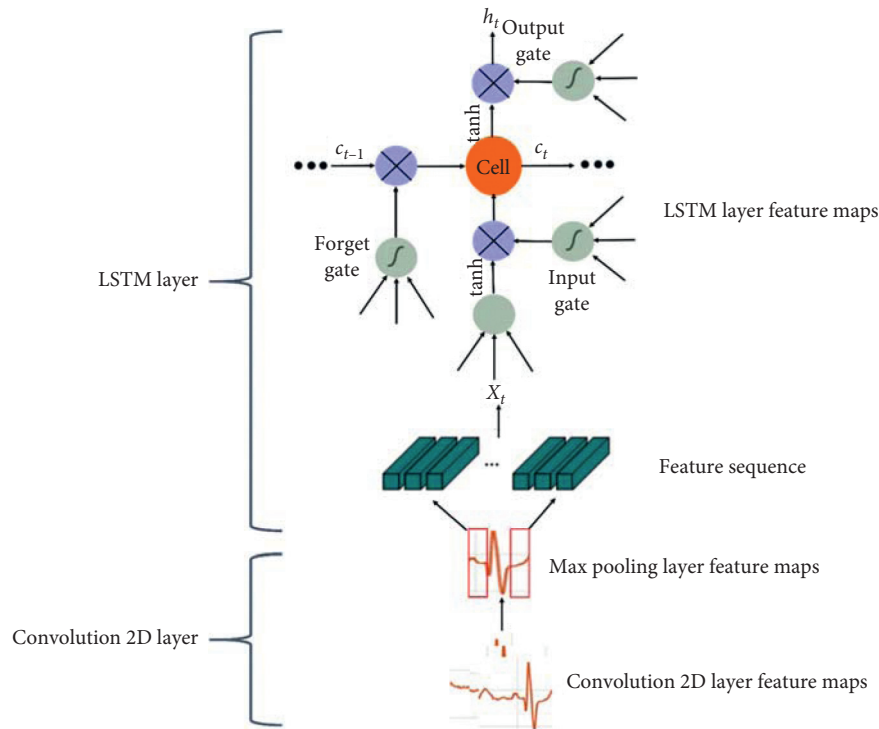


FIGURE 5: Components of the convolution 2D layer and the LSTM layer.

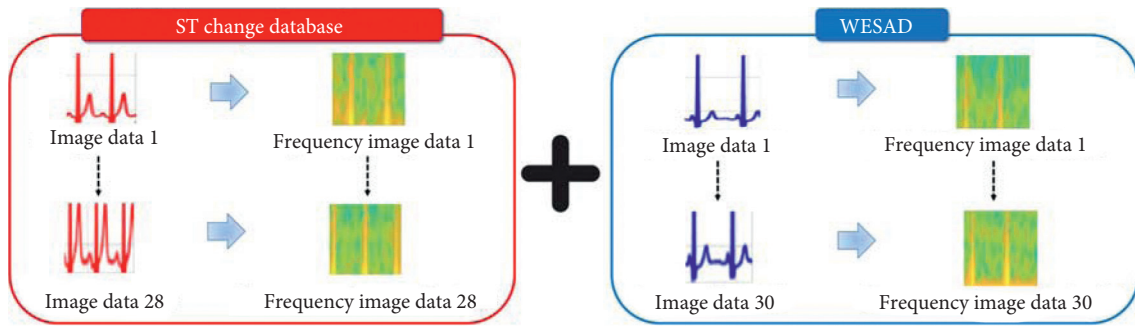


FIGURE 6: FFT transformation to increase input data quantity.

TABLE 2: Classification performance assessment of stress signals in time and frequency domains.

Stress	Time domain				
	Precision	Sensitivity	Specificity	Negative predictive value	Accuracy
Performance (%)	93.1%	96.4%	93.3%	96.6%	94.8%
Error (%)	6.9%	3.6%	6.7%	3.4%	5.2%
Stress	Frequency domain				
	Precision	Sensitivity	Specificity	Negative predictive value	Accuracy
Performance (%)	96.6%	100%	96.7%	100%	98.3%
Error (%)	3.4%	0.0%	3.3%	0.0%	1.7%

Figure 8 shows the classification performance according to the epochs for the time and frequency domains. The graph shows the mean squared error (MSE) according to the number of epochs. The time domain yielded the lowest MSE at 219 epochs (the validation curve shown), while the frequency domain yielded the lowest MSE at 223 epochs.

Figure 9 shows the ROC curves according to the epochs for the time and frequency domains of the ECG data. The ROC curve is a performance evaluation technique applicable to a binary classifier system that indicates how the performance of the classification model changes as the threshold changes [30]. The area under the curve (AUC) (the area under the ROC curve) is an index used to evaluate the

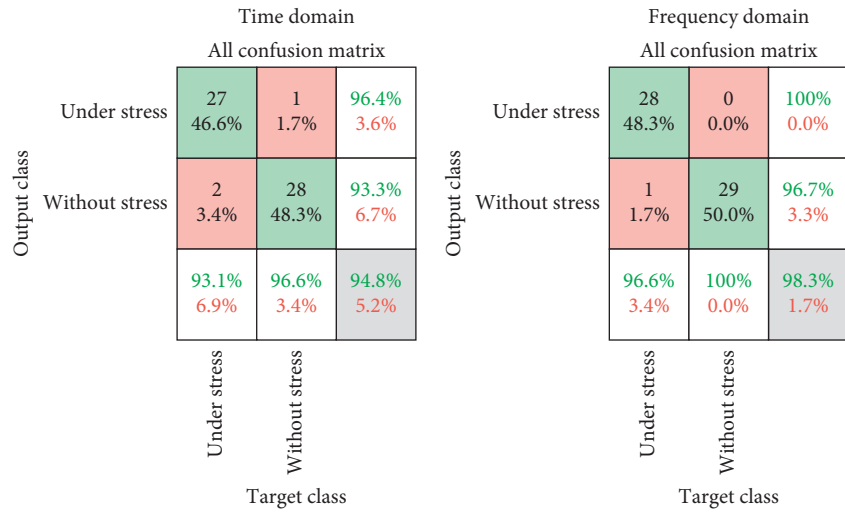


FIGURE 7: Classification performance evaluation of stress signals using a confusion matrix.

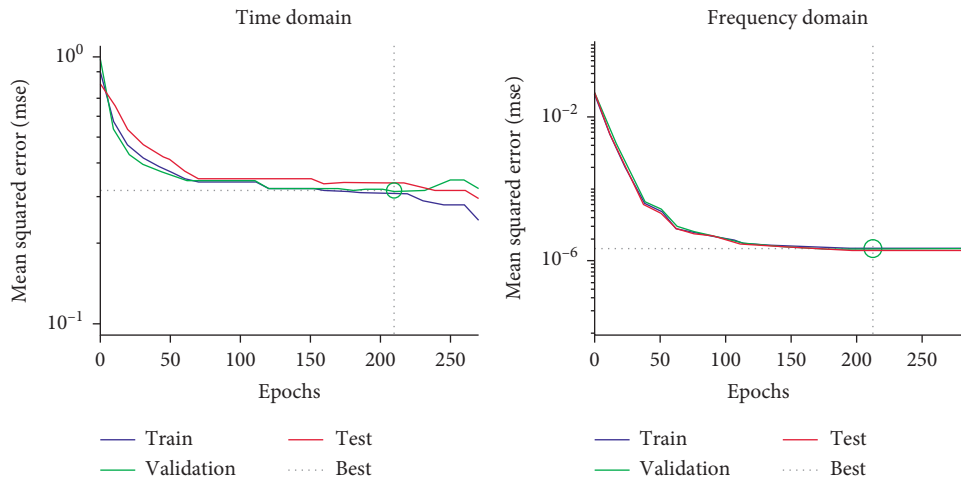


FIGURE 8: Classifier performance evaluation according to epoch.

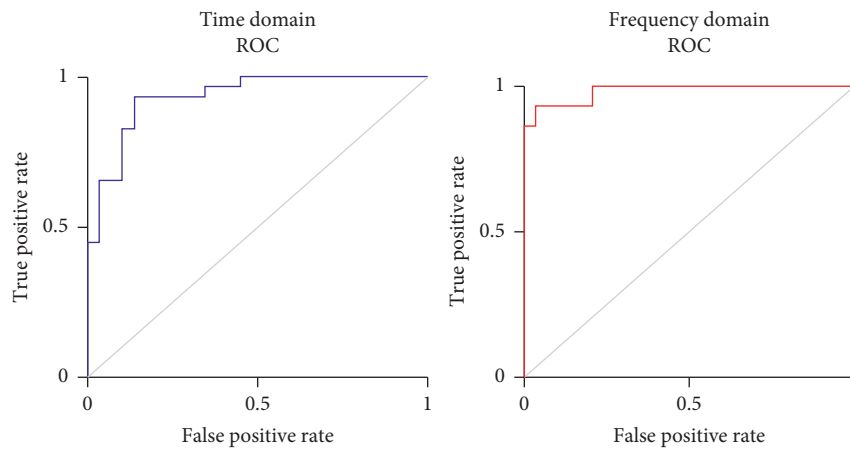


FIGURE 9: Stress classification performance evaluation using ROC curves.

classification performance of different types of signals (stress signals in this study). When the AUC range falls between 0.9 and 1.0 (90%–100%), the classification performance is excellent, and when the AUC range falls between 0.8 and 0.9 (80%–90%), the classifier's performance is low. In the time domain, the AUC of the ROC curve was 94.67%, and it was 98.12% in the frequency domain. The AUC of a previous study was 85.7% [14], and it was confirmed that the ensemble proposed in this study represents a 12.42% improvement. The AUC value of the frequency domain was 3.45% higher than that of the time domain in our model indicating that the classification performance of the stress signal is better in the former.

Figure 10 shows the PR curves for the ECG data according to the epochs for the time and frequency domains. When considering the ROC curve, if the dataset is unbalanced, the shape of the curve is skewed to one side, and the classifier performance cannot be accurately evaluated [31]. The PR curve can be used to overcome the shortcomings of the ROC curve and shows the correlation between precision and recall. The average precision (AP) of the PR curve is an index that can be used to evaluate the classification performance of stress signals [32].

The X-axis represents the recall, and the Y-axis represents the precision. In the PR curve, the larger the AP is, the better the stress signal classification performance. The PR curve AP of the time domain was 93.8%, and it was 97.6% for the frequency domain. The AP obtained using the PR curve in [32] was 84.2%. Therefore, compared to the previously proposed stress signal classifier, the AP of the PR curve is improved by 13.4% using the proposed classifier. In addition, the AP value of the frequency domain was 3.8% higher than that of the time domain in our model, indicating that the stress classification performance is better in the former.

In previous studies using the time domain or frequency domain of ECG data, the epochs were set to 10, and the batch size was set to 64. As a result, the time domain and frequency domain accuracies were 83.6% and 74.5%, respectively [33]. However, the architectures used are susceptible to overfitting, and the accuracies achieved after 10 epochs may reflect this problem. Figure 11 shows the accuracy of stress classification using the proposed CNN-LSTM. After setting the epochs to 20 and the batch size to 64, the classification accuracies involving ECG stress signals in the time and frequency domains were measured. Under these settings, the time required for the time domain classification was 7 min 48 s and the verification accuracy was 94.13%. The elapsed time for the frequency domain was 7 min 31 s and the verification accuracy was 98.26%, which represents 10.53% and 23.76% improvements in accuracy compared to previous results [33].

For comparison purposes, we evaluated the stress classification performance of the CNN, LSTM, and CNN-LSTM models. First, stress signals were classified using CNN. After inputting the time series data values from the DBs into the image input layer, feature maps were extracted using

convolutional, batch normalization, and max pooling layers. The stress was classified using a fully connected layer and a softmax layer under stress and without stress as the final classification. The classification accuracy of the stress signals using CNN was 88.35%.

In addition, stress signals were classified using LSTM. LSTM is a type of recurrent neural network (RNN), which is an artificial neural network that recognizes patterns in data that can be represented as an array and is used for tasks such as text and gene signal analysis. After inputting the sequence data of the ECG DBs into the sequence input layer, the output was calculated using the LSTM layer (with the ReLU activation function). The signal was then classified as under stress or without stress using a fully connected layer. The classification accuracy of the stress signals using LSTM was 86.25%.

Table 3 compares the stress classification accuracies of the CNN, LSTM, and CNN-LSTM models. We set the epochs to 20 and the batch size to 64 and then determined the elapsed time and accuracy. The results confirmed that the CNN-LSTM model was approximately 1 min faster than the CNN and LSTM models, and accuracy was improved by 9.91% and 12.01%, respectively.

Figure 12 shows the AUC and AP curves for each model based on the ROC and PR results. The AUC of CNN-LSTM was 98.12%, while those of CNN and LSTM were 87.5% and 84.3%, respectively. Therefore, the AUC of the CNN-LSTM model was 10.62% and 13.82% higher than that of the CNN and LSTM models, respectively, confirming that its stress classification performance is better. The AP of CNN-LSTM was 97.6%, and it was 88.2% and 86.02%, respectively, for CNN and LSTM. The CNN-LSTM model achieved AP values that were 9.4% and 11.58% higher than the CNN and LSTM models, respectively, further confirming improved classification performance.

4. Discussion

In this study, to improve the performance of stress classification and prevent overfitting, an optimized ensemble model was developed by generating additional data using spectrograms and adding layers such as batch normalization, a flattening layer, and a fully connected layer. The performance of the classifier was evaluated using a confusion matrix, ROC, and other measures. By applying the average value of the $R - S_{\text{peak}}$ of the ECG signal, the characteristics of under-stress and without-stress signals are extracted to improve the stress classification accuracy. In the time domain, a precision of 93.1%, a sensitivity of 96.4%, and a specificity of 93.3% were achieved. In the frequency domain, a precision of 96.6%, a sensitivity of 100%, and a specificity of 96.7% were achieved. The CNN-LSTM achieved 94.8% accuracy for time domain signals and 98.3% accuracy for frequency domain signals. The best stress classification accuracy of the proposed CNN-LSTM algorithm is 98.3%, which is approximately 14.7% higher than the best

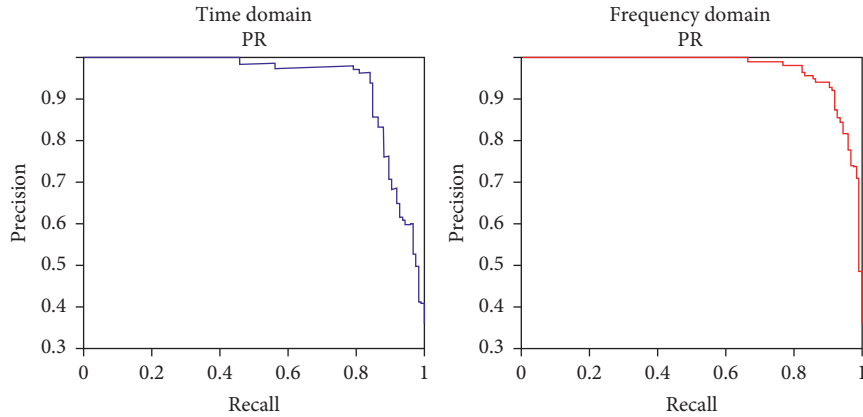


FIGURE 10: Evaluation of classification performance using PR curve.

