

## Research Article

# A Fused Multidimensional EEG Classification Method Based on an Extreme Tree Feature Selection

Ruijing Lin <sup>1,2</sup> Chaoyi Dong <sup>1,2</sup> Pengfei Ma <sup>1,2</sup> Shuang Ma <sup>1,2</sup> Xiaoyan Chen <sup>1,2</sup>  
and Huanzi Liu <sup>1,2</sup>

<sup>1</sup>College of Electric Power, Inner Mongolia University of Technology, Hohhot 0100801, China

<sup>2</sup>Intelligent Energy Technology and Equipment Engineering Research Centre of Colleges and Universities in Inner Mongolia Autonomous Region, Inner Mongolia, Hohhot 010051, China

Correspondence should be addressed to Chaoyi Dong; [dongchaoyi@imut.edu.cn](mailto:dongchaoyi@imut.edu.cn)

Received 2 April 2022; Accepted 19 July 2022; Published 8 August 2022

Academic Editor: António Dourado

Copyright © 2022 Ruijing Lin 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.

When a brain-computer interface (BCI) is designed, high classification accuracy is difficult to obtain for motor imagery (MI) electroencephalogram (EEG) signals in view of their relatively low signal-to-noise ratio. In this paper, a fused multidimensional classification method based on extreme tree feature selection (FMCM-ETFS) is proposed for discerning motor imagery EEG tasks. First, the EEG signal was filtered by a Butterworth filter for preprocessing. Second, C3, C4, and CZ channels were selected to extract time-frequency domain and spatial domain features using autoregressive (AR), common spatial pattern (CSP), and discrete wavelet transform (DWT). The extracted features were fused for a further feature elimination. Then, the features were selected using three feature selection methods: recursive feature elimination (RFE), principal component analysis method (PCA), and extreme trees (ET). The selected feature vectors were classified using support vector machines (SVM). Finally, a total of twelve subjects' EEG data from Inner Mongolia University of Technology (IMUT data), the 2nd BCI competition in 2003, and the 4th BCI competition in 2008 were employed to show the effectiveness of this proposed FMCM-ETFS method. The results show that the classification accuracy using the multidimensional fused feature extraction (AR + CSP + DWT) is 3%–20% higher than those using the aforementioned three single feature extractions (AR, CSP, and DWT). Extreme trees (ET), which is a sort of tree-based model method, outperforms RFE and PCA by 1%–9% in term of classification accuracies, when these three methods were applied to the procedure of feature extraction, respectively.

## 1. Introduction

As a new method of human-computer interaction, brain-computer interface (BCI) no longer relies on the output pathways of conventional peripheral cerebral nerve and muscle systems but directly uses peripheral auxiliary devices such as sensors, amplifiers, and computers to collect and analyze electroencephalogram (EEG) signals [1, 2]. Thus, the action intentions contained in brains can be extracted and deciphered for the purpose of information interaction between brains and the outside world [3]. Motor imagery (MI), as a typical BCI, refers to the act of imagining a specific action but not actually performing that action, and has

received widespread attention in fields such as neuroscience and artificial intelligence [4, 5].

The EEG signals of different MI tasks are usually accompanied by different sensory-motor rhythm (SMR), i.e., increased and decreased power in specific frequency bands in various brain regions, and mathematical feature vectors can be extracted from the SMR using a feature extraction algorithm and fed into a classifier for classification [6, 7]. The classical feature extraction method on the time domain is the autoregressive (AR). Wang and Chen used a hierarchical vector AR [8], and Liu et al. proposed a feature extraction method based on the combination of phase synchronization and AR model coefficients [9]. These two

feature extraction methods mainly focused on the feature extractions in the time domain and obtained a satisfactory classification accuracy for the investigated MI tasks [8, 9]. The typical feature extraction method in the time-frequency domain is the wavelet transform (WT). WT can characterize EEG signals in both time and frequency domains and decompose these EEG signals into multiple frequency bands to reveal different time-frequency features. Nevertheless, WT relies heavily on a priori knowledge to extract the frequency bands of interest [10]. This fact limits the application scope of WT. In another aspect, the common spatial pattern (CSP) extracts the features reflecting the different power distributions in the spatial domain for different MI tasks. Later, researchers combined CSP with other optimization algorithms or classification methods to improve its performance in classification accuracy. For example, Feng et al. proposed a new correlation-based time window selection (CTWS) algorithm [11]. The algorithm used correlation analysis to select an optimal reference signal and the starting point of time windows for each class, so that the average classification accuracy of MI-based BCI was substantially improved in combination with the traditional CSP. Bao Liu et al. proposed a PSO-CSP-SVM that was applied to MI feature extraction [12]. Gu et al. combined CSP with a convolutional neural network (CNN) to extract high-level characteristics of original data [13]. Pei et al. considered that the CSP algorithm usually cannot extract adequate frequency band features [14]. To overcome this shortcoming of the CSP, the paper proposed a tensor-based frequency feature combination (TFFC) method to construct a new feature set by fusing broadband features with narrowband features. Thus, the dependence on a classifier can be reduced and the adaptiveness of the features can be increased.

Each of the aforementioned methods has its own advantages, but they also have limitations when they are applied to further improve the classification accuracy of MI tasks in fact. These methods only consider some partial aspects of features underlying MI EEG signals and cannot fully reflect data characteristics and network information flows. This paper proposes a fused multidimensional classification method based on extreme tree feature selection (FMCM-ETFS) in classifying motor imagery (MI) EEG signals. The specific process is as follows: First, the EEG signals are filtered by a Butterworth filter to obtain a purified EEG signal. Second, AR, CSP, and DWT are used to extract the AR model coefficients, the variance and mean after CSP filtering, and the wavelet coefficients. These three methods can effectively extract time, frequency, and spatial domain features of EEG signals together to meet the signal dimensional diversity and yield more comprehensive information. Therefore, the three sorts of features were fused to obtain an augmented feature vector including all features. Then, the complexity of the classifier models is further decreased by three feature selection methods according to the support vector machine (SVM) classifier. The three methods are recursive feature elimination (RFE), principal component analysis method (PCA), and extreme tree (ET). Finally, a total of twelve subjects' EEG data from Inner

Mongolia University of Technology (IMUT data), the 2nd BCI competition in 2003 [15], and the 4th BCI competition in 2008 [16] are applied to validate the effectiveness of the proposed FMCM-ETFS. The experimental results show that the FMCM combined with ET obtained 1%–9% higher accuracy than the FMCM combined with RFE or PCA on average.

## 2. Feature Extraction Methods

*2.1. AR Model.* The AR model is a fundamental technique in time-series analysis and is widely used in BCI data processing [17]. The definition of the AR model is as follows: if there is a time series  $\{y(t); t = 1, 2, \dots, n\}$  consisting of  $n$  sample points at an equal time period, the time series can be regressed (or predicted) on its values at multiple previous moments, using the following formula:

$$y(t) = - \sum_{i=1}^p a_i y(t-i) + e(t), \quad (1)$$

where  $e(t)$  is a white noise series with mean 0,  $p$  is the order of the AR model, and  $a_i$  are the coefficients of the AR model [18].

*2.2. CSP.* The CSP algorithm is a spatial domain filtering and feature extraction algorithm for two categories tasks, capable of extracting the spatially distributed characteristics of each class from multi-channel EEG signals. The basic principle of CSP algorithm is to find a set of optimal spatial filters for yielding a projection. This projection uses matrix diagonalization to maximize the difference between the variances of the two categorical signals, thus obtaining feature vectors with a high degree of discrimination [19].

The experimentally measured EEG data is represented as an  $M \times N$  matrix  $X$ , where  $M$  is the number of channels and  $N$  is the number of sampling points per channel [20]. Using  $R_1$  and  $R_2$  to denote the normalized covariance matrices of the left-handed motion imagery and the right-handed motion imagery. Then, a normalized covariance matrix of EEG data is as follows:

$$R_1 = \frac{X_1 X_1^T}{\text{trace} X_1 X_1^T}, \quad (2)$$

$$R_2 = \frac{X_2 X_2^T}{\text{trace} X_2 X_2^T}.$$

Respectively, the mixed-space covariance matrix can be represented as follows:

$$R_m = \overline{R_1} + \overline{R_2}. \quad (3)$$

An eigenvalue decomposition is carried out according to the mixed space covariance matrix by the following equation(4):

$$R_m = U \lambda U^T, \quad (4)$$

where matrix  $\lambda$  is the diagonal matrix with the eigenvalues of  $R_m$  and  $U$  is the corresponding eigenvector matrix of  $\lambda$ . Thus, the whitened matrix  $P$  of  $R_m$  is derived as follows:

$$P = \sqrt{\lambda^{-1}}U^T. \quad (5)$$

Then, two transformations of  $R_1$  and  $R_2$  is performed as follows :

$$\begin{aligned} S_1 &= PR_1P^T, \\ S_2 &= PR_2P^T. \end{aligned} \quad (6)$$

After principal component decompositions of  $S_1$  and  $S_2$ , it can be proved that the eigenvector matrices of  $S_1$  and  $S_2$  are equal, and the sum of  $\lambda_1$  and  $\lambda_2$  is the identity matrix. Thus, the spatial filter  $W$  is constructed as follows:

$$W = B^T P, \quad (7)$$

where  $B$  is the eigenvector matrix of  $S_1$  and  $S_2$ .

**2.3. DWT.** WT is a transform analysis method, inheriting and developing the idea of localization in a short-time Fourier transform. As an ideal tool for the time-frequency analysis of signals, WT overcomes the shortcoming that the time-frequency window is fixed and cannot be adjusted with frequency [21]. The continuous wavelet transform (CWT) is defined as follows:

$$W_x(a, b) = |a|^{-1/2} \int x(t) \psi\left(\frac{t-b}{a}\right) dt = \langle x(t), \psi_{a,b}(t) \rangle, \quad (8)$$

where  $\psi_{a,b}(t) = |a|^{-1/2} \psi(t-b/a)$  is a wavelet function. The parameter  $a$  is the scaling factor of the wavelet function, and  $b$  is the translation parameter of the wavelet function. The two parameters adjust the frequency scale and the time scale, respectively. The subsequent wavelet transforms of the same signal can vary for different mother wavelets. The discrete wavelet transform (DWT) requires the discretization of CWT:

$$W_x(j, r) = 2^{-j/2} \int x(t) \psi(2^{-j}t - r) dt = \langle x(t), \psi_{j,r}(t) \rangle, \quad (9)$$

where  $\Psi_{j,r}(t) = 2^{-j/2} \Psi(2^{-j}t - r)$  is called a dyadic wavelet.

### 3. Feature Selection Methods

**3.1. Recursive Feature Elimination.** RFE works by searching a subset of features starting from all features of the training data and successfully removing features until a desired number of features are retained in performing classifications by SVM [22]. This is achieved by fitting a given mathematical model, ranking the features by their importance, discarding unimportant features, and refitting the model recursively. This process is repeated until a specific number of features are retained [23]. The description of the RFE algorithm is described as follows:

- (1) The training samples are  $x = \{x_1, x_2, \dots, x_k, \dots, x_n\}^T$ . The class labels are

$y = \{y_1, y_2, \dots, y_k, \dots, y_n\}^T$ . External estimators are selected as the basis of constructing an SVM.

- (2) The estimators are trained using the squared weight coefficients  $w^2$  as the feature importance criterion. The formula for the weight vector is as follows:

$$w = \sum_k \alpha_k y_k x_k, \quad (10)$$

where  $\alpha_k$  is the Lagrange multiplier.