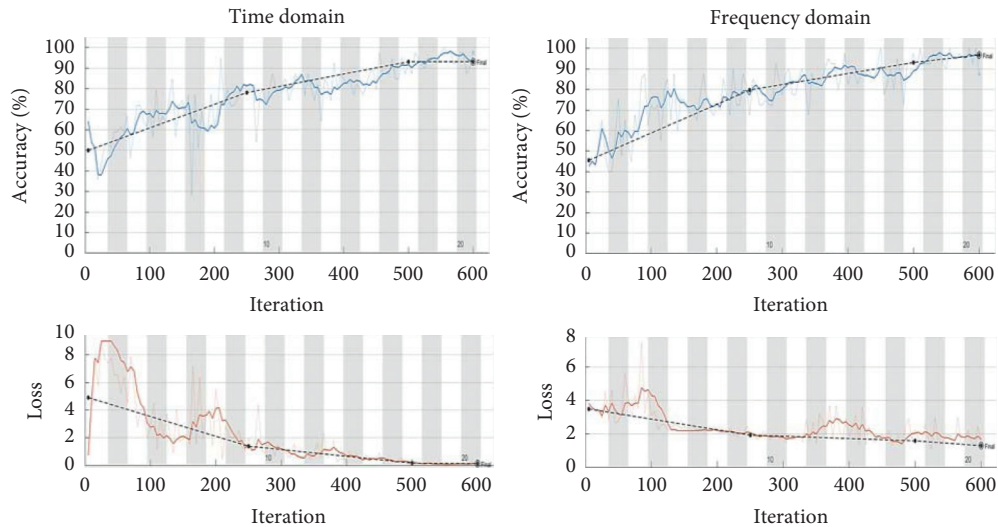


FIGURE 11: Time and frequency domain stress signal classification.

TABLE 3: Classification accuracy comparison of stress signals using CNN, LSTM, and CNN-LSTM.

	CNN	LSTM	CNN-LSTM
Elapsed time	8 min 32 s	8 min 45 s	7 min 31 s
Accuracy	88.35%	86.25%	98.26%

accuracies reported in previous studies. The proposed stress classifier achieves optimal stress signal classification performance when the number of epochs is 219 in the time domain and 223 in the frequency domain. In addition, the

model’s performance was evaluated using ROC and PR curves. It was confirmed that improvements of 12.42% and 13.4%, respectively, were obtained compared to previous study results.

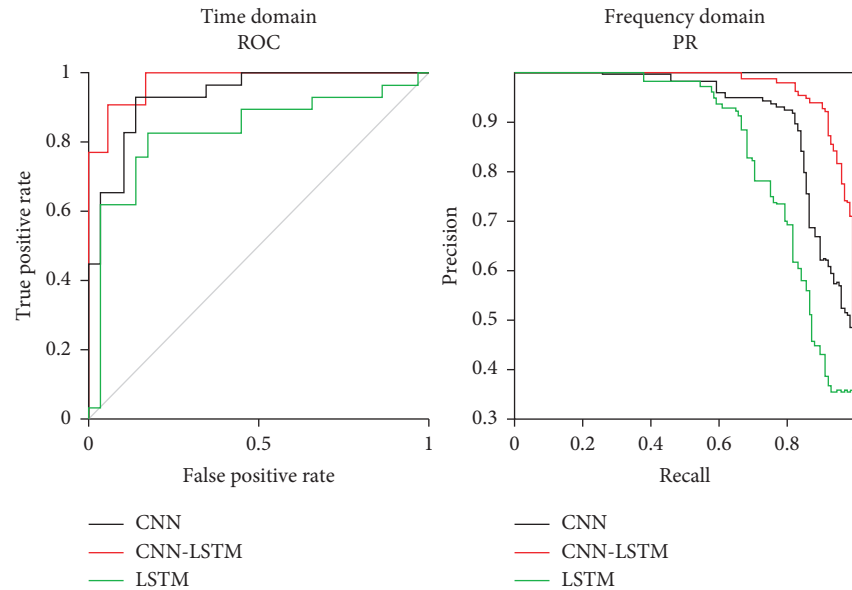


FIGURE 12: Evaluation of classification performance of stress signals using ROC and PR curves.

5. Conclusions

In this study, we proposed an improved ensemble model based on CNN-LSTM to accurately classify stress states. To prevent the overfitting of the algorithm and improve the accuracy of the classifier, ECG signals were classified separately in the time domain and frequency domain. The proposed ensemble model achieved a stress classification accuracy of 98.3%. These results exhibit an approximate 14.7% improvement in accuracy compared to earlier studies that classify the existing under stress and without stress. In the future, we plan to improve the preprocessing method, such as a subtle noise removal of biological signals, and to improve accuracy by applying a wearable transform filter that will remove baseline fluctuations and noise using Fourier transforms. The stress classifier proposed by us is expected to be helpful in mental health management as it can quickly and accurately classify the stress experienced by modern people. It is also expected to assist in preventing various diseases such as depression, high blood pressure, and diabetes through periodic stress management.

Data Availability

The data are available at <https://physionet.org/content/stdb/1.0.0/>, <https://archive.ics.uci.edu/ml/datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29>.

Conflicts of Interest

The authors declare no conflicts of interest in this research.

Acknowledgments

This research was supported by the Healthcare AI Convergence R&D Program through the National IT Industry Promotion Agency of Korea (NIPA) funded by the Ministry

of Science and ICT (No. S1601-20-1041) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2018R1A6A1A03015496).

References

- [1] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, vol. 2019, Article ID 2927337, 1 page, 2019.
- [2] B. S. McEwen and E. Stellar, "Stress and the individual. Mechanisms leading to disease," *Archives of Internal Medicine*, vol. 153, no. 18, pp. 2093–2101, 1993.
- [3] M. N. Dar, M. U. Akram, S. G. Khawaja, and A. N. Pujari, "CNN and LSTM-based emotion charting using physiological signals," *Sensors*, vol. 20, no. 16, pp. 4551–4552, 2020.
- [4] Y. Jammes, J. G. Steinberg, O. Mambrini, F. Bregeon, and S. Delliaux, "Chronic fatigue syndrome: assessment of increased oxidative stress and altered muscle excitability in response to incremental exercise," *Journal of Internal Medicine*, vol. 257, no. 3, pp. 299–310, 2005.
- [5] S. Elzeiny and M. Qaraq, "Stress classification using photoplethysmogram-based spatial and frequency domain images," *Sensors*, vol. 17, pp. 312–313, 2020.
- [6] A. Arsalan, "Classification of perceived human stress using physiological signals," in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1247–1250, IEEE, Berlin, Germany, July 2019.
- [7] S. M. U. Saeed, S. M. Anwar, H. Khalid, M. Majid, and U. Bagci, "EEG based classification of long-term stress using psychological labeling," *Sensors*, vol. 7, pp. 886–887, 2020.
- [8] S. M. U. Saeed, S. M. Anwar, and M. Majid, "Psychological stress measurement using low cost single channel EEG headset," in *Proceedings of the 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 581–585, IEEE, Abu Dhabi, UAE, December 2015.

- [9] R. Luijckx, "Experimentally induced stress validated by EMG activity," *PLoS One*, vol. 9, pp. 215–216, 2014.
- [10] P. Zontone, A. Affanni, R. Bernardini et al., "Car driver's sympathetic reaction detection through electrodermal activity and electrocardiogram measurements," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 12, pp. 3413–3424, 2020.
- [11] G. Giannakakis, K. Marias, and M. Tsiknakis, "A stress recognition system using HRV parameters and machine learning techniques," in *Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 269–272, IEEE, Cambridge, UK, September 2019.
- [12] S.-Y. Ye and D.-H. Kim, "Evaluation on the stress using HRV according to elapsed time of MRI noise," *Journal of the Institute of Convergence Signal Processing*, vol. 16, pp. 50–55, 2015.
- [13] S.-Y. Dong, "Stress resilience measurement with heart-rate variability during mental and physical stress," in *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5290–5293, IEEE, Honolulu, HI, USA, July 2018.
- [14] H.-M. Cho, H. Park, S.-Y. Dong, and I. Youn, "Ambulatory and laboratory stress detection based on raw electrocardiogram signals using a convolutional neural network," *Sensors*, vol. 19, no. 20, pp. 4408–4410, 2019.
- [15] B. Hwang, J. You, T. Vaessen, I. Myin-Germeys, C. Park, and B.-T. Zhang, "Deep ECGNet: an optimal deep learning framework for monitoring mental stress using ultra short-term ECG signals," *Telemedicine and E-Health*, vol. 24, no. 10, pp. 753–772, 2018.
- [16] B. S. Zheng, M. Murugappan, and S. Yaacob, "FCM clustering of emotional stress using ECG features," in *Proceedings of the 2013 International Conference on Communication and Signal Processing*, pp. 305–309, IEEE, Melmaruvathur, India, April 2013.
- [17] J. He, K. Li, X. Liao, P. Zhang, and N. Jiang, "Real-time detection of acute cognitive stress using a convolutional neural network from electrocardiographic signal," *IEEE Access*, vol. 7, pp. 42710–42717, 2019.
- [18] P. Zontone, A. Affanni, R. Bernardini, L. D. Linz, A. Piras, and R. Rinaldo, "Supervised learning techniques for stress detection in car drivers," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 22–29, 2020.
- [19] Z. Zhao, L. Yang, D. Chen, and Y. Luo, "A human ECG identification system based on ensemble empirical mode decomposition," *Sensors*, vol. 13, no. 5, pp. 6832–6864, 2013.
- [20] P. Schmidt, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 400–408, Boulder, CO, USA, October 2018.
- [21] V. Bhateja, S. Urooj, R. Mehrotra, R. Verma, A. Lay-Ekuakille, and V. D. Verma, "A composite wavelets and morphology approach for ECG noise filtering," in *Proceedings of the International Conference on Pattern Recognition and Machine Intelligence*, pp. 361–366, Springer, Tezpur, India, December 2013.
- [22] B. Porr and L. Howell, "R-peak detector stress test with a new noisy ECG database reveals significant performance differences amongst popular detectors," *BioRxiv*, vol. 4, pp. 722397–722399, 2019.
- [23] R. Laszlo, H. J. Weig, S. Weretka, and J. Schreieck, "Narrow complex tachycardia with alternating R-R intervals during physical stress: double ventricular excitation," *Indian Pacing and Electrophysiology Journal*, vol. 8, pp. 129–132, 2008.
- [24] C. I. Garcia, F. Grasso, A. Luchetta, M. C. Piccirilli, L. Paolucci, and G. Talluri, "A comparison of power quality disturbance detection and classification methods using CNN, LSTM and CNN-LSTM," *Applied Sciences*, vol. 10, no. 19, pp. 6755–6757, 2020.
- [25] S. B. Shafiei, L. Zaeem, S. E. Ahmed, A. H. Ahmed, and A. G. Khurshid, "Identifying mental health status using deep neural network trained by visual metrics," *Translational Psychiatry*, vol. 10, pp. 1–8, 2020.
- [26] T. Marwick, A.-M. D'hondt, T. Baudhuin et al., "Optimal use of dobutamine stress for the detection and evaluation of coronary artery disease: combination with echocardiography or scintigraphy, or both?" *Journal of the American College of Cardiology*, vol. 22, no. 1, pp. 159–167, 1993.
- [27] D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData Mining*, vol. 14, pp. 1–22, 2021.
- [28] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalanced data classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1806–1819, 2017.
- [29] K. Behfarnia and F. Khademi, "A comprehensive study on the concrete compressive strength estimation using artificial neural network and adaptive neuro-fuzzy inference system," *Iran University of Science & Technology*, vol. 7, pp. 71–80, 2017.
- [30] D. J. Vining and G. W. Gladish, "Receiver operating characteristic curves: a basic understanding," *Radiographics*, vol. 12, no. 6, pp. 1147–1154, 1992.
- [31] M. Mounir, *Acoustic Event Detection: Feature, Evaluation and Dataset Design*, pp. 1–171, Ku Leuven, Leuven, Belgium, 2020.
- [32] S. Kumar, "StressNet: detecting stress in thermal videos," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 999–1009, Waikola, HI, USA, April 2021.
- [33] Z. Ahmad and M. K. Naimul, "Multi-level stress assessment using multi-domain fusion of ECG signal," in *Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4518–4521, IEEE, Montreal, CA, USA, July 2020.

Research Article

Reduce Surface Electromyography Channels for Gesture Recognition by Multitask Sparse Representation and Minimum Redundancy Maximum Relevance

Yali Qu , Haoyan Shang , Jing Li , and Shenghua Teng 

College of Electronic and Information Engineering, Shandong University of Science and Technology, Qingdao, China

Correspondence should be addressed to Jing Li; peilianglj@163.com and Shenghua Teng; shteng@sdust.edu.cn

Received 18 March 2021; Revised 9 May 2021; Accepted 21 May 2021; Published 28 May 2021

Academic Editor: G R Sinha

Copyright © 2021 Yali Qu 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.

Surface electromyography- (sEMG-) based gesture recognition is widely used in rehabilitation training, artificial prosthesis, and human-computer interaction. The purpose of this study is to simplify the sEMG devices by reducing channels while achieving comparably high gesture recognition accuracy. We propose a compound channel selection scheme by combining the variable selection algorithms based on multitask sparse representation (MTSR) and minimum Redundancy Maximum Relevance (mRMR). Specifically, channelwise features are first extracted to compose channel-feature paired variables, for which variable selection procedures by MTSR and mRMR are carried out, respectively. Then, we rank all the channels according to their occurrences in each variable selection procedure and figure out a certain number of informative channels by fusing these rankings of channels. Finally, the gesture classification performance using the selected channels is evaluated by the support vector machine (SVM) classifier. Experiment results validate the effectiveness of this proposed method.

1. Introduction

Surface electromyography (sEMG) is commonly used in clinical and engineering areas with the advantages of being noninvasive and convenient in signal acquisition. For example, sEMG reveals the information in diagnosing neuromuscular disorders [1, 2]. More generally, it may play important roles in the controlling of artificial assistance robots, arm prostheses, rehabilitation equipment, and some other instruments [3, 4].

Most of the related works have been carried out with sEMG of multiple channels to guarantee satisfactory recognition performance [5]. However, the increase of channels makes not only a high cost in engineering but also the great complexity of the sEMG devices and data processing burden. In addition, it could suffer from performance deterioration due to signal crosstalk [6, 7]. To overcome these problems due to multiple channels of sEMG, it is rewarding to select a reduced group of channels in a myoelectric control system.

This is just the aim of our work which is to simplify the sEMG device by removing some redundant electrodes on the premise of desired classification performance.

2. Related Research and Motivation

Feature extraction is a routine procedure to describe the sEMG signals with a feature vector. Multitudinous features of time domain, frequency domain, and time-frequency domain have been widely applied in sEMG-based classification tasks. When multiple features are extracted for channels one by one, we could get a feature set with a quite large size (the number of features per channel times the number of channels). Hence, feature selection can be followed to reduce the feature redundancy and alleviate the curse of dimensionality, where metrics including scatter plot of features, statistical analysis, and recognition rate are applied to evaluate the effectiveness of features [8, 9], and feature search strategies including sequential forward

selection (SFS), sequential backward selection (SBS), or bidirectional searching are adopted to find out the most informative features [10].

Like feature selection in the point of lowering the feature size, channel selection will, in addition, remove those channels unnecessary or irrelevant to classify different gestures. In fact, channel selection is highly related to feature selection since features coming from all the channels are generally combined to create a set of channel-feature paired variables. Hence, channel selection can be the successor operation after feature selection, using the selected or fixed features.