- (3) The feature weight values are ranked and the features with the smallest contribution are removed, one feature at a time.
- (4) Steps 2 and 3 are repeated until the number of features reaches a specified threshold.

**3.2. Principal Component Analysis.** The main purpose of PCA is to explain most of the variation in the original data with fewer variables by transforming many highly correlated variables into variables that are uncorrelated with each other [24]. Usually, a few new variables, called principal components, are selected to explain most of the variation in the data instead of using all the original variables. PCA attempts to reduce the dimensionality of the original variable space, simultaneously losing less information as possible. The detailed procedure of PCA was surveyed in [25].

**3.3. Tree-Based Model.** As a sort of embedded feature filtering algorithm, tree-based model algorithms are based on machine learning theory to analyze the importance of features, so the most important feature can be preserved and selected. Random forests (RF) and extreme trees (ET) are two sorts of typical tree-based model algorithms.

RF is an integrated learning method based on bagging method, and the advantage of this integrated algorithm is that each decision tree is constructed by random variables [26]. The randomness of RF lies in: sample randomization, feature randomization, parameter randomization, and model randomization. ET is a variant of the RF algorithm with a different stochastic process [27]. ET uses all training samples as training samples to build each decision tree, and randomly draws segmentation rules on each node to select the optimal segmentation rules by scores [26]. Therefore, another characteristic of randomness, i.e., split randomness, is introduced. This split randomness greatly enhances the independence between each decision tree, thus, it improves the training speed and generalization ability of the classifiers. ET algorithm is described as follows:

- (1) All original  $D$  training samples are selected as training data input.
- (2) A decision tree is constructed from  $m$  features  $\{f_1, f_2, \dots, f_m\}$  randomly selected from the whole  $M$  features without replacement. In general,  $m$  is smaller than or equal to  $M$ .
- (3) The division values  $\{d_1, d_2, \dots, d_m\}$  are randomly selected between their maximum and minimum

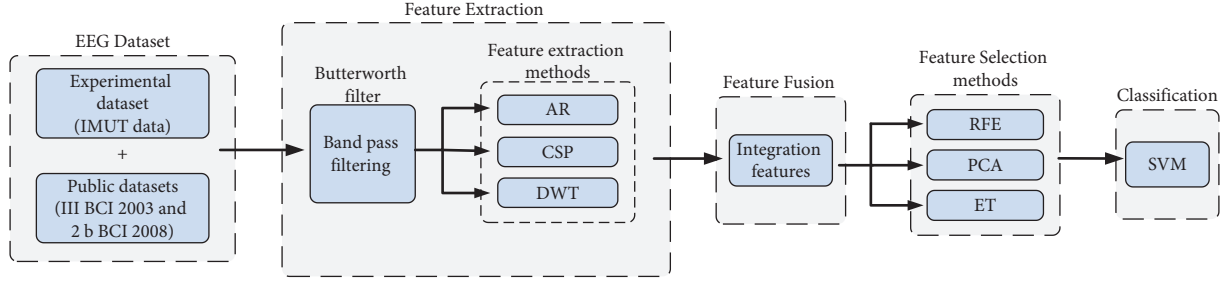


FIGURE 1: Scheme of FMCM-ETFS.

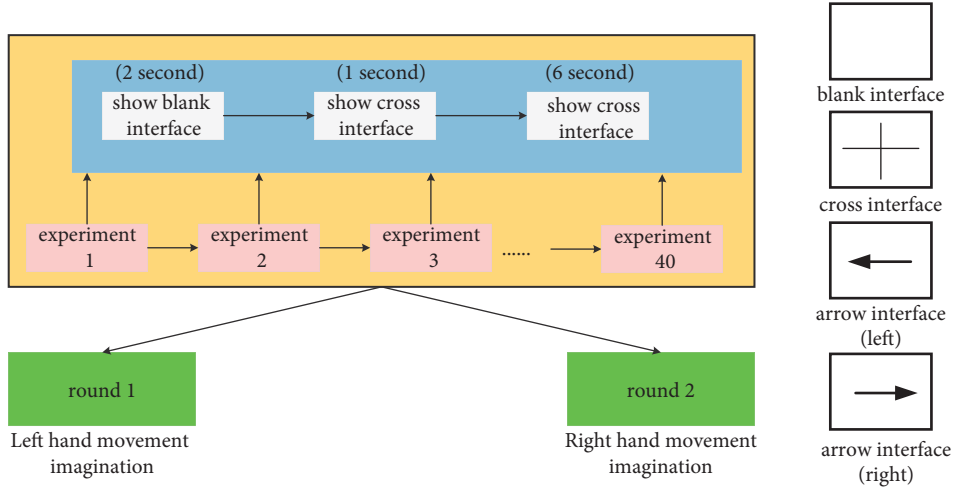


FIGURE 2: Experimental timing diagram of IMUT EEG data recording.

values for each of these  $m$  features sequentially, and the optimal split value  $d_j$  ( $j = 1, 2, \dots, m$ ) is selected by a particular normalization of the information gain. The formula for the particular normalization of the information gain is as follows:

$$\text{Score}_C(d, D) = \frac{2I_C^d(D)}{H_d(D) + H_C(D)}, \quad (11)$$

where  $H_C(D)$  is the (log) entropy of the classification in  $D$ ,  $H_d(D)$  is the split entropy, and  $I_C^d(D)$  is the mutual information of the split outcome and the classification [27].

- (4) Steps 2 and 3 are repeated until the multiple decision trees are constructed to obtain an extreme forest.

#### 4. Algorithm Structure of FMCM-ETFS

**4.1. Overview of FMCM-ETFS Algorithm Structure.** The proposed FMCM-ETFS includes two procedures: 1. The experiment procedure collecting MI EEG dataset from IMUT (6 subjects) or the direct use of dataset III from the 2nd BCI competition 2003 (1 subject) and dataset 2b from the 4th BCI competition 2008 (5 subjects); 2. The data processing procedure consisting of data processing, feature extraction, feature selection, and pattern classification. A scheme of FMCM-ETFS is shown in Figure 1. For the data

preprocessing, the collected EEG signals are band-pass filtered by a Butterworth filter since most of the response frequency band of MI EEG signal is 8–30 Hz. Therefore, the passband frequency of the filter is set at 8–30 Hz. At the same time, the order of the filter is set at 4. For the feature extraction, three methods: AR, CSP, and DWT, are used to obtain the features of EEG signals both in the time-frequency domain and the spatial domain. Then, all the extracted features are fused to obtain a whole feature vector. In the subsequent feature extraction, the fused features are selected by three methods: RFE, PCA, and ET to obtain the optimal subset of features and eliminate irrelevant and redundant features. Finally, the screened features are classified by a SVM, and the performance of the three feature selection methods is compared in terms of classification accuracy.

**4.2. Feature Fusion Method.** MI EEG signals always include complex and diverse information from various brain regions. However, extracting features from one dimension alone does not reflect this comprehensive information. Therefore, this paper uses a feature fusion algorithm to obtain fused features in time, frequency, and spatial domains, which represent three-dimensional characteristics of brain networks. First, the coefficients of the AR model are extracted by the AR algorithm to construct a feature vector, so the time-domain features of the EEG signals can be

TABLE 1: The statistics of experimental datasets from IMUT and BCI competitions (Graz University of Technology).

Dataset	Experimental dataset		Public datasets	
	IMUT data	III BCI 2003	2b BCI 2008	
Number of subjects	6	1	5	
Number of channels	3	3	3	
Number of experiments (times/person)	80	280	240	

TABLE 2: Classification accuracy for IMUT data using different feature extraction methods.

Dataset	Experimental dataset					
	IMUT data					
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
AR + SVM	0.660	0.641	0.575	0.486	0.720	0.721
CSP + SVM	0.638	0.567	0.656	0.569	0.511	0.778
DWT + SVM	0.558	0.630	0.498	0.518	0.650	0.780
CSP + DWT + SVM	0.613	0.763	0.825	0.613	0.625	0.800
AR + DWT + SVM	0.725	0.800	0.513	0.600	0.825	0.838
AR + CSP + SVM	0.675	0.850	0.788	0.525	0.813	0.875
AR + CSP + DWT + SVM	<b>0.700</b>	<b>0.837</b>	<b>0.812</b>	<b>0.600</b>	<b>0.825</b>	<b>0.875</b>

TABLE 3: Classification accuracy for datasets III BCI 2003 and 2b BCI 2008 using different feature extraction methods.

Dataset	Public datasets					
	III BCI 2003			2b BCI 2008		
	aa	bb	cc	dd	ee	ff
AR + SVM	0.622	0.416	0.397	0.649	0.551	0.596
CSP + SVM	0.420	0.493	0.580	0.786	0.541	0.622
DWT + SVM	0.451	0.573	0.668	0.742	0.657	0.585
DWT + CSP + SVM	0.693	0.659	0.775	0.808	0.617	0.658
AR + DWT + SVM	0.636	0.725	0.667	0.792	0.758	0.700
AR + CSP + SVM	0.650	0.783	0.633	0.808	0.683	0.675
AR + CSP + DWT + SVM	<b>0.650</b>	<b>0.725</b>	<b>0.742</b>	<b>0.817</b>	<b>0.733</b>	<b>0.675</b>

TABLE 4: Classification accuracy for IMUT data of using different feature selection methods.

Dataset	Experimental dataset					
	IMUT data					
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
RFE + SVM	0.700	0.850	0.775	0.600	0.825	0.875
PCA + SVM	0.700	0.837	0.825	0.600	0.825	0.875
ET + SVM	<b>0.737</b>	<b>0.862</b>	<b>0.862</b>	<b>0.625</b>	<b>0.850</b>	<b>0.887</b>

obtained initially. Second, the spatial domain features of the EEG signals are obtained by CSP with the variance of each channel as the feature vector. Finally, the time-frequency domain features of the EEG signals were extracted by DWT using Daubechies class db4 wavelets for 3-layer decomposition, and the mean value of the squared coefficients in layer 3 of each channel was used as the feature vector.

**4.3. Feature Selection Method.** The fused features usually have some irrelevant and redundant features while the information content accumulates. This makes the classification model too complex and overfitted, which causes the prediction accuracy to decrease. To address this difficulty, this

TABLE 5: Classification accuracy for datasets III BCI 2003 and 2b BCI 2008 using different feature selection methods.

Dataset	Public datasets					
	III BCI 2003			2b BCI 2008		
	aa	bb	cc	dd	ee	ff
RFE + SVM	0.650	0.750	0.742	0.833	0.675	0.708
PCA + SVM	0.664	0.725	0.742	0.808	0.733	0.675
ET + SVM	<b>0.692</b>	<b>0.783</b>	<b>0.775</b>	<b>0.850</b>	<b>0.750</b>	<b>0.717</b>

paper adopts three feature selection algorithms: RFE, PCA, and ET, to screen the fused features and reduce the dimensionality of the features while ensuring the necessary information content of the features is preserved.