To select useful channels from multielectrode, Nagata and his colleagues [11] used the recognition rate to evaluate each measurement channel and found out the best combination of channels by the Monte Carlo method. Huang et al. [12] applied SFS search strategy for expected channels where four kinds of time-domain features and an LDA classifier are used in the searching iteration. Khushaba and Al-Jumaily [13] also adopted a wrapper method, particle swarm optimization, in channel selection where the importance of subsets was measured using the error rates acquired from a multilayer perceptron trained with back-propagation neural network. Similar work by Oskoei et al. [14] employed a multiobjective genetic searching algorithm with the objective function of data separability index or classification rate. Besides, filter methods have also been applied to rank the channels, where the minimum Redundancy Maximum Relevance (mRMR) [15] was used by Liu et al. [16] and Gupta et al. [17], the Relief-F by Qu et al. [18], and the Markov random field (MRF) by Qu et al. [16] as well.

As shown in these aforementioned pieces of literature, channel selection could be conducted by fixing the feature subset. That means we cannot simultaneously select the best features and channels, which can be improved in the way as follows. Features and channels are combined to construct feature-channel pairs, leading to a hybrid feature-channel selection problem. By finding the least redundant and most informative group of feature-channel pairs among all the possible ones, the best channels should be the most repeated ones. In these aspects, some classic or modified ranking methods have been applied to select channel-feature variables, such as mRMR-FCO [19] and certain correlation-based or distance-based evaluation function in the work by Al-Angari et al. [20].

Channel selection can follow a feature-channel filtering pipeline, but differing in specific ranking scores or search strategies. Our work is just under this kind of framework where we resort to the multitask sparse learning [21] together with mRMR filtering to pursue the discriminative sEMG channels across the classification for multiple gestures.

Since the classic least square regression model in sparse learning does not pursue the class-discriminative power of features, certain type of discriminative regularization terms is preferred to make up this limitation. Zhu et al. [22] put forward a group-sparsity-based least square regression framework integrating linear discriminant analysis and

locality preserving projection. Similarly, to better capture the discriminative information among subjects, a multitask feature selection method was proposed to incorporate the intraclass and interclass Laplacian matrices [23]. But this kind of work will generally lead to a complicated optimization problem and most likely suffer from heavy computation cost.

Inspired by the works related to multitask sparse learning, for channel selection, we propose a channel selection method that combines the multitask sparse representation (MTSR) and mRMR algorithms. Instead of superimposing discriminative regularization terms in the MTSR framework, we evaluate the sEMG channels using the MTSR and mRMR, respectively, and then fuse their results to figure out the ideal channels in the end. The flowchart of this paper is shown in Figure 1.

3. Methods

3.1. Dataset and Evaluation Metrics. The sEMG dataset [24] contains thirty healthy normal-limbed subjects, who were kept relaxed and performed 7 distinct hand gestures including hand open, hand close, supination, pronation, wrist flexion, wrist extension, and rest. Eight surface electrodes were used for sEMG acquisition. In other words, we have signals with eight channels.

In this work, three classic measures, that is, precision, recall, and accuracy, are selected as indicators to evaluate the performance of gesture classification. These metrics are defined as follows:

$$\text{precision} = \frac{TP}{TP + FP}, \quad (1)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (2)$$

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (3)$$

where TP, FP, TN, and FN are True Positive, False Positive, True Negative, and False Negative, respectively. An average of classification metrics in the experiments below will be obtained by 5-fold cross validation.

3.2. Feature Extraction. To analyze the sEMG signal, a sliding window is adopted for the 8 channels. Totally 11 time-domain features, as listed in Table 1, are extracted which have been proved effective for myoelectric pattern recognition [16]. Thus, we have channel-feature paired variables with the size of 8 times 11.

L is the signal length, and x_i is the signal in an analysis window. SD is the standard deviation. p is the order of autoregressive model, ε_i is a white noise term, and the coefficients a_p are used as features.

3.3. Channel Selection Scheme. This study aims to reduce the sEMG channels by finding the least and best electrode locations to discriminate different hand motions. For channel-feature variables, we first perform a composite variable

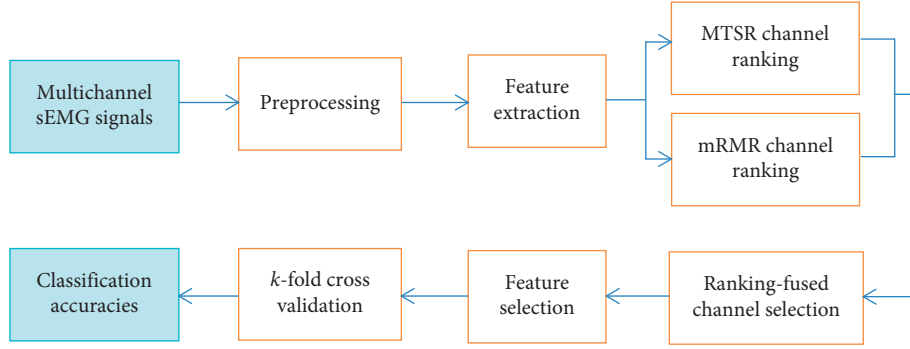


FIGURE 1: Gesture recognition by reducing sEMG channels.

TABLE 1: Features extracted for each analysis window.

Acronym	Name	Number of features	Formula
WL	Waveform length	1	$WL = \sum_{i=1}^{L-1} x_{i+1} - x_i $
IAV	Integrated absolute value	1	$IAV = \sum_{i=1}^L x_i $
RMS	Root mean square	1	$RMS = \sqrt{\frac{1}{L} \sum_{i=1}^L x_i^2}$
SSI	Simple square integral	1	$SSI = \sum_{i=1}^L x_i ^2$
Kurtosis	Kurtosis	1	$Kurtosis = \frac{1}{L-1} \sum_{i=1}^L ((x_i - \bar{x})^4 / SD^4) - 3$
Skewness	Skewness	1	$Skewness = \frac{1}{L-1} \sum_{i=1}^L ((x_i - \bar{x})^3 / SD^3)$
ZC	Zero crossing (threshold $T=10$)	1	$ZC = \sum_{i=1}^{L-1} \varphi(x_i, x_{i+1})$ $\varphi(x_i, x_{i+1}) = \begin{cases} 1, & x_i x_{i+1} < 0, x_i - x_{i+1} > T, \\ 0, & \text{otherwise,} \end{cases}$
4AR	4th-order autoregressive model	4	$x_i = \sum_{p=1}^4 a_p x_{i-p} + \varepsilon_i$

selection for the task of gesture motion recognition. Then, all channels will be ranked according to their occurrences in the selection of channel-feature variables, where MTSR and the mRMR variable ranking method are used, respectively. By fusing these two ranking results, we can finally get the ideal channels but with high recognition capability for hand gesture motions.

3.3.1. MTSR-Based Variable Selection. Given a feature matrix $X \in R^{d \times n}$, where d and n are the numbers of features and samples, respectively, we also have a class indicator matrix $Y \in R^{c \times n}$ with the class number c . Since multiple response variables are included in the class indicator matrix Y , for each response variable, we can find a regression coefficient vector individually. By regularizing a least square regression model with an $\ell_{2,1}$ -norm, the multiclass feature selection problem can be formulated as a sparse least square regression model as follows [21]:

$$\min_W \frac{1}{2} \|Y - W^T X\|_F^2 + \lambda \|W\|_{2,1}, \quad (4)$$

where $W \in R^{d \times c}$ is a coefficient matrix for regression and the parameter λ is adopted to adjust the sparsity of W . By enforcing the group sparsity on the coefficient matrix with a $\ell_{2,1}$ -norm, some rows in W will be zero. The first term in equation (4) controls the data fitting error, and the regularization parameter λ balances the relative importance of

both terms. The larger λ results in more zero rows in the coefficient matrix. It can be assumed that the optimal solution would assign large weights to the important features and zero or small weights to the less important features.

3.3.2. mRMR Variable Ranking. The above MTSR method mainly focuses on the relationship between labels and features but ignores the relationship between features to some extent. Hence, we resort to mRMR algorithm to select features from a different perspective.

The mRMR criteria [15] aim to choose features that are mutually dissimilar to each other and marginally similar to the classification labels, ranking candidate component features based on compromise between relevance and redundancy. In this paper, we use mutual information to measure both redundancy and relevance.

Mutual information is defined as follows:

$$I(X; Y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (5)$$

where X and Y denote two feature vectors and $p(x, y)$ is the joint probabilistic density, while $p(x)$ and $p(y)$ are the marginal probabilistic densities. The goal is to find a subset S with m features, and the maximum relevance and the minimum redundancy are defined by equations (6) and (7):

$$\max D(S, c),$$

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c), \quad (6)$$

$$\min R(S),$$

$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j), \quad (7)$$

where x_i is the i -th feature, c is the class variable, and S is the feature subset. The maximum relevance and the minimum redundancy are integrated by equation (8) or (9).

$$\max \Phi(D, R), \quad (8)$$

$$\Phi = D - R,$$

$$\max \Phi(D, R), \quad (9)$$

$$\Phi = \frac{D}{R}.$$

The incremental search method is used to find the approximate optimal feature. Supposing that we already have the feature set S_{m-1} , the next step is to find the m -th feature from the feature set $X - S_{m-1}$ maximizing $\Phi(\cdot)$. The incremental algorithm optimizes the formula [15]

$$\max_{x_j \in X - S_{m-1}} \left[I(x_j; c) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I(x_j; x_i) \right]. \quad (10)$$

3.4. To Fuse the Channel Rankings. As stated above, we successively select the effective channel-feature pairs by MTSR and mRMR ranking method. Thus, we can get two groups of rankings for all channels according to their occurrences in the screened channel-feature variables. These two channel ranking methods work with different principles, but their corresponding results share common informative components even if they differ to a certain extent. We combine the two channel ranking results in the hope of avoiding decision faults to the utmost extent.

4. Results

4.1. Channel Selection. Considering that 11 features are extracted for 8 channels each, we have 88 channel-feature paired variables in total for each analysis window. We apply the multiclass sparse representation model for the training data. According to equation (4), the parameter λ controls the sparsity of the coefficient matrix W , namely, the number of the screened channel-feature variables. The gesture recognition performance would be affected by features and classifiers we employed.

Let λ varies from 0.01 to 0.1, and channel-feature variables corresponding to nonzero rows of the coefficient matrix W are kept and fed to a support vector machine (SVM) classifier with radial basis function [25]. We hope to

achieve a high recognition rate (accuracy is used in Section 4.1 and 4.3) while using only a few feature variables.

We make a comparison to show how to decide a proper value for λ . When λ varies from 0.01 to 0.1, the screened channel-feature number varies greatly but the recognition rate does not decrease too much. The changing of recognition rate and channel-feature number along with λ is shown in Figure 2. We can also see that a good balance between the recognition rate and channel-feature dimension can be achieved when λ equals 0.03. Accordingly, we will keep 36 channel-feature variables in the following channel selection procedure.

And for mRMR-based channel-feature selection, we also keep the top 36 variables which will be fused with the results of MTSR.

Table 2 lists the selected 36 channel-feature variables (features for each channel) by MTSR and mRMR, respectively. It is obvious that there is a certain difference between the screened results by these two methods. For instance, autoregressive features AR1 and AR2 play important roles in MTSR modal, being used by most channels. However, for mRMR, the two features only appear in channel ⑧. Therefore, we select channels based on channel utilization rather than analyzing the features. We count the number of times that any two channels occupy a common feature, namely, the number of features shared by a channel pair. The more frequently a channel is utilized, the more important the channel will be. The corresponding statistical results for MTSR and mRMR are shown in Tables 3 and 4.

From Tables 3 and 4, we sort channels by the number of times which are used. For MTSR, the order is ② > ③ = ⑧ > ⑤ > ⑦ > ① = ④ > ⑥ and ① = ⑤ > ⑧ > ③ > ⑦ > ② = ⑥ > ④ for mRMR. By decision-making level fusion for channel selection, three channels ③, ⑤, and ⑧ are adopted for the subsequent gesture recognition.

4.2. Feature Selection. Also based on the screened channel-feature variables by MTSR and mRMR, we list all the channels occupying a given feature (shown in Table 5). If a feature is shared by over half channels (>4), it will be selected for the gesture recognition task. Specifically, we have WL, AR1, and AR2 from MTSR-based results, and WL, IAV, SSI, and Kurtosis by mRMR. These six features, WL, IAV, SSI, Kurtosis, AR1, and AR2, will be fed into classifier in the following experiments.

4.3. Classification Performance Based on Channel and Feature Selection. According to Section 4.1, three channels (③, ⑤, and ⑧) are jointly selected by fusing MTSR and mRMR. We first compare the gesture classification performance using these three channels with those by MTSR or mRMR individually. For MTSR-based results, the top three channels are ②, ③, and ⑧, and the three channels ①, ⑤, and ⑥ are for mRMR. Their corresponding gesture recognition accuracies are shown in Figure 3. By combining MTSR and mRMR, channels ③, ⑤, and ⑧ are used and the average recognition rate is 98.68%, which is higher than that using channels ②, ③, and ⑧ or ①, ⑤, and ⑥ (the average classification

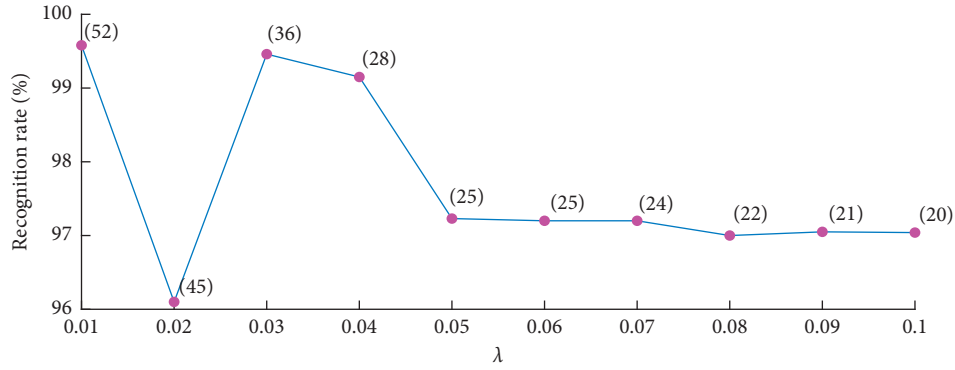
FIGURE 2: Changes of recognition rate and channel-feature number when λ varies from 0.01 to 0.1.

TABLE 2: The selected 36 channel-feature variables by MTSR and mRMR. AR1~AR4 are four coefficients in the fourth-order autoregressive model, respectively.

Channel	Features by MTSR	Features by mRMR
①	RMS, AR1, AR2, and AR4	WL, IAV, Kurtosis, SSI, and AR3
②	WL, Skewness, AR1, AR2, and AR3	WL, IAV, Kurtosis, and Skewness
③	WL, IAV, AR1, AR2, and AR3	WL, IAV, Kurtosis, and SSI
④	WL, IAV, AR1, and AR4	WL, IAV, and Kurtosis
⑤	RMS, Kurtosis, AR1, AR2, AR3, and AR4	WL, IAV, Kurtosis, SSI, and AR3
⑥	WL and Skewness	WL, IAV, SSI, and AR3
⑦	IAV, Skewness, AR1, and AR2	WL, IAV, Kurtosis, and AR3
⑧	WL, SSI, Kurtosis, Skewness, AR1, and AR2	WL, IAV, SSI, Kurtosis, Skewness, AR1, and AR2

TABLE 3: The number of times that two given channels occupy a common feature by MTSR (e.g., channel ① and channel ② share 2 features: AR1 and AR2). The best channels are in *italics*.