REF uses a support vector regression (SVR) model to train the EEG data and removes one feature at a time, which is based on the weight coefficients in the models, until 20 features are left. In PCA, the number of principal components is set at 20, and then dimensionality reduction on the fused features is performed subsequently. The last applied approach is ET, which is a sort of tree-based model algorithm. The performance of ET depends on the adjustment of three parameters, i.e.  $m$ ,  $n_{\min}$ , and  $R$ . The notation  $m$  denotes

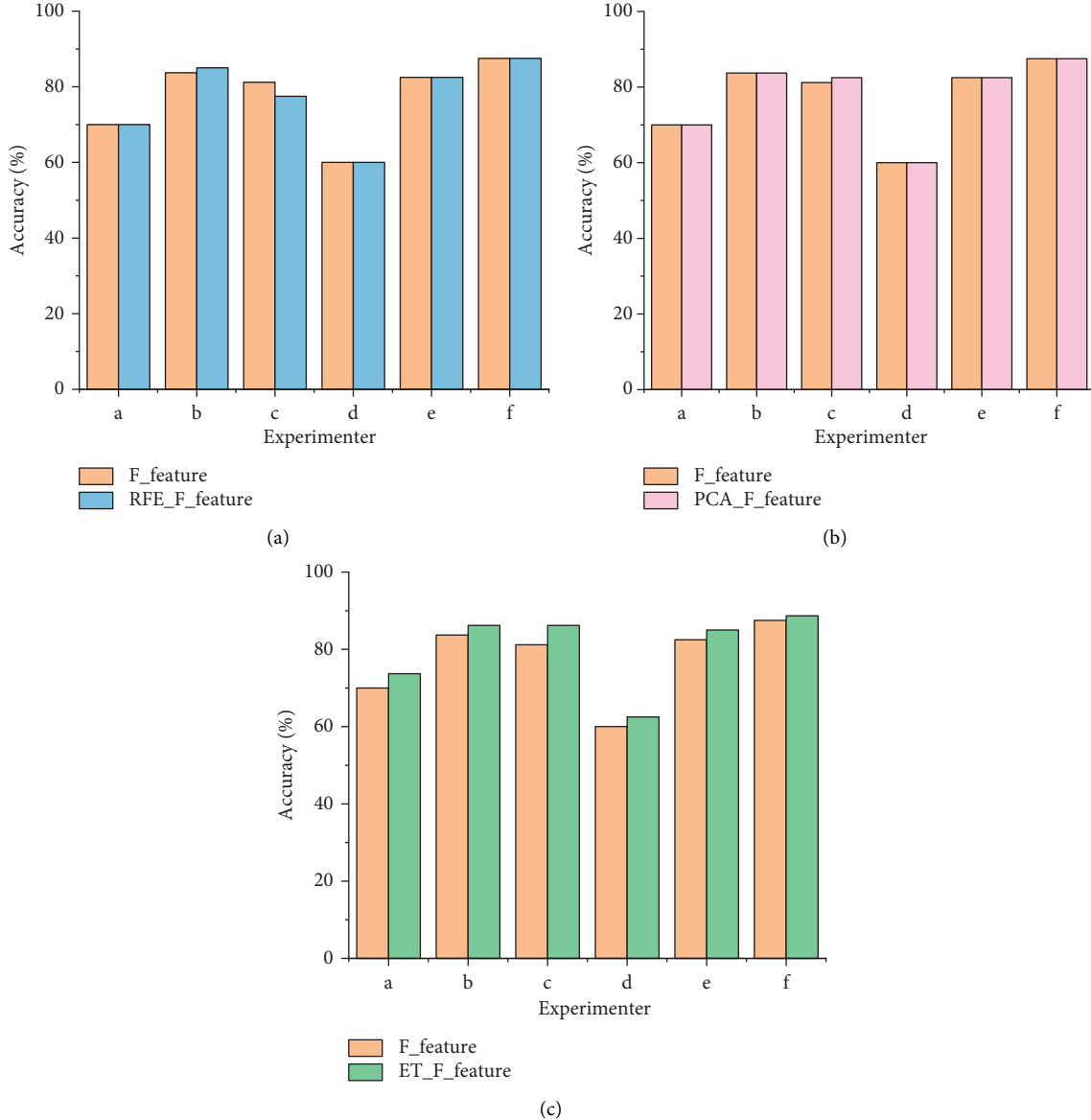


FIGURE 3: The comparison of classification accuracies with feature selection and without any feature selections for IMUT data. (a) Before and after RFE; (b) before and after PCA; (c) before and after ET.

the total number of randomly selected features for each node, which is usually set to  $\sqrt{M}$ . The notation  $M$  denotes the total number of features. This default setting has proven to be the optimal way to solve various problems [27]. The notation  $n_{\min}$  represents the minimum sample size for the splitting nodes. Smaller values of  $n_{\min}$  result in deeper forests, and in this paper  $n_{\min} = 2$  is used as a default value.  $R$  denotes the number of trees. In practice, the larger the value of  $R$  is, the higher the accuracy can be arrived [24]. However, a large tree number usually causes the algorithm difficult to converge. In this paper,  $R$  is at 10. Finally, the classification accuracies of the three feature elimination methods are compared to determine the choice of elimination methods.

**4.4. Classification Algorithm.** There are four kernel functions, for examples, linear kernel, polynomial kernel,

Sigmoid kernel, and radial basis kernel, commonly used to construct different classifiers in SVM methods [28]. In this paper, the following linear kernel function is used:

$$k(x_i, x_j) = x_i \cdot x_j, \quad (12)$$

where  $x_i$  and  $x_j$  denote the  $i$ -th and  $j$ -th samples, respectively. For the SVM classifier, the error penalty factor  $C$  is a major parameter that affects the performance of the SVM classifier. The parameter is determined by a grid search and validated by a 10-fold cross-validation.