	Channel ①	Channel ②	Channel ③	Channel ④	Channel ⑤	Channel ⑥	Channel ⑦	Channel ⑧
Channel ①	—	2	2	2	4	0	2	2
Channel ②	2	—	4	2	3	2	3	4
Channel ③	2	4	—	3	3	1	3	3
Channel ④	2	2	3	—	2	1	2	2
Channel ⑤	4	3	3	2	—	0	2	3
Channel ⑥	0	2	1	1	0	—	1	2
Channel ⑦	2	3	3	2	2	1	—	3
Channel ⑧	2	4	3	2	3	2	3	—
Sum	14	20	19	14	17	7	16	19

TABLE 4: The number of times that two given channels occupy a common feature by mRMR (e.g., channel ① and channel ② share 3 features: WL, IAV, and Kurtosis). The best channels are in *italics*.

	Channel ①	Channel ②	Channel ③	Channel ④	Channel ⑤	Channel ⑥	Channel ⑦	Channel ⑧
Channel ①	—	3	4	3	5	4	4	4
Channel ②	3	—	3	3	3	2	3	4
Channel ③	4	3	—	3	4	3	3	4
Channel ④	3	3	3	—	3	2	3	2
Channel ⑤	5	3	4	3	—	4	4	4
Channel ⑥	4	2	3	2	4	—	3	3
Channel ⑦	4	3	3	3	4	3	—	3
Channel ⑧	4	4	4	3	4	3	3	—
Sum	27	21	24	20	27	21	23	25

TABLE 5: Channels used by each feature.

Feature	Channels by MTSR	Selected	Channels by mRMR	Selected
WL	②③④⑥⑧	√	①②③④⑤⑥⑦⑧	√
IAV	③④⑦		①②③④⑤⑥⑦⑧	√
RMS	①⑤		-	
SSI	⑧		①③⑤⑥⑧	√
Kurtosis	⑤⑧		①②③④⑤⑦⑧	√
Skewness	②⑥⑦⑧		②⑧	
ZC	-		-	
AR1	①②③④⑤⑦⑧	√	⑧	
AR2	①②③⑤⑦⑧	√	⑧	
AR3	②③⑤		①⑤⑥⑦	
AR4	①④⑤		-	

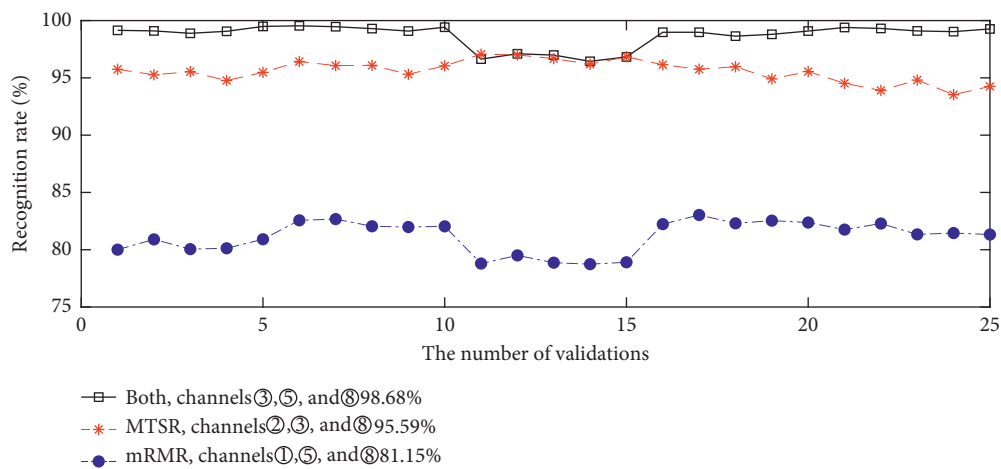


FIGURE 3: The classification accuracies using three channels selected by different methods (channels ②, ③, and ⑧ are selected by MTSR, channels ①, ⑤, and ⑧ are selected by mRMR, and channels ③, ⑤, and ⑧ are jointly selected by the two methods).

accuracy is 95.59% for channels ②, ③, and ⑧ and 81.15% for channels ①, ⑤, and ⑧).

In addition, comparative experiments for gesture classification are carried out using two or four channels selected by different methods. When choosing two channels, we have channels ⑤ and ⑧ by fusing MTSR and mRMR. For MTSR-based method, the top two channels are ② and ③ or ② and ⑧; for mRMR, the selected two channels are ① and ⑤. The gesture classification accuracies are illustrated in Figure 4, where channels selected by jointly using MTSR and mRMR achieve the highest classification accuracy.

As for choosing four channels, channels ③, ⑤, ⑦, and ⑧ are selected by fusing MTSR and mRMR. For MTSR-based method, the top four channels are ②, ③, ⑤, and ⑧; for mRMR, the four channels are ①, ③, ⑤, and ⑧. Correspondingly, the gesture classification accuracies are drawn in Figure 5. It also verifies that channels selected by jointly using MTSR and mRMR achieve the highest classification accuracy.

4.4. Performance Evaluation and Comparison. To evaluate the performance of our method by fusing MTSR and mRMR for channel selection, comparative experiments are conducted in two aspects. Firstly, we further compare the proposed method with MTSR and mRMR in the task of channel selection. For the number of selected channels varying from 2 to 4, precision and recall for gesture classification corresponding to different method are listed in Table 6 where the selected channels are in square brackets.

Compared with only 2 channels used, the recognition performance improves significantly when 3 channels are selected. It reveals that even 2 informative channels cannot capture enough information to distinguish different hand gestures in the experiment, where the best combination of 2 channels [5 8] is picked out by the proposed method. With more channels added in a certain range, the recognition performance will increase overall. In all cases, as shown in the table, our MTSR- and mRMR-fused methods outperform each of the two base methods alone.

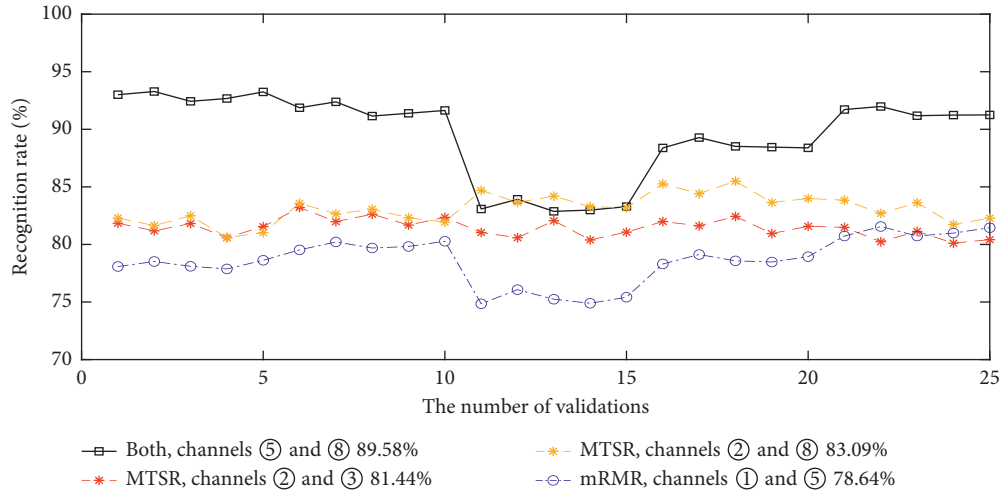


FIGURE 4: The classification accuracies using three channels selected by different methods (channels ② and ③ and channels ② and ⑧ are selected by MTSR, channels ① and ⑤ are selected by mRMR, and channels ⑤ and ⑧ are jointly selected by the two methods).

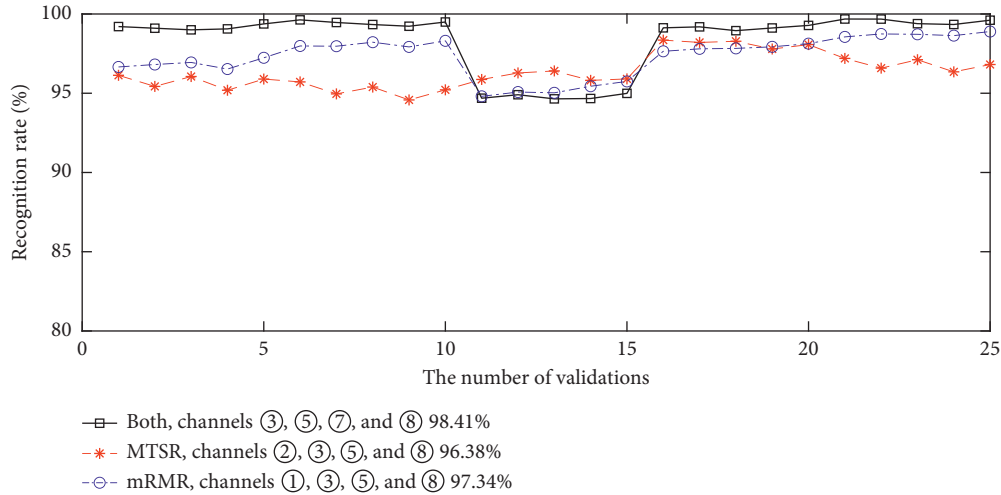


FIGURE 5: The classification accuracies using three channels selected by different methods (channels ②, ③, ⑤, and ⑧ are selected by MTSR, channels ①, ③, ⑤, and ⑧ are selected by mRMR, and channels ③, ⑤, ⑦, and ⑧ are jointly selected by the two methods).

TABLE 6: Comparison with different research methods.

Method	Channel	Recall	Precision	Channel	Recall	Precision	Channel	Recall	Precision
MTSR	[2 8]	83.69	84.59	[2 3 8]	95.30	95.60	[2 3 5 8]	96.82	97.18
mRMR	[1 5]	80.52	83.90	[1 5 8]	85.94	89.81	[1 3 5 8]	98.61	98.75
MRCS [18]	[1 5]	80.52	83.90	[1 5 7]	83.64	88.64	[1 3 5 7]	96.99	97.09
Our method	[5 8]	93.42	93.74	[3 5 8]	98.92	98.95	[3 5 7 8]	99.12	99.19

Besides, a latest work proposed a mean Relief-F-based channel selection method (MRCS) [18]. Under the same experimental conditions including dataset and features, its classification performance is shown as the third row in Table 6. As for selecting four channels, the channel combination [1 3 5 7] is obtained by MRCS, and the

corresponding classification rate is lower than our work here by selecting 3 channels or 4 channels. It should be noted that the classification performance can be further improved by using more informative features as demonstrated in the work [18], which will be our focus in the work later.

5. Conclusion

Recent developments in sEMG instrumentation have made it possible to record many channels from single or multiple muscles simultaneously. The current study combines MTSR and mRMR to process the channel-feature variables, aiming to reduce the channel number without degrading the gesture recognition performance.

For a gesture recognition task, sEMG dataset of 8 channels is recorded for 7 hand motions. Given the channel-features pairs obtained from time-domain features, the most informative channels are decided by the MTSR- and mRMR-combined variable selection method. The combination of MTSR and mRMR makes the selected variables not only reflect the relationship between labels and feature vectors but also try to meet the requirement of maximum relevance and minimum redundancy between vectors. Experimental results have verified the effectiveness of the proposed method.

It is worth noting that only time-domain features are extracted for sEMG signals in this paper. The channel selection operation is dependent on these features. More features generated in the frequency domain or time-frequency domain are to be used to test this feature/variable selection method in the coming work. In addition, this proposed method for feature selection can also be used in other pattern recognition and machine learning applications.

Data Availability

The pattern recognition library is available at <http://www.sce.carleton.ca/faculty/chan>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of Shandong Province (ZR2020MF086).

References

- [1] P. A. Kaplanis, C. S. Pattichis, and D. Zazula, "Multiscale entropy-based approach to automated surface EMG classification of neuromuscular disorders," *Medical & Biological Engineering & Computing*, vol. 48, no. 8, pp. 773–781, 2010.
- [2] S. Koçer, "Classification of EMG signals using neuro-fuzzy system and diagnosis of neuromuscular diseases," *Journal of Medical Systems*, vol. 34, no. 3, pp. 321–329, 2010.
- [3] N. Das, N. Nagpal, and S. S. Bankura, "A review on the advancements in the field of upper limb prosthesis," *Journal of Medical Engineering & Technology*, vol. 42, no. 7, pp. 532–545, 2018.
- [4] A. Krasoulis, I. Kyranou, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements," *Journal of NeuroEngineering and Rehabilitation*, vol. 14, no. 1, pp. 1–14, 2017.
- [5] M. Rojas-Martínez, M. A. Mañanas, and J. F. Alonso, "High-density surface EMG maps from upper-arm and forearm muscles," *Journal of NeuroEngineering & Rehabilitation*, vol. 9, no. 1, pp. 1–17, 2012.
- [6] L. Mesin, S. Smith, S. Hugo, S. Viljoen, and T. Hanekom, "Effect of spatial filtering on crosstalk reduction in surface EMG recordings," *Medical Engineering & Physics*, vol. 31, no. 3, pp. 374–383, 2009.
- [7] D. A. Winter, A. J. Fuglevand, and S. E. Archer, "Crosstalk in surface electromyography: theoretical and practical estimates," *Journal of Electromyography and Kinesiology*, vol. 4, no. 1, pp. 15–26, 1994.
- [8] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7420–7431, 2012.
- [9] N. M. Kakoty, S. M. Hazarika, and J. Q. Gan, "EMG feature set selection through linear relationship for grasp recognition," *Journal of Medical and Biological Engineering*, vol. 36, no. 6, pp. 883–890, 2016.
- [10] A. Phinyomark, R. N. Khushaba, and E. Scheme, "Feature extraction and selection for myoelectric control based on wearable EMG sensors," *Sensors*, vol. 18, no. 5, p. 1615, 2018.
- [11] K. Nagata, K. Ando, K. Magatani, and M. Yamada, "Development of the hand motion recognition system based on surface EMG using suitable measurement channels for pattern recognition," in *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 5214–5217, Lyon, France, August 2007.
- [12] H. Huang, P. Zhou, G. Li, and T. A. Kuiken, "An analysis of EMG electrode configuration for targeted muscle reinnervation based neural machine interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, vol. 16, no. 1, pp. 37–45, 2008.
- [13] R. N. Khushaba and A. Al-Jumaily, "Channel and feature selection in multifunction myoelectric control," in *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pp. 5182–5185, Lyon, France, August 2007.
- [14] M. A. Oskoei, o. Hu, and J. Q. Gan, "Feature-channel subset selection for optimising myoelectric human-machine interface design," *International Journal of Biomechanics and Biomedical Robotics*, vol. 2, no. 2/3/4, pp. 195–208, 2013.
- [15] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [16] J. Liu, X. Li, G. Li, and P. Zhou, "EMG feature assessment for myoelectric pattern recognition and channel selection: a study with incomplete spinal cord injury," *Medical Engineering & Physics*, vol. 36, no. 7, pp. 975–980, 2014.
- [17] R. Gupta, S. Saxena, and A. Sazid, "Channel selection in multi-channel surface electromyogram based hand activity classifier," in *Proceedings of the International Conference on computational Intelligence & Communication Technology*, pp. 1–7, Ghaziabad, India, February 2018.
- [18] Y. Qu, H. Shang, and S. Teng, "Reduce sEMG channels for hand gesture recognition," in *Proceedings of the 2020 IEEE 3rd International Conference on Information Communication and Signal Processing*, pp. 215–220, Shanghai, China, September 2020.
- [19] I. Mesa, A. Rubio, I. Tubia, J. De No, and J. Diaz, "Channel and feature selection for a surface electromyographic pattern

- recognition task,” *Expert Systems with Applications*, vol. 41, no. 11, pp. 5190–5200, 2014.
- [20] H. M. Al-Angari, G. Kanitz, S. Tarantino, and C. Cipriani, “Distance and mutual information methods for EMG feature and channel subset selection for classification of hand movements,” *Biomedical Signal Processing and Control*, vol. 27, pp. 24–31, 2016.
- [21] A. Argyriou, T. Evgeniou, and M. Pontil, “Convex multi-task feature learning,” *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.
- [22] X. Zhu, H.-I. Suk, and D. Shen, “Sparse discriminative feature selection for multi-class Alzheimer’s disease classification,” in *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, pp. 157–164, Cham, Switzerland, September 2014.
- [23] T. Ye, C. Zu, C. Zu, B. Jie, D. Shen, and D. Zhang, “Discriminative multi-task feature selection for multi-modality classification of Alzheimer’s disease,” *Brain Imaging and Behavior*, vol. 10, no. 3, pp. 739–749, 2016.
- [24] A. D. Chan and G. C. Green, “Myoelectric control development toolbox,” in *Proceedings of the 30th Conference of the Canadian Medical & Biological Engineering Society*, Toronto, Canada, June 2007.
- [25] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

Research Article

Optimizing Residual Networks and VGG for Classification of EEG Signals: Identifying Ideal Channels for Emotion Recognition

Kit Hwa Cheah ¹, Humaira Nisar ¹, Vooi Voon Yap ¹, Chen-Yi Lee ²,
and G. R. Sinha ³

¹Department of Electronic Engineering, Faculty of Engineering and Green Technology, Universiti Tunku Abdul Rahman, Kampar 31900, Malaysia

²Department of Electronics Engineering and Institute of Electronics, National Chiao Tung University, Hsinchu, Taiwan

³Myanmar Institute of Information Technology (MIIT), Mandalay, Myanmar

Correspondence should be addressed to Humaira Nisar; humaira@utar.edu.my

Received 3 February 2021; Accepted 19 March 2021; Published 30 March 2021

Academic Editor: Saverio Maietta

Copyright © 2021 Kit Hwa Cheah 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.

Emotion is a crucial aspect of human health, and emotion recognition systems serve important roles in the development of neurofeedback applications. Most of the emotion recognition methods proposed in previous research take predefined EEG features as input to the classification algorithms. This paper investigates the less studied method of using plain EEG signals as the classifier input, with the residual networks (*ResNet*) as the classifier of interest. *ResNet* having excelled in the automated hierarchical feature extraction in raw data domains with vast number of samples (e.g., image processing) is potentially promising in the future as the amount of publicly available EEG databases has been increasing. Architecture of the original *ResNet* designed for image processing is restructured for optimal performance on EEG signals. The arrangement of convolutional kernel dimension is demonstrated to largely affect the model's performance on EEG signal processing. The study is conducted on the Shanghai Jiao Tong University Emotion EEG Dataset (SEED), with our proposed *ResNet18* architecture achieving 93.42% accuracy on the 3-class emotion classification, compared to the original *ResNet18* at 87.06% accuracy. Our proposed *ResNet18* architecture has also achieved a model parameter reduction of 52.22% from the original *ResNet18*. We have also compared the importance of different subsets of EEG channels from a total of 62 channels for emotion recognition. The channels placed near the anterior pole of the temporal lobes appeared to be most emotionally relevant. This agrees with the location of emotion-processing brain structures like the insular cortex and amygdala.

1. Introduction

Emotion is the conscious or subconscious neuropsychological response generated upon external or internal stimuli which are of major concern to the person.

Emotion involves the interrelated synchronization of a number of organismic subsystems encompassing the central nervous system, the autonomous nervous system, the neuroendocrine system, the somatic nervous system, and the conscious or subconscious reactions of the respective effectors [1].

Expression of emotion includes the linguistic choice of words, speaking rate, intonation, facial expression, gesture, and posture. Emotion can also be reflected via the

autonomous nervous system and neuroendocrine system into the cardiovascular dynamics [2, 3], respiratory patterns [4], and electrodermal activity [5]. Nevertheless, all the peripheral emotion reactions arise from the neurological activities in the brain. The cerebral neuronal activities can be recorded as the electrical potentials on the scalp with the electroencephalography (EEG) technique [6].

Emotion recognition algorithms are useful in human-machine interaction, allowing machines to identify the emotional or affective mental states of humans [7]. Affective computing [8] and the “empathetic” capability of the machine can have an important role in the development of many applications such as neurofeedback therapies for

mood and cognition improvement [9]. Also, affective computing has substantial potential in helping victims with a neurodevelopmental disorder and improving their ability to emote and identify emotional expressions [10].

Emotion recognition can be achieved by analyzing the abovementioned nonphysiological expression such as the vocal signals, facial expression and bodily gestures, and the physiological signals such as the photoplethysmogram (PPG), electrocardiogram (ECG), electrodermal activity (EDA), electromyogram (EMG), and electroencephalogram (EEG).

Emotion recognition methods can be classified as multimodal or single modal. The combined information from two or more of the physiological and nonphysiological aspects is required in the multimodal emotion recognition methods, while the single-modal recognition approach typically utilizes one type of physiological signal. EEG is among the most widely used single-modal signal for emotion recognition as it carries the information of the brain neuronal activities from which almost all other physiological and nonphysiological reactions arise [11–14].

A recent extensive 5-year review by Craik et al. (2019) [15] on the current research state of deep neural networks for EEG classification reported the finding of only about 22% of the emotion recognition research using EEG signal values as the input data, with the remaining vast majority using the precalculated EEG features or images constructed from the EEG features as the input data to the classifiers.

Another 8-year review (2010–2018) by Roy et al. (2019) [16] on deep learning architectures for EEG analysis covering 154 publications reported that only about 12% of the deep learning architectures used for the affective mental domain were the convolutional neural networks (CNNs).

In recent years, there are an increasing number of publicly shared EEG databases among the research community. With this trend ongoing, increasingly sufficient variations of input EEG samples will be available soon for the training of very-deep neural networks. The projected future availability of sufficiently large variation of input EEG samples can promisingly solve the problem of overfitting of very-deep neural networks to the small research pool of EEG samples which are currently insufficient to serve as a good representative of the population. Therefore, our work studies the application of variants of very-deep CNN (*ResNet18* and *VGG16*) on the plain EEG signal (instead of precalculated EEG features) classification, with emotion recognition as the case study. We will present the architectural optimization of Residual Network (*ResNet*) for EEG signal classification. The performance of the variants of *ResNet* will be compared with the *VGG* variants. We will also compare the significance of different EEG channel subsets for emotion recognition and present the relevance of different subsets of EEG channels to emotion recognition.

2. Methodology

2.1. Experiment Design of SEED Dataset. This study uses the EEG signals recorded in the SEED experiment by the Shanghai Jiao Tong University (SJTU). The SEED dataset [17, 18] is an emotion-related EEG dataset publicly available for research purposes. The stimuli in the SEED experiment

were 15 film clips carefully chosen such that each elicits a single desired target emotion. Each film clip lasts about 4 minutes and is coherent to either positive, neutral, or negative valence emotion as described in Table 1.

SEED experiment had 15 participants. Every participant underwent 3 sessions of the experiment, with at least one-week interval between every 2 sessions. Each experiment session contained 15 trials, each playing one of the 15 film clips followed by self-assessment and a short rest. Figure 1 shows the structure of the experiment session.

The play sequence of the film clips was arranged such that no two consecutive trials carried the clips of the same emotion category.

The EEG signals were recorded with 62 active AgCl electrodes of the ESI NeuroScan System at a sampling frequency of 1000 Hz. The electrode placement was based on the international 10–20 system as shown in Figure 2.

The recorded EEG signals were then downsampled to 200 Hz and a bandpass frequency filter of 0.5 Hz to 70 Hz was applied.

2.2. SEED Dataset Literature Review. The research working on the SEED dataset in the recent three years (2018–2020) was reviewed and is summarized in Table 2. Although many of the research works were using one or another kind of neural network classifier, almost all of the attention had been placed on using manually extracted EEG features, instead of plain EEG signals.

Using plain EEG signals as the input data to the emotion classifiers has currently received relatively much lower research attention. Although the number of currently available public EEG research databases may not yet be sufficiently representative of the general population, the trend of an increasing number of publicly available EEG databases shall warrant more research works into the application of very-deep neural networks on plain EEG signals.

In line with this, the focus of this work is on eliciting the architectural modification on the original image-oriented *ResNet* and *VGG* that results in a vast improvement of their performance on plain EEG signal. In addition, we have also proposed the location of EEG channels that are most useful for emotion recognition.

2.3. EEG Dataset Preprocessing. As the target emotion caused by watching the film clip would not likely be successfully induced immediately at the start of the film clip, we have set a buffering period of 90 seconds for the emotion establishment. Therefore, the initial 90 seconds of each of the 4-minute EEG trials were discarded.

The remaining EEG recording is split into 2-second nonoverlapping segments, with each EEG segment assuming the length of 400 sampling points for the sampling frequency of 200Hz. Each of the nonoverlapping segments is then normalized along the time axis, respectively, using the Euclidean normalization method. All the generated EEG segments are split into five subpools for 5-fold cross-validation of the model performance.

TABLE 1: Film clips in SEED dataset.

Source film name	Emotion	Number of clips
Tangshan earthquake	Negative	2
Back to 1942	Negative	3
Lost in Thailand	Positive	2
Flirting scholar	Positive	1
Just another Pandora's box	Positive	2
World heritage in China	Neutral	5

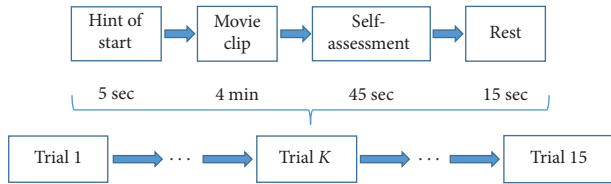


FIGURE 1: Data collection session design of SEED experiment.

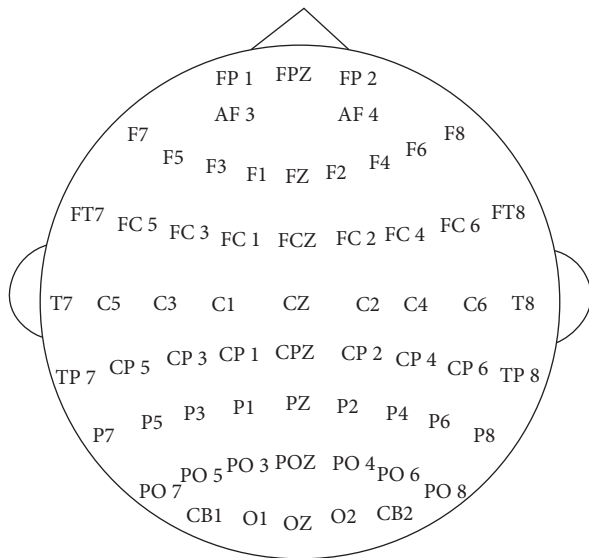


FIGURE 2: EEG channel layout of SEED dataset.

2.4. *Optimizing ResNet and VGG for EEG Signals.* Figures 3 and 4, respectively, illustrate the architectural details of different versions of *ResNet18* and *VGG16* examined in this study.

2.4.1. *ResNet Optimization.* The original architecture of *ResNet18* consisting of 17 convolutional layers and 1 layer of the fully connected network is depicted in Figure 3(a).

As the original *ResNet18* is designed for image processing, the convolutional kernels within the model are all 2-dimensional kernels. It has 3-by-3 kernels throughout its convolutional path, except for the very first convolutional layer (*Conv 0*) which has 7-by-7 kernels.

The color coding of Figure 3 denotes the major convolutional blocks of the ResNet. The convolutional layers of the same color have the same number of kernels (e.g., orange for 64 kernels, yellow for 128 kernels, green for 256 kernels, and blue for 512 kernels). The darker color layers are

convolutional layers, while the lighter layers are the other functional layers in the block, such as the batch normalization (BN) function, the Rectified Linear Unit (ReLU) activation function, the summation (Sum) of the by-passed feature map and the main convolution feature map, and the adaptive average pooling (AvgPool). The adaptive AvgPool layer before the fully connected (FC) layer allows the model to process EEG signals of different numbers of channels without the need to reassign the number of connections in the FC network.

The last layer of the *ResNet18* is a single layer of a fully connected (FC) network with three output nodes, corresponding to the three emotion classes.

There are two types of bypass connection in the ResNet, i.e., the identity bypass and the downsampling bypass. The identity bypass has its feature map being passed on, skipping two convolutional layers without any further processing before the summation function. The downsampling bypass happens at the initial stage of every major convolutional block, where the input feature maps will have their map size reduced due to kernel stride and the number of feature maps will increase due to the increment of convolutional kernels. Therefore, the downsampling bypass is necessary in order to have the dimension of the shortcut data matching the data dimension of the main convolutional path. While the identity bypass performs no additional processing on the data passed onwards, the downsampling bypass has 1-by-1 convolutional kernels which introduce an additional small number of trainable parameters as reported in Figure 3.

In this study, three variants of the original *ResNet18* were constructed and investigated. Two of the three *ResNet18* variants are illustrated in Figures 3(b) and 3(c). The 2D kernels of the *ResNet* were all restructured into 1D kernels along either the temporal(time)-dimension or the spatial(channel)-dimension.

The variant in Figure 3(b) has alternating temporal and spatial-dimension convolution. Eckart and Young [28] and Maji and Mullins [29] reported that the matrix such as the convolution filters can be well approximated with an arbitrary number of lower rank matrices. Maji and Mullins (2018) [29] had also demonstrated the feasibility of separating the 2D kernels of the well-established CNNs (e.g., *AlexNet*, *VGG-16*, *Inception-v1*, *ResNet-152*) into alternating 1D vertical and horizontal kernels, achieving near baseline accuracy for image classification with a significant speedup of training.

Nevertheless, given the different format and nature of EEG signals from the images, the alternating arrangement of 1D horizontal (time-dimension) kernel and 1D vertical (spatial-dimension) kernel may not be the optimal design for EEG signal processing. Therefore, we have constructed another variant of *ResNet18* (Figure 3(c)) with the initial two major convolutional blocks (all the nine initial convolutional layers) operating purely in the temporal dimension before introducing the spatial convolutional kernels. The spatial-dimension convolution of this *ResNet* variant appears only in the final two convolutional blocks.

In addition, we have investigated the effect of initializing the convolutional path with spatial-dimension

TABLE 2: Recent research on SEED dataset.

Classifier algorithm/year	Data input	Accuracy (%)
Dynamic graph CNN [19]/2018	Differential entropy (DE)	79.95
Logistic regression classifier [20]/2018	DE	72.47
GRSLR (graph regularized sparse linear regression) [21]/2018	DE, Hjorth features	88.41
Bidirectional LSTM [22]/2019	DE/Power spectral density (PSD)	94.96/86.27
Graph convolutional broad network (GCBN) [23]/2019	DE	94.24
CNN + LSTM [24]/2019	DE	89.88
Variational pathway reasoning (VPR) [25]/2019	DE	94.3
Sequential backward selection SVM [26]/2019	Hjorth features, standard deviation, sampling entropy, wavelet entropy	89
Spiking NN [27]/2020	DWT, FFT, variance	96.67

convolution, by making only a single change in the initial layer (*Conv 0*) of *ResNet18-1D-kernel-(T-S-alternate)* in Figure 3(b), from time-dimension convolution into spatial-dimension convolution. We have name-coded this variant as *ResNet18-1D-kernel-(S-T-alternate)*, for comparison with the model in Figure 3(b) to highlight the great impact of the abovementioned single minor architectural change on the model's performance which is presented in Figure 5.