## 5. Experiment

**5.1. Experimental Data Set.** The dataset used in this paper comes from three sources: 1. MI EEG dataset III in 2003 BCI competition from Graz University of Technology; 2. MI EEG

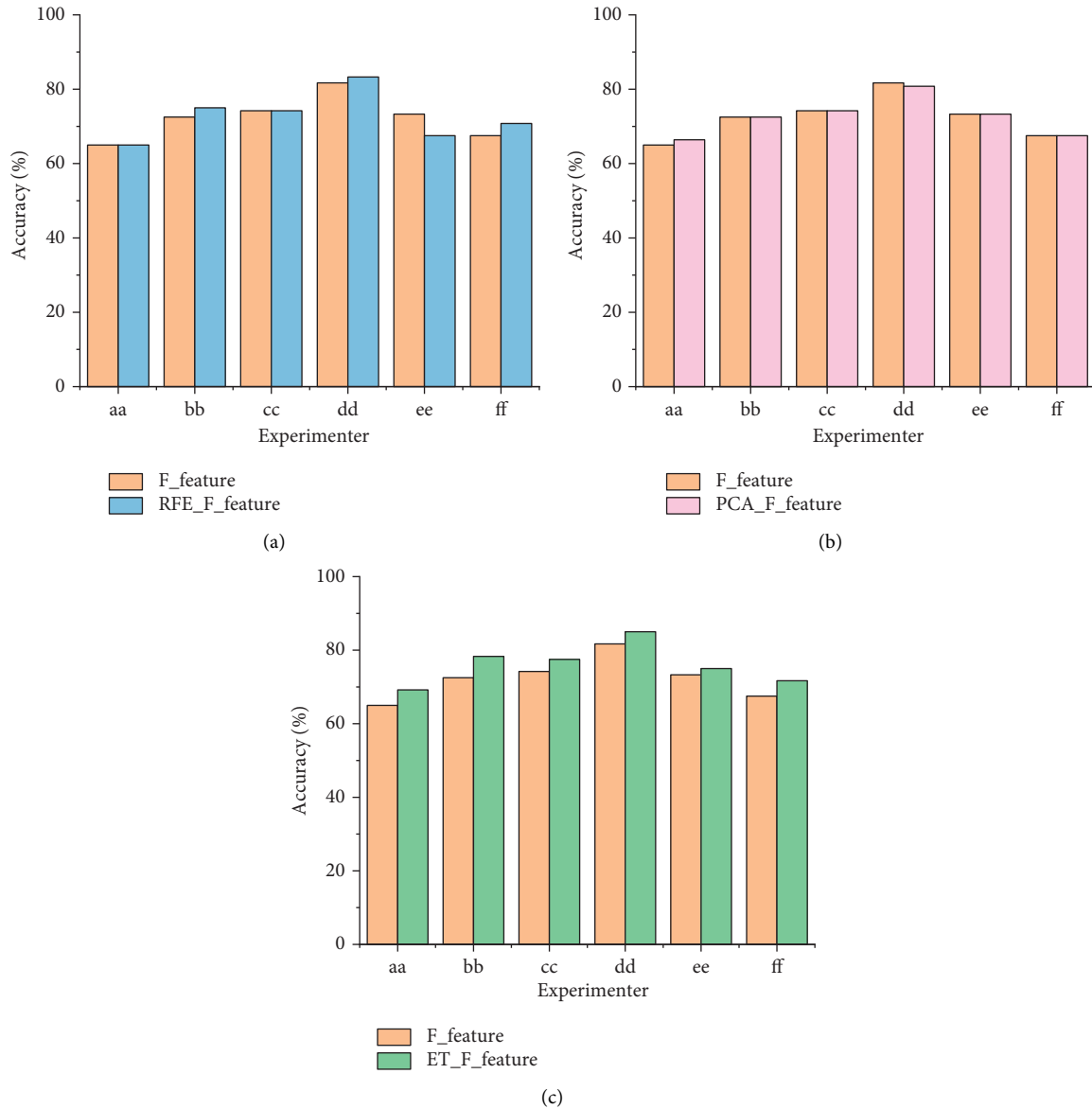


FIGURE 4: The comparison of classification accuracies with feature selection and without any feature selections for datasets III BCI 2003 and 2b BCI 2008. (a) Before and after RFE; (b) before and after PCA; (c) before and after ET.

dataset 2b in 2008 BCI competition from Graz University of Technology; 3. EEG dataset from Inner Mongolia University of Technology (IMUT data) recorded by a 32-channel EEG acquisition device from Brain Products (BP) Inc., Germany.

The 2003 BCI competition dataset III contains 7 sets of experiments with 40 trials for each set of experiments, totally yielding 280 trials of MI data. The EEG acquisition device consisted of a G.tec amplifier and Ag/AgCl electrodes to acquire EEG data with a sampling frequency of 128 Hz, and the EEG data of 3 channels namely C3, C4, and Cz were recorded. The experimental data were divided into two parts, i.e., the training set and the testing set, each of which included 140 trials of experiments (70 experiments for left-handed MI and 70 experiments for right-handed MI).

The 2008 BCI competition dataset 2b contains 6 runs with ten trials per run. Each subject participated in two

screening sessions without feedback, recorded on two different days within two weeks. This resulted in 20 trials per run and 120 trials per session. Data of 240 repetitions of each MI class were available for each subject in total. An EEG acquisition device acquires the EEG data with a sampling frequency of 250 Hz, and the EEG data of 3 channels C3, C4, and Cz were recorded.

The data from IMUT were collected using a 32-channel EEG acquisition device from BP Inc. at a sampling frequency of 500 Hz. The subjects include six males around 25 years of age, and the experiments were conducted in a quiet environment. The experimental timing diagram is shown in Figure 2. At the beginning of the experiment ( $t=0$ s), the screen was black and the subject remained at rest; two seconds later, the screen appeared a “+” cross, prompting the subject to get ready; one second later, the screen appeared a

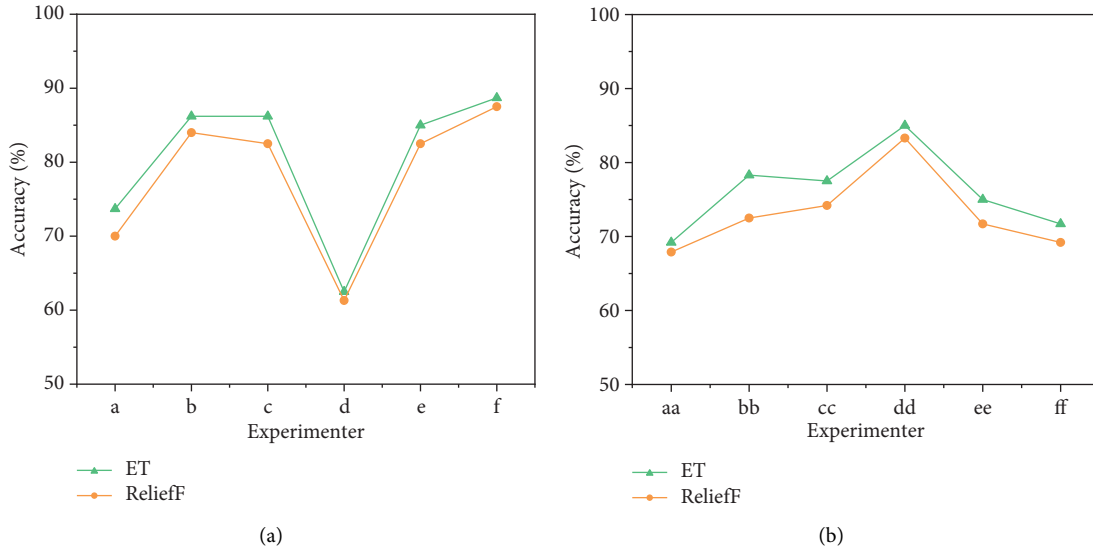


FIGURE 5: Comparison of classification accuracies between ET and ReliefF for IMUT data, datasets III BCI 2003 and 2b BCI 2008. (a) IMUT data; (b) datasets III BCI 2003 and 2b BCI 2008.

left or right arrow lasting 6 s prompting the subject to start MI of the left or right hand. A total of 80 experiments were performed for each set of experimental data, including 40 left-handed MI experiments and 40 right-handed MI experiments. The details of the data are shown in Table 1.

**5.2. Experimental Results.** Table 2 provides a comparison of the classification accuracies between using AR features, CSP features, DWT features, and AR + CSP + DWT features, which reflect the multidimensional fused features for the laboratory-collected IMUT data. The accuracy of using AR + CSP + DWT features is on average 14.1% higher than that of using AR features, 15.5% higher than that of using CSP features, and 16.9% higher than that of using DWT features. The datasets III BCI 2003 and 2b BCI 2008 also validates a similar trend that the average classification accuracy using AR + CSP + DWT features is higher than the average classification accuracy using a single category of features. The results for the two public datasets are presented in Table 3. The accuracy of using AR + CSP + DWT features is on average 18.5% higher than that of using AR features, 15% higher than that of using CSP features, and 11.1% higher than that of using DWT features. We have deleted the feature categories one by one in the sequence of AR, CSP, and DWT. For IMUT data, the average classification accuracies are as follows: 0.707 for CSP + DWT + SVM, 0.717 for AR + DWT + SVM, and 0.754 for AR + CSP + SVM (see Table 2). Compared to AR + CSP + DWT + SVM, CSP + DWT + SVM shows a decrease of 0.068, which is the largest decrease magnitude among the other two decreases of CSP + DWT + SVM and AR + DWT + SVM. Therefore, the feature category of AR contributes most to the proposed AR + CSP + DWT + SVM in terms of classification accuracy. The same conclusion can also be obtained for datasets III BCI 2003 and 2b BCI 2008 (see Table 3).

Table 4 shows a comparison of the classification accuracies using three feature selection algorithms (RFE, PCA, and ET) for the laboratory-collected IMUT data. ET works better than REF and PCA because the classification accuracy of ET + SVM is on average 3.3% ( $p = 0.0346 < 0.05$ ,  $t$ -test) higher than that of RFE + SVM and 2.68% ( $p = 8.9771e-4 < 0.01$ ,  $t$ -test) higher than that of PCA + SVM. A similar trend also applied to the datasets III BCI 2003 and 2b BCI 2008. The experimental results are demonstrated in Table 5, where ET + SVM outperforms 3.5% ( $p = 0.014 < 0.05$ ,  $t$ -test) than RFE + SVM and 3.65% ( $p = 0.0014 < 0.01$ ,  $t$ -test) higher than PCA + SVM. The advantage of feature selections also is shown in Figures 3 and 4. Figures 3 and 4 show a comparison of the classification accuracy with and without the feature selection algorithm under the IMUT data, datasets III BCI 2003 and 2b BCI 2008, respectively. In Figures 3 and 4, the classification accuracy using the fused features selected by ET is significantly higher than that using the fused features without any feature selection. F\_feature represents using the fused features without any selection. RFE\_F\_feature represents using the fused features with RFE model selection, PCA\_F\_feature represents using the fused features with PCA selection, and ET\_F\_feature represents using the fused features with ET model selection. A similar situation also happened when using RFE and PCA as feature selection methods.

To further verify the effectiveness of the proposed FMCM-ETFS, another feature selection algorithm ReliefF proposed in the literature [29], was selected to compare with the ET feature selection method. The comparison result is detailed in Figure 5, where the average classification accuracy of the ET + SVM algorithm for IMUT data is 2.43% ( $p = 0.0029 < 0.01$ ,  $t$ -test) higher than that of ReliefF + SVM algorithm (see Figure 5(a)). The same trend retains in datasets III BCI 2003 and 2b BCI 2008, where the average classification accuracy of the ET + SVM algorithm was 2.99%



( $p = 0.0061 < 0.01$ ,  $t$ -test) higher than that of the ReliefF + SVM algorithm (see Figure 5(b)).