The right columns of the Figures 3(a)–3(c) indicate the number of trainable parameters in each architectural layer of the *ResNet* variants.

2.4.2. VGG Optimization. As illustrated in Figure 4, variants of *VGG16* are also constructed for performance comparison with the variant of *ResNet18*. The *VGG* models have classical convolutional pathways without data bypassing. The *VGG16* has five major convolutional blocks, with two convolutional layers in each of its first two major convolutional blocks and three convolutional layers in each of its last three convolutional blocks. These thirteen convolutional layers together with the final three FC layers have made up the 16 main functional layers in the *VGG16*.

Figure 4(a) shows the structure of the *VGG16* with all the original 2D kernels being modified into 1D kernels along either the temporal or spatial dimension. The model in Figure 4(b) is named *VGG14-1D* with the removal of the two hidden FC layers from the *VGG16-1D*, such that the fully connected network is more closely resemble and comparable to that of the *ResNet18*.

The *VGG* architectures in Figure 4 are color-coded such that the transition between different color blocks is preceded by max-pooling (*MaxPool*) operation along the dimension of the previous convolution operation. The adaptive *AvgPool* layer before the FC networks is for the same purpose as described for the *ResNet18*.

We have also investigated the importance of batch normalization in CNN for EEG processing by removing the BN layers of the *VGG16* as in Figure 4(c). The performance analysis is presented in the Results section.

2.5. Model Training. The objective function for model optimization during training was set as the cross-entropy loss of the CNN outputs. Adam optimizer was used to update the trainable

parameters of the CNN at the learning rate of 0.001, based on the backpropagated error from the output cross-entropy loss.

The model training process was conducted with stochastic minibatches, with the size of each minibatch being one 200th of the total training pool. Thus, one complete training epoch consists of 200 training iterations. The training data pool will be reshuffled after every complete training epoch to ensure the different combinations of minibatch samples in the subsequent training epochs. Stochastic minibatch training serves to prevent the training process from being stuck at the local minima of the objective function.

3. Results and Discussion

3.1. Variants of ResNet18. Figure 5 presents the averaged 5-fold cross-validation classification accuracy of the *ResNet* variants, using different subsets of EEG channels as their data input.

The *ResNet* variant with 1D kernels has generally outperformed the original *ResNet18*, particularly in the scenario of using a lower number of EEG channels (10 channels for each subset). Not only has the classification improved with the *ResNet18* architectural restructuring from 2D-kernel convolution to 1D-kernel convolution, the total number of trainable parameters (obtainable by summing up the layer-wise parameters in Figure 3) in the *ResNet18* has also seen a reduction of more than 50% from the original 11.17 million parameters down to the range of 4.27 to 5.34 million parameters.

As pointed out in Section IV-B, the models *ResNet18-1D-(S-T-alternate)* and *ResNet18-1D-(T-S-alternate)* differ in only their very first convolutional layer (the *Conv-0* of Figure 3(b)), where the *ResNet18-1D-(T-S-alternate)* model has *Conv-0* as temporal convolution while the *ResNet18-1D-(S-T-alternate)* model has its *Conv 0* as spatial convolution. Although this single change in *Conv-0* has resulted in the difference in parameter count by only 256 ($(1 \times 9 - 5 \times 1) \times 64 = 256$), the performance in EEG signal classification has seen substantial improvement by about 10% elevation (using either all 62 channels, the outermost 10 channels, or outer 10 channels), as presented in Figure 5. This strongly indicates that the convolution operation on

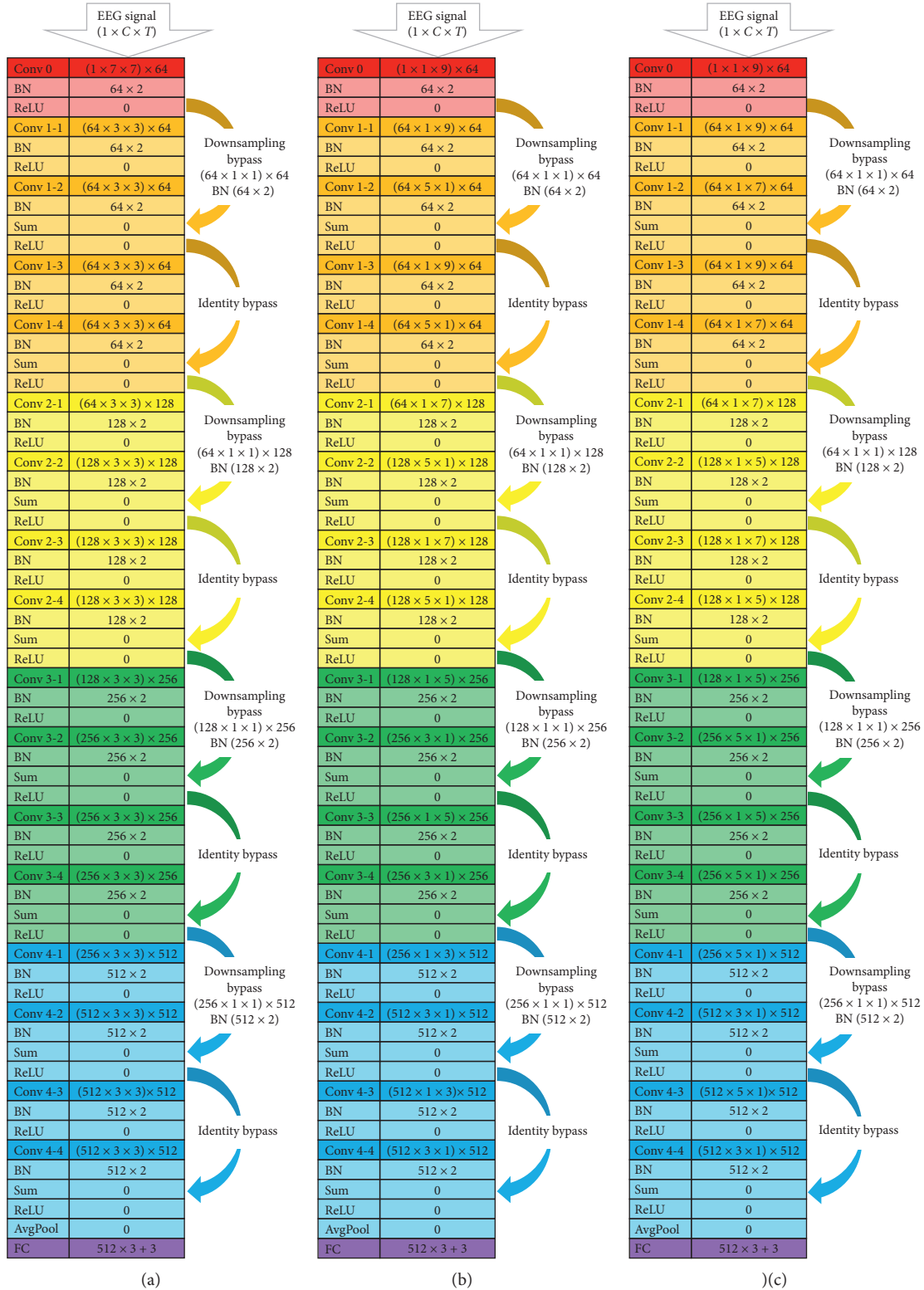


FIGURE 3: Architectural details of (a) original *ResNet18* and its modified variants (b) *ResNet18-1D-kernel-(T-S-alternate)* and (c) *ResNet18-1D-kernel-(T-then-S)* for EEG signal processing.

plain EEG signal should not be initiated with spatial(channel)-dimension convolution.

Some other previous works that used CNN for plain EEG signal processing had also forced the convolution process to

operate only along either the temporal or spatial dimension for every single convolutional layer. Most of the works [30–36] applying 1D-kernel CNN on EEG signals had initiated the convolutional path with temporal convolution.

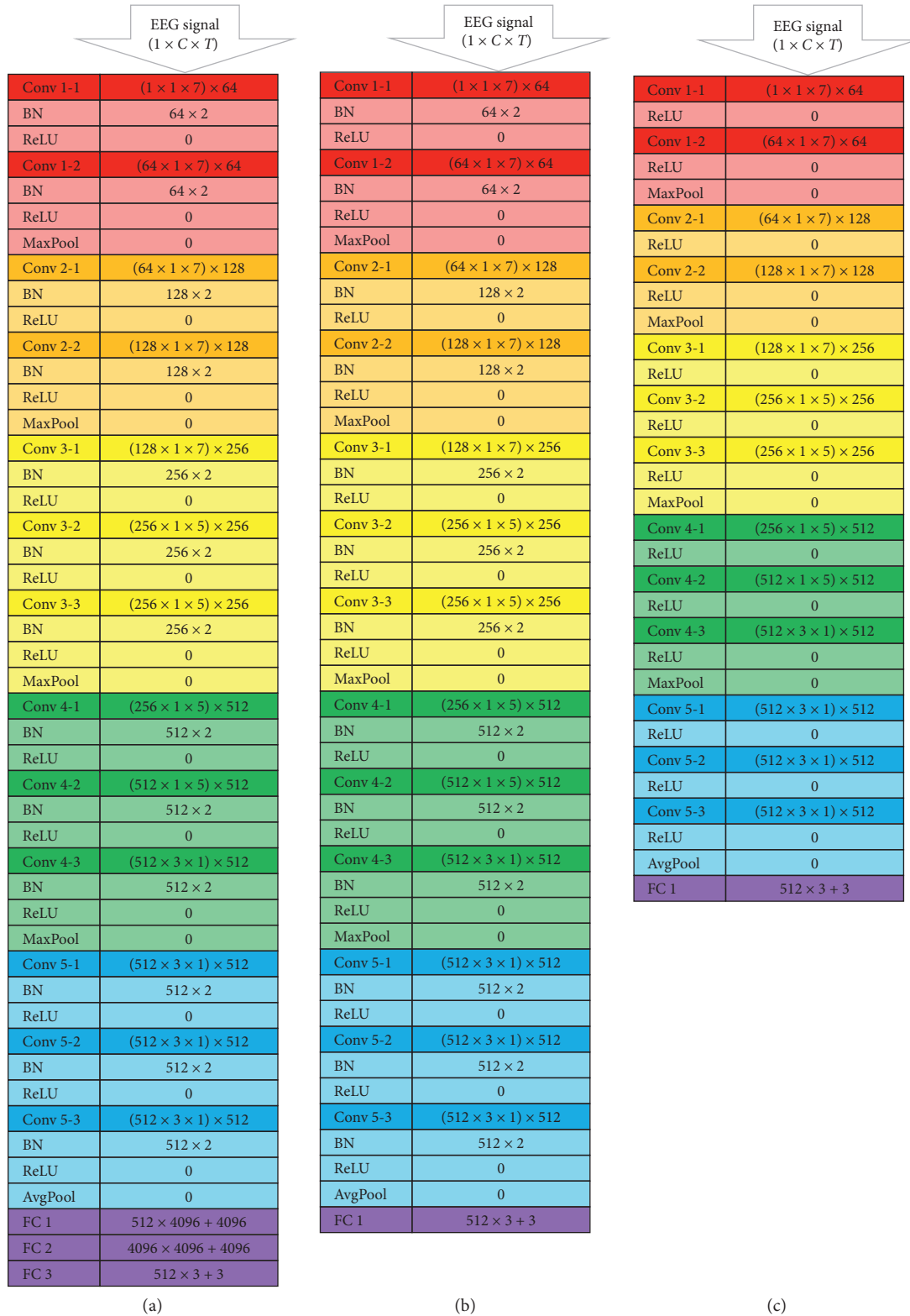


FIGURE 4: Architectural details of (a) VGG16-1D-kernel and its modified variants (b) VGG14-1D-kernel and (c) VGG14-1D-kernel (no batch norm) for EEG signal processing.

However, they had not provided the performance comparison with the models that did otherwise, as we highlighted in this study.

We took a further step of increasing the number of layers of pure temporal convolution before starting spatial convolutional operation, as in the architecture of *ResNet18-1D-(T-then-S)* in

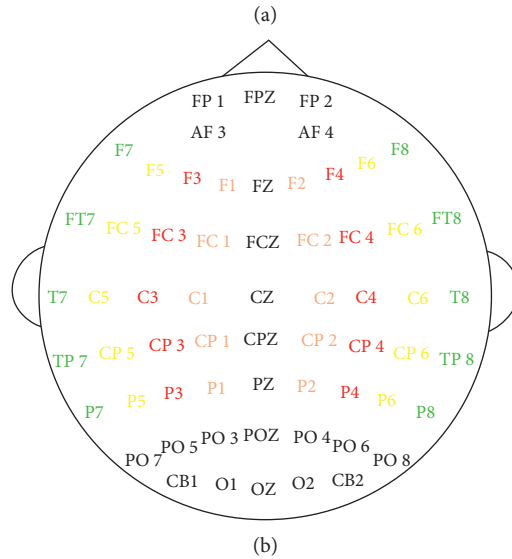
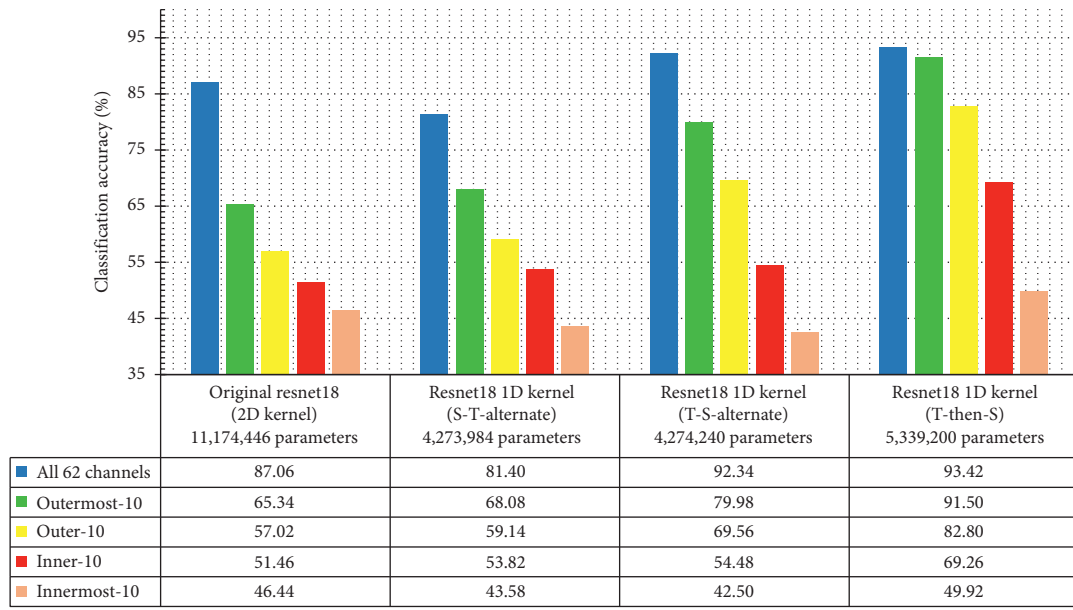


FIGURE 5: SEED 3-class emotion recognition accuracy by variants of *ResNet18* using different subsets of EEG channels. (a) Classification accuracy and the total number of model parameters. (b) Different subsets EEG channels.