## 6. Conclusion

In this paper, we propose a fused multidimensional classification method based on extreme tree feature selection in the task of discerning MI EEG. The fused multidimensional features include the features extracted by AR, CSP, and DWT algorithms that reflect the complex information of MI EEG signals from various dimensions: time, frequency, and space domain. Furthermore, to avoid the overfitted model problem, three feature selection methods, RFE, PCA, and ET are applied to lower the complexity of the machine learning models. The experiment results for the datasets from IMUT and Graz University show that the accuracy of using AR + CSP + DWT features is 3%–23% higher than that of using the single-dimensional features and ET has 3%–4% higher accuracy than the other two feature selection algorithms. This result of the study convergingly verifies that a fused multidimensional classification method based on extreme tree feature selection (FMCM-ETFS) can significantly improve the performance of the MI EEG classifiers.

## Data Availability

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

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61364018 and 61863029), Inner Mongolia Natural Science Foundation (2016JQ07, 2020MS06020, and 2021MS06017), Inner Mongolia Scientific and Technological Achievements Transformation Project (CGZH2018129), and Science and Technology Plan Project of the Inner Mongolia Autonomous Region (Key Technology Project).

## References

- [1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks et al., “Brain-computer interface technology: a review of the first international meeting,” *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2000.
- [2] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [3] J. R. Millán and J. Carmena, “Invasive or noninvasive: understanding brain-machine interface technology,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, pp. 16–22, 2010.
- [4] M. Radman, A. Chaibakhsh, N. Nariman-zadeh, and H. He, “Feature fusion for improving performance of motor imagery brain-computer interface system,” *Biomedical Signal Processing and Control*, vol. 68, Article ID 102763, 2021.
- [5] Y. Wang, Bo Hong, X. Gao, and G. Shangkai, “Implementation of a brain-computer interface based on three states of motor imagery,” in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5059–5062, IEEE, Lyon, France, October 2007.
- [6] G. Pfurtscheller, “Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest,” *Electroencephalography and Clinical Neurophysiology*, vol. 83, no. 1, pp. 62–69, 1992.
- [7] J. Li, J. Liang, Q. Zhao, J. Li, K. Hong, and L. Zhang, “Design of assistive wheelchair system directly steered by human thoughts,” *International Journal of Neural Systems*, vol. 23, no. 3, Article ID 1350013, 2013.
- [8] J. Wang and C. Chen, “Multi-channel EEG feature extraction using hierarchical vector autoregression,” *Acta Automatica Sinica*, vol. 42, no. 08, pp. 1215–1226, 2016.
- [9] L. Liu, J. Chen, Li Song, and L. XU, “Feature extraction of EEG in motion imagery based on phase synchronization and AR,” *Software guide*, vol. 17, no. 03, pp. 7–10, 2018.
- [10] N. Yahya, H. Musa, Z. Y. Ong, and I. Elamvazuthi, “Classification of motor functions from electroencephalogram (EEG) signals based on an integrated method comprised of common spatial pattern and wavelet transform framework,” *Sensors*, vol. 19, no. 22, p. 4878, 2019.
- [11] J. Feng, E. Yin, J. Jin et al., “Towards correlation-based time window selection method for motor imagery BCIs,” *Neural Networks*, vol. 102, pp. 87–95, 2018.
- [12] B. Liu, M. Cai, Y. Bo, and X. Zhang, “A feature extraction and classification algorithm based on PSO-CSP-SVM for motor imagery EEG signals,” *Journal of Central South University (Science and Technology)*, vol. 51, no. 10, pp. 2855–2866, 2020.
- [13] X. Gu, Z. Wei, H. Liu, and P. Shen, “EEG signal classification based on typical spatial pattern and convolutional neural network,” *Laser Journal*, vol. 42, no. 04, pp. 100–104, 2021.
- [14] Yu Pei, Z. Luo, H. Zhao et al., “A tensor-based frequency features combination method for brain-computer interfaces,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 465–475, 2022.
- [15] L. Wu, H. Lu, and N. Gao, “The comparison of EEG feature extraction of motor imagery between CSP algorithm and wavelet packet analysis,” *Journal of Biomedical Engineering Research*, vol. 36, no. 3, pp. 224–228, 2017.
- [16] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, “Brain-computer communication: motivation, aim, and impact of exploring a virtual apartment,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 4, pp. 473–482, 2007.
- [17] E. Maiorana, D. La Rocca, and P. Campisi, “On the permanence of EEG signals for biometric recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 163–175, 2016.
- [18] S. Ma, C. Dong, T. Jia, X. Zhiyun, C. Xiaoyan, and L. Zhang, “A feature extraction algorithm of brain network of motor imagination based on a directed transfer function,” *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 4496992, 2022.
- [19] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, “Designing optimal spatial filters for single-trial EEG classification in a movement task,” *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
- [20] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand

- movement,” *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [21] S. Ma, C. Dong, T. Jia, and C. Xiaoyan, “Classifying motor-imagination signals in brain-computer interface based on feature extraction of parametric AR model,” in *Proceedings of the In 2020 39th Chinese Control Conference (CCC)*, pp. 6291–6294, IEEE, Shenyang, China, July 2020.
- [22] 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.
- [23] S. Liu, J. Meng, J. Yang, X. Zhao, and H. Feng, “Within-stimulus Emotion Recognition May Inflate the Classification Accuracies Based on EEG signals,” in *Proceedings of the in 2015 IEEE 7th International Conference on Awareness Science and Technology (iCAST)*, pp. 115–118, IEEE, Qinhuangdao, China, September 2015.
- [24] J. Lever, M. Krzywinski, and N. Altman, “Points of significance: principal component analysis,” *Nature Methods*, vol. 14, no. 7, pp. 641–642, 2017.
- [25] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [26] C. Désir, C. Petitjean, L. Heutte, M. Salaun, and L Thiberville, “Classification of endomicroscopic images of the lung based on random subwindows and extra-trees,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2677–2683, 2012.
- [27] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [28] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [29] R. Moein, C. Ali, Z. N. Nariman, and H. Huiguang, “Feature fusion for improving performance of motor imagery brain-computer interface system,” *Biomedical Signal Processing and Control*, vol. 68, 2021.