Figure 3(c). The *ResNet18-1D-(T-then-S)* model has outperformed all the other *ResNet18* variants substantially, in every classification scenario as reported in Figure 5.

This supports that constructing multiple consecutive layers of temporal convolution before starting spatial convolution is beneficial for extracting distinctive information from the EEG signals. Although *ResNet* had been reported with inferior performance than the typical CNN at EEG classification in [34], their *ResNet* architecture was, however, designed with spatial convolution very early on as the second convolutional layer. If more temporal convolutional layers were introduced before the spatial convolution, the *ResNet* presented in [34] could potentially have seen significant performance improvement.

With the presence of multiple consecutive temporal convolutional layers before spatial convolution, higher hierarchical

features within each EEG channel could be extracted before comparing across different channels. Direct cross-channel convolution of rudimentary EEG voltages may not carry as much distinctive information as that of the higher hierarchical features.

Plain EEG signals carry only voltage levels measured over the scalp. Every single sampling point of the voltage level in an EEG channel is not as meaningful as a sequence of sampling points along the channel. The excessively short receptive field over a single channel is susceptible to recording artifacts and other nonessential signal variations.

Therefore, with multiple consecutive temporal convolutional layers, the initial stages of the model can cover a larger receptive field over the raw signal, at the same time extracting features of a higher level of abstraction from the

particular channel. Comparing the rudimentary EEG signal sampling point by sampling point across the channels may have considerably taken into account the undesired meaningless voltage variations, resulting in lower classification accuracy in the *ResNet18-1D-(S-T-alternate)* model.

We have also constructed and examined another variant of the *ResNet18-1D-(S-then-T)* model with its several initial convolutional layers all being spatial-dimension convolution followed by temporal convolution only. This model which was not presented in Figure 3 had presented worse performance than even the *ResNet18-1D-(S-T-alternate)* model, which further supports the proposal above that EEG signal convolution for emotion recognition should ideally be started with temporal-dimension convolution.

Figure 6 reports the training-validation performance log of the four variants of *ResNet-1D*, using the 10 outermost channels. Based on the training-validation cross-entropy loss plot, the *ResNet18-1D-(T-then-S)* model, which had outperformed all the rest, was clearly less susceptible to overfitting. The other three *ResNet18-1D* models all had started to experience overfitting after around eight to ten training epochs, with the models *ResNet18-1D-(S-then-T)* and *ResNet18-1D-(S-T-alternate)* experiencing the greatest degree of overfitting.

3.2. ResNet versus VGG. We have compared the performance of *ResNet18* with the more classical CNN architecture (the *VGG16*) from the aspects of classification accuracy, the number of trainable parameters, and the model training convergence speed.

Figure 7 shows that the classification accuracy achieved by *ResNet18-1D-(T-then-S)*, *VGG14-1D*, and *VGG16-1D* models is very close to each other. The *ResNet18-1D-(T-then-S)* achieves 93.42% classification accuracy, outperforming the VGG at using all 62 EEG channels. The VGG models have achieved higher accuracy at the less significant subsets of EEG channels (e.g., using the innermost 10 channels).

Given the almost negligible difference in the classification accuracy, the *ResNet18-1D-(T-then-S)* model contains only 5.34 million parameters, which is only about 36.3% of that in the *VGG14-1D* model which has 14.72 million parameters. The *VGG16-1D* has an even staggering greater number of parameters (at 46.18 million) due to a large number of fully connected perceptrons in its original 3-layer FC networks. This densely connected FC network containing over 31 million parameters does not appear to be essential to the classification accuracy.

Another aspect of performance measurement investigated is the convergence speed of the model under training. With reference to Table 3, using all 62 EEG channels, the *ResNet18-1D-(T-then-S)* and the *VGG14-1D* models are able to converge to above 95% training accuracy in 11 epochs and 10 epochs, respectively. The *VGG16-1D* requires a greater number of training epochs (14 complete rounds) to reach its training accuracy of 95%. The lower convergence speed of *VGG16-1D* is likely due to its complex FC network.

The *ResNet18-1D-(T-then-S)* model completes a training epoch with (1665/11 \approx 151) seconds, while the VGG models require a much greater amount of time to complete a training epoch (*VGG14-1D* using about 249 seconds, and *VGG16-1D* using about 250 seconds).

Similarly, the *ResNet18-1D-(T-then-S)* uses only about 38 seconds for a complete training epoch with 10 EEG channels, while the two VGG models use about 50 seconds for completing a training epoch.

The *VGG14-1D-(no batch norm)* illustrated in Figure 4(c) is the version of *VGG14-1D* without the batch normalization function after every convolutional layer. This model without the batch normalization had failed to progress well even in its training phase. The training accuracy of this model had stayed at around 35%, with the training loss staying at around the initial value.

The failure of this *VGG14-1D-(no batch norm)* has indicated the importance of batch normalization in training deep CNN on EEG signals, even with the EEG signals being prenormalized before being passed into the CNN model. All the *ResNet18* variants in Figure 3 are also equipped with batch normalization at the output of their convolutional layers.

In our model, each layer of the batch normalization function introduces two additional trainable parameters per feature map. The dimension of the feature map depends on the number of convolutional kernels immediately preceding the batch norm function.

The short EEG segments being passed into the classifier may contain large signal amplitude variations from segment to segment. Different batches of the EEG segments may also encounter the problem of large internal covariate shift [37] which is a notorious reason for the diverging loss during model optimization [38].

This does not only slow down the training speed by demanding a very low learning rate but also potentially disrupt altogether the convergence of the model optimization process as experienced in our model (Figure 4(c)) without batch normalization.

3.3. Channel Significance in Emotion Recognition. Identifying the most critical subsets of EEG channels can reduce the input data redundancy and ease the design and mounting of portable consumer-friendly EEG recording hardware. Therefore, previous works [39–41] had tried to identify the subsets of EEG channels that are most crucial for emotion recognition. In line with the purpose, we have looked into the emotion EEG channel significance with regard to lateral-medial placement, along the nasion-inion axis, and in terms of the left-versus-right hemispheric discrepancy.

3.3.1. Electrode Distance to the Midline. With reference to Figures 5(a) and 5(b), the relevance of different subsets of EEG channels for emotion recognition is investigated, with respect to the channels' distance from the midline.

The trend of classification accuracy as reported in Figure 5 follows that the more laterally placed the EEG channels

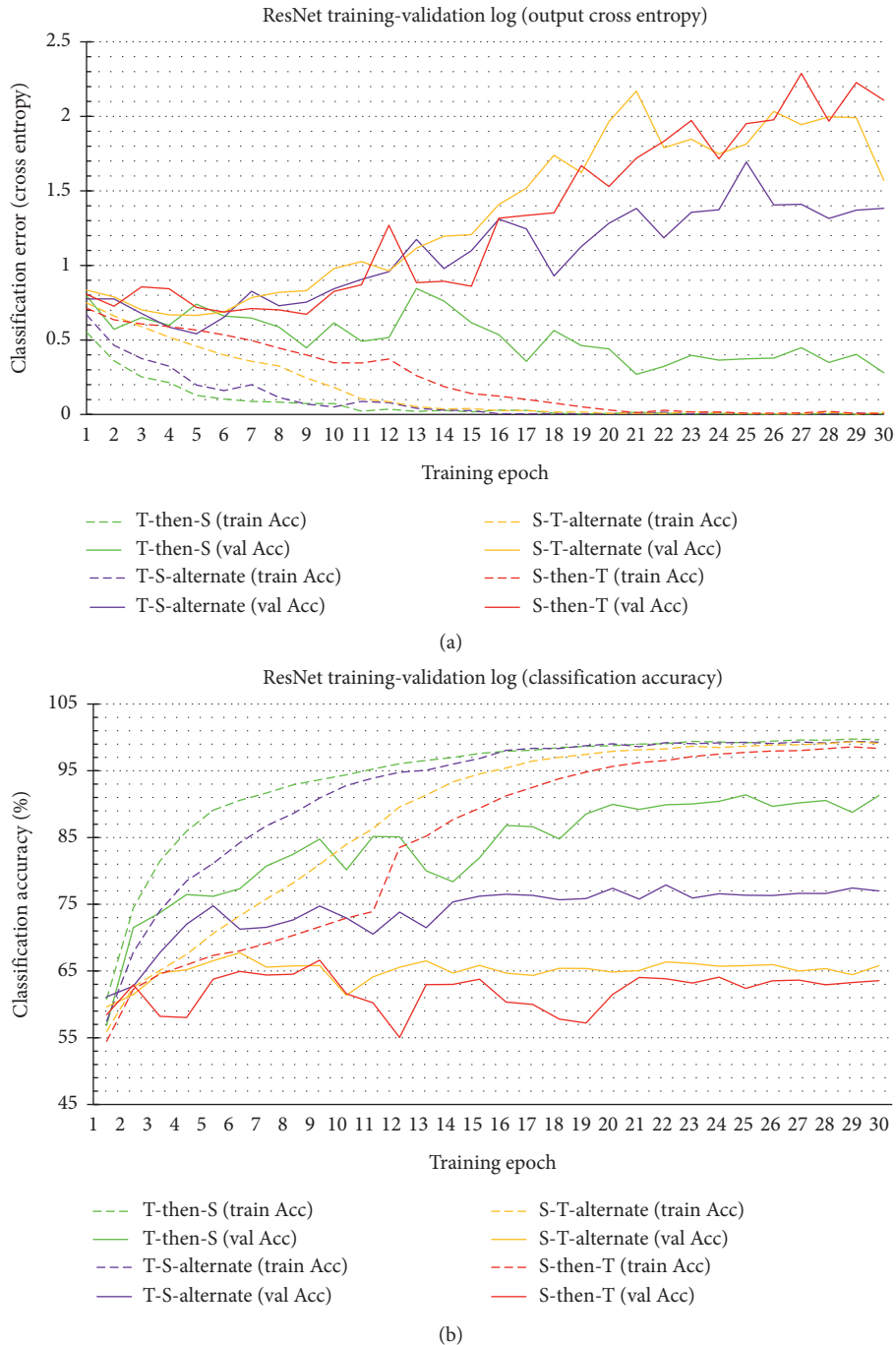


FIGURE 6: Training-validation performance log of variants of ResNet18-1D.

are, the higher the classification accuracy they deliver. This implies that more emotionally distinctive information is carried in the laterally placed (farther away from the midline) EEG channels than the medially placed channels.

The possible reason for this channel significance distribution pattern is that the lateral channels are in fact placed over or close to the temporal region above the ears on the scalp. These electrode locations are closer to the brain structures that are highly involved in emotional response. These structures (such as the anterior temporal pole, the

insular cortex, the amygdala, and the hippocampus [42–44]) are either part of the temporal lobe itself or lying at just the medial side of the temporal lobe. Hence, the more medially placed EEG electrodes are located higher up on top of the scalp and are hence farther away from these emotionally important brain structures.

3.3.2. *Along the Nasion-Inion Axis.* Figure 8 shows the 5-fold cross-validated emotion classification accuracy of

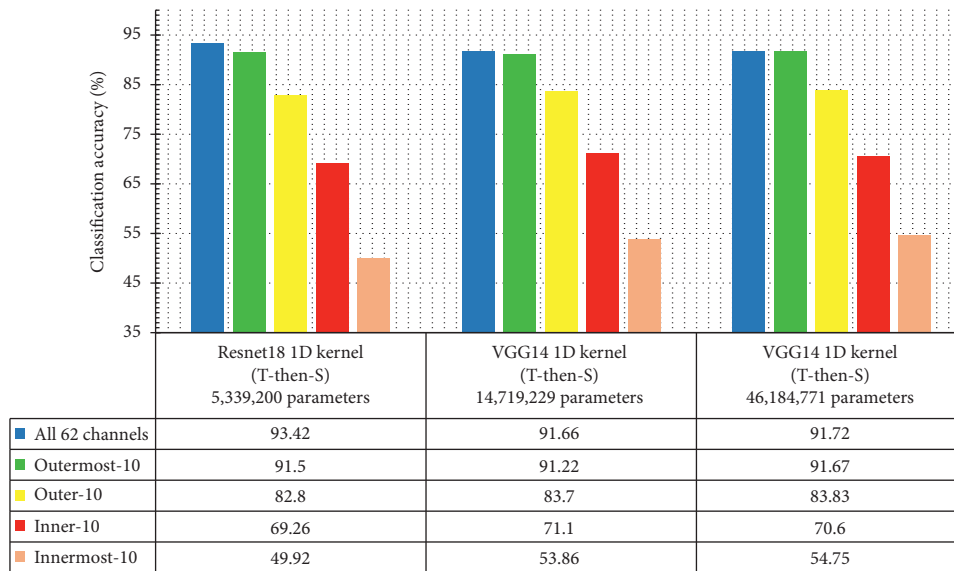


FIGURE 7: Classification accuracy of ResNet18-1D and VGG16 variants.

TABLE 3: Model training convergence efficiency comparison between *ResNet* and *VGG*.

	Training length to reach 95% training accuracy (epochs//seconds)	
	Using all 62 channels	Using outermost 10 channels
ResNet18-1D (T-then-s)	11//1665	11//416
VGG14-1D (T-then-s)	10//2488	10//503
VGG16-1D (T-then-s)	14//3505	12//622

ResNet18-1D(T-then-S) model, using four different subsets of EEG channels along the nasion-inion axis.

As indicated by Figure 8(b), these subsets of EEG channels, respectively, cover the frontal region (blue), centrotemporal region (green), centroparietal region (yellow), and the parietooccipital region (red).

In coherence with the distribution of emotionally important brain structures (e.g., the anterior temporal pole, the insular cortex, and the amygdala) discussed above, the three different emotion classes are best classified with the twelve centrotemporal channels (green color coded) because these twelve channels are located nearest to these structures, relative to the other three subsets.

The twelve-frontal-channel subset gives the same accuracy as the twelve parietal channels. The occipital channels are the least emotionally correlated set of EEG channels.

3.3.3. Cerebral Lateralization of Emotion. Figure 9 shows the 5-fold cross-validation accuracy using EEG channels of the left hemisphere versus the right hemisphere. The left channels present around 4-5% higher accuracy than the right channels. Using only 10 lateral channels of the left hemisphere has resulted in 88.48% average accuracy which is still even better than using all 27 right-hemispheric channels which give 86.96%.

This lateralized significance of EEG channels in emotion recognition can be due to the fundamental cerebral

lateralization [45, 46] or simply because of the nature of the SEED experiment design.

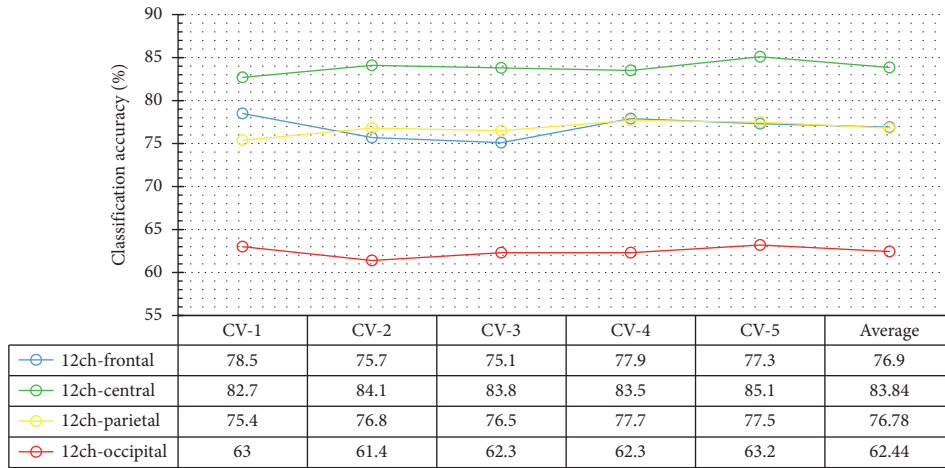
The stimuli of the SEED experiment were movie clips, and the mode of content delivery of movies can be heavily verbal or language-based. The center of language processing and understanding is located exactly in the lateral side of the left temporal lobe, known as Wernicke's area [47]. Therefore, the imbalanced activation of Wernicke's area in comparison to its right-hemispheric counterpart area can be a compounding factor resulting in the classification accuracy discrepancy.

3.3.4. Comparing across all the Channel Subsets.

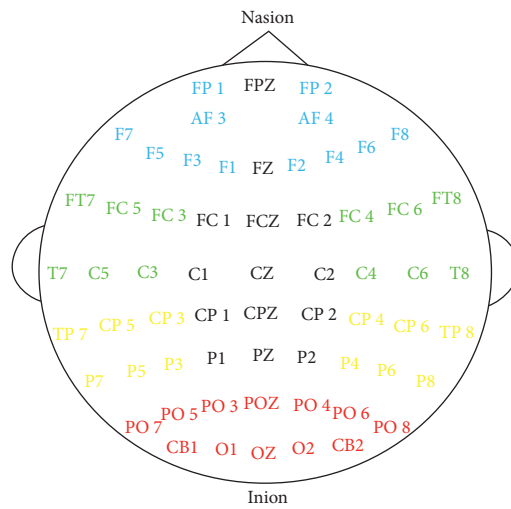
Reviewing the classification results using various EEG channel subsets presented in Figures 5, 8, and 9, the ten lateral-most left and right EEG channels in Figure 5 achieved the highest accuracy (91.5%), compared to using the ten lateral left channels in Figure 9 which has achieved 88.48% recognition accuracy and the twelve centrotemporal channels in Figure 8 which have achieved 83.84% accuracy.

With a comparable number of channels used in the subsets, the above result implies that there is additional distinctive information for emotion recognition retrievable from the left-versus-right channel feature cross-correlation, in view of the pairing of 10 left and right channels giving better classification result than the 10 lateral-most left channels.

The highly emotion-correlated subsets of EEG channels identified by this work are close to the 12-channel (FT7, FT8,

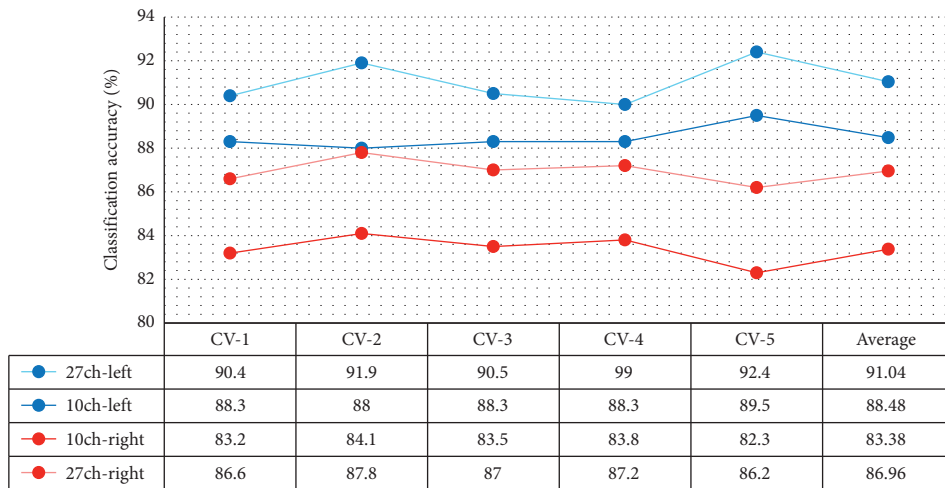


(a)



(b)

FIGURE 8: SEED 3-class emotion recognition accuracy using different subsets of EEG channels along the nasion-inion axis. (a) 5-fold cross-validation classification accuracy. (b) Electrode placement.



(a)

FIGURE 9: Continued.

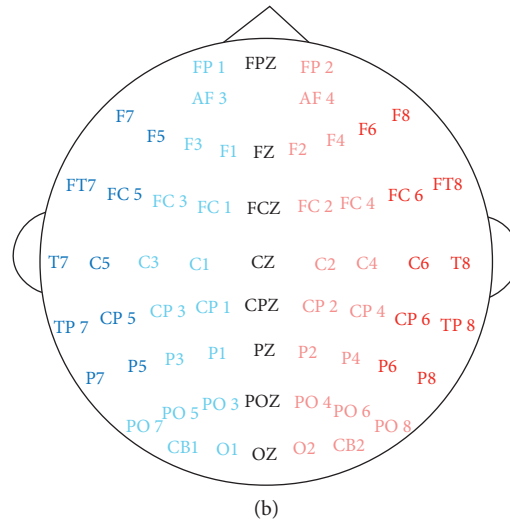


FIGURE 9: SEED 3-class emotion recognition accuracy comparison using left and right hemispheric EEG channels. (a) 5-fold cross-validation classification accuracy. (b) Electrode placement.

T7, T8, C5, C6, TP7, TP8, CP5, CP6, P7, and P8) subsets used by Zheng and Lu (2015) [41] which was reported to have achieved even higher emotion recognition accuracy than using all 62 channels.

4. Conclusion

This study has investigated the applicability of a very-deep convolutional neural network for plain emotion-related EEG signal classification, which is an area of relatively low research attention as most of the emotion EEG classification tasks were based on preextracted EEG features. With the future availability of a greater pool of EEG data that better represents the population, the very-deep CNNs can potentially outperform the feature-based algorithms, although they do not yet show accuracy superiority over feature-based algorithms with the current size of publicly available EEG database.

ResNet18 and *VGG16* originally constructed for image processing were modified for EEG signal processing. The original *ResNet18* and *VGG16* designed for image processing are not ideal for direct application onto EEG signals. Our modified variants of *ResNet18* with 1D kernels have shown significant performance improvement in both the aspects of classification accuracy and reduced model parameters. The modified *ResNet18* variants have shown better training convergence speed than the *VGG16* variants.

The sequence of convolutional dimension arrangement within the *ResNet18-1D* has also been investigated for optimal EEG signal processing performance. The result findings have suggested against initiating the convolutional operation with spatial-dimension convolution. Multiple layers of consecutive temporal-dimension convolution should ideally be placed before the operation of spatial-dimension convolution. Using the SEED dataset, our best performing model [*ResNet18-1D-(T-then-S)*] has achieved a 3-class emotion classification accuracy of 93.42%.

Not of less importance, the batch normalization function proves to be essential in tackling the problem of internal

covariate shift which can result in model optimization convergence failure.

Investigating the EEG channel significance for emotion recognition from the neurological aspects, the laterally placed channels around the temporal lobe show greater importance than the channels placed over other brain regions. This finding is consistent with the fact that many emotionally important brain structures are located within or nearby the temporal lobe.

Data Availability

The data are available at <http://bcmi.sjtu.edu.cn/home/seed/seed.html>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is financially supported by the Universiti Tunku Abdul Rahman Research Fund (UTARRF) (Grant No. IPSR/RMC/UTARRF/2020-C2/H01) and Excellent Research Center Award Fund, Centre for Healthcare Science and Technology, UTAR, Malaysia.

References

- [1] K. R. Scherer, "What are emotions? And how can they be measured?" *Social Science Information*, vol. 44, no. 4, pp. 695–729, 2016.
- [2] R. Sinha, W. R. Lovallo, and O. A. Parsons, "Cardiovascular differentiation of emotions," *Psychosomatic Medicine*, vol. 54, no. 4, pp. 422–435, 1992.
- [3] G. Valenza, L. Citi, A. Lanata, E. P. Scilingo, and R. Barbieri, "Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics," *Scientific Reports*, vol. 4, no. 4998, 2014.

- [4] I. Homma and Y. Masaoka, "Breathing rhythms and emotions," *Experimental Physiology*, vol. 93, pp. 1011–1021, 2008.
- [5] N. Ganapathy and R. Swaminathan, "Emotion recognition using electrodermal activity signals and multiscale deep convolution neural network," *Studies in Health Technology and Informatics*, vol. 258, p. 140, 2019.
- [6] P. Olejniczak, "Neurophysiologic basis of EEG," *Journal of Clinical Neurophysiology*, vol. 23, no. 3, pp. 186–189, 2006.
- [7] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis et al., "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32–80, 2001.
- [8] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, MA, USA, 2000.
- [9] O. Alkoby, A. Abu-Rmileh, O. Shriki, and D. Todder, "Can we predict who will respond to neurofeedback? A review of the inefficacy problem and existing predictors for successful EEG neurofeedback learning," *Neuroscience*, vol. 378, pp. 155–164, 2018.
- [10] H. Kalantarian, K. Jedoui, P. Washington et al., "Labeling images with facial emotion and the potential for pediatric healthcare," *Artificial Intelligence in Medicine*, vol. 98, pp. 77–86, 2019.
- [11] C. Hondrou and G. Caridakis, "Affective, natural interaction Using EEG: sensors, application and future directions," *Artificial Intelligence: Theories and Applications. SETN*, pp. 331–338, 2012.
- [12] M. Ali, A. H. Mosa, F. Al Machot, and K. Kyamakya, "EEG-based emotion recognition approach for e-healthcare applications," in *Proceedings of the 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 946–950, Vienna, Austria, July 2016.
- [13] O. A. Zoubi, M. Awad, and N. K. Kasabov, "Anytime multipurpose emotion recognition from EEG data using a Liquid State Machine based framework," *Artificial Intelligence in Medicine*, vol. 86, pp. 1–8, 2018.
- [14] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 910–926, 2020.
- [15] A. Craik, Y. T. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," *Journal of Neural Engineering*, vol. 16, no. 3, Article ID 031001, 2019.
- [16] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of Neural Engineering*, vol. 16, no. 5, Article ID 051001.
- [17] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential Entropy Feature for EEG-Based Emotion Classification," in *Proceedings of the 6th International IEEE EMBS Conference on Neural Engineering (NER)*, pp. 81–84, San Diego, CA, USA, November 2013.
- [18] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development (IEEE TAMd)*, vol. 7, no. 3, pp. 162–175, 2015.
- [19] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2018.
- [20] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 1, pp. 85–94, 2018.
- [21] Y. Li, W. Zheng, Z. Cui, Y. Zong, and S. Ge, "EEG emotion recognition based on graph regularized sparse linear regression," *Neural Processing Letters*, vol. 49, pp. 555–571, 2019.
- [22] Y. Wang, S. Qiu, J. Li et al., "EEG-based emotion recognition with similarity learning network," in *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, July 2019.
- [23] T. Zhang, X. Wang, X. Xu, C. L. P. Chen, and "GCB-Net, "Graph convolutional broad network and its application in emotion recognition," *IEEE Transactions on Affective Computing*, 2019.
- [24] S. Hwang, K. Hong, G. Son, and H. Byun, "Learning CNN features from DE features for EEG-based emotion recognition," *Pattern Analysis and Applications*, vol. 23, pp. 1323–1335, 2019.
- [25] T. Zhang, Z. Cui, C. Xu, W. Zheng, and J. Yang, "Variational pathway reasoning for EEG emotion recognition," in *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY, USA, February 2020.
- [26] F. Yang, X. Zhao, W. Jiang, P. Gao, and G. Liu, "Multi-method fusion of cross-subject emotion recognition based on high-dimensional EEG features," *Frontiers in Computational Neuroscience*, vol. 13, p. 53, 2019.
- [27] Y. Luo, Q. Fu, J. Xie et al., "EEG-based emotion classification using spiking neural networks," *IEEE Access*, vol. 8, pp. 46007–46016, 2020.
- [28] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.
- [29] P. Maji and R. Mullins, "On the reduction of computational complexity of deep convolutional neural networks," *Entropy*, vol. 20, no. 4, p. 305, 2018.
- [30] J. Behncke, R. T. Schirrmeister, W. Burgard, and T. Ball, "The signature of robot action success in EEG signals of a human observer: decoding and visualization using deep convolutional neural networks," in *Proceedings of the 6th International Conference on Brain-Computer Interface (BCI)*, Gangwon, South Korea, January 2018.
- [31] S. Chambon, M. N. Galtier, P. J. Arnal, G. Wainrib, and A. Gramfort, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 758–769, 2017.
- [32] N. S. Kwak, K. R. Müller, and S. W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS One*, vol. 12, no. 2, Article ID e0172578, 2017.
- [33] R. Manor and A. B. Geva, "Convolutional neural network for multi-category rapid serial visual presentation BCI," *Frontiers in Computational Neuroscience*, vol. 9, 2015.
- [34] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederera et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [35] R. Zafar, S. C. Dass, and A. S. Malik, "Electroencephalogram-based decoding cognitive states using convolutional neural network and likelihood ratio based score fusion," *PLOS ONE*, vol. 12, no. 5, Article ID e0178410, 2017.
- [36] K. H. Cheah, H. Nisar, V. V. Yap, and C.-Y. Lee, "Convolutional neural networks for classification of music-listening EEG:

- comparing 1D convolutional kernels with 2D kernels and cerebral laterality of musical influence,” *Neural Computing and Applications*, 2019.
- [37] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” 2015, <https://arxiv.org/abs/1502.03167>.
- [38] J. Bjorck, C. Gomes, B. Selman, and K. Q. Weinberger, “Understanding batch normalization,” 2018, <https://arxiv.org/abs/1806.02375>.
- [39] K. Ansari-Asl, G. Chanel, and T. Pun, “A channel selection method for EEG classification in emotion assessment based on synchronization likelihood,” in *Proceedings of the 15th European Signal Processing Conference (EUSIPCO)*, Poznan, Poland, September 2007.
- [40] M. S. Ozerdem and H. Polat, “Emotion recognition based on EEG features in movie clips with channel selection,” *Brain Informatics*, vol. 4, no. 4, pp. 241–252, 2017.
- [41] W.-L. Zheng and B.-L. 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, 2015.
- [42] R. Dolan, R. Lane, P. Chua, and P. Fletcher, “Dissociable temporal lobe activations during emotional episodic memory retrieval,” *NeuroImage*, vol. 11, no. 3, pp. 203–209, 2000.
- [43] F. Dolcos, K. S. LaBar, and R. Cabeza, “Remembering one year later: role of the amygdala and the medial temporal lobe memory system in retrieving emotional memories,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 102, no. 7, pp. 2626–2631, 2005.
- [44] T. Iidaka, T. Okada, T. Murata et al., “Age-related differences in the medial temporal lobe responses to emotional faces as revealed by fMRI,” *Hippocampus*, vol. 12, no. 3, pp. 352–362, 2002.
- [45] M. C. Corballis, “Left brain, right brain: facts and fantasies,” *PLoS Biology*, vol. 12, no. 1, Article ID e1001767, 2014.
- [46] H. Liu, S. M. Stufflebeam, J. Sepulcre, T. Hedden, and R. L. Buckner, “Evidence from intrinsic activity that asymmetry of the human brain is controlled by multiple factors,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 106, no. 48, pp. 20499–20503, 2009.
- [47] I. DeWitt and J. P. Rauschecker, “Wernicke’s area revisited: parallel streams and word processing,” *Brain and Language*, vol. 127, no. 2, pp. 181–191, 2013.