

Research Article

Motor Imagery EEG Decoding Based on New Spatial-Frequency Feature and Hybrid Feature Selection Method

Yuan Tang,¹ Zining Zhao,¹ Shaorong Zhang ,^{1,2} Zhi Li,^{1,2} Yun Mo,² and Yan Guo²

¹School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi, China

²School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin, Guangxi, China

Correspondence should be addressed to Shaorong Zhang; zsrong@guat.edu.cn

Received 21 October 2021; Accepted 27 November 2021; Published 7 January 2022

Academic Editor: Xiaofeng Li

Copyright © 2022 Yuan Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Feature extraction and selection are important parts of motor imagery electroencephalogram (EEG) decoding and have always been the focus and difficulty of brain-computer interface (BCI) system research. In order to improve the accuracy of EEG decoding and reduce model training time, new feature extraction and selection methods are proposed in this paper. First, a new spatial-frequency feature extraction method is proposed. The original EEG signal is preprocessed, and then the common spatial pattern (CSP) is used for spatial filtering and dimensionality reduction. Finally, the filter bank method is used to decompose the spatially filtered signals into multiple frequency subbands, and the logarithmic band power feature of each frequency subband is extracted. Second, to select the subject-specific spatial-frequency features, a hybrid feature selection method based on the Fisher score and support vector machine (SVM) is proposed. The Fisher score of each feature is calculated, then a series of threshold parameters are set to generate different feature subsets, and finally, SVM and cross-validation are used to select the optimal feature subset. The effectiveness of the proposed method is validated using two sets of publicly available BCI competition data and a set of self-collected data. The total average accuracy of the three data sets achieved by the proposed method is 82.39%, which is 2.99% higher than the CSP method. The experimental results show that the proposed method has a better classification effect than the existing methods, and at the same time, feature extraction and feature selection time also have greater advantages.

1. Introduction

Motor imagery electroencephalogram (EEG) signal is widely used in brain-computer interface (BCI) system, but it has strong randomness and low signal-to-noise ratio and is easily disturbed by physiological and nonphysiological noises, which makes it difficult to decode [1]. In EEG decoding, feature extraction and selection are the core components [2]. On the one hand, extracting discriminative and stable features can effectively improve the performance of EEG decoding [3]. On the other hand, the extracted features usually contain noise and redundant information, so feature selection is required to eliminate invalid information [4]. In addition, feature selection can reduce the feature dimension and the complexity of the classification model and avoid dimension disaster and overfitting. Therefore, feature extraction and selection have always been the focus and difficulty of BCI system research.

Common spatial pattern (CSP) is a relatively effective method for feature extraction of motor imagery EEG among many methods [5]. The traditional CSP method extracts logarithmic variance as features after spatial filtering [6], but some studies have shown that this feature extraction method is not necessarily optimal. For example, literature [7] proposed the logarithmic band power (LBP) feature based on the CSP transform, which is called CSP-LBP in this paper. The experimental results show that CSP-LBP is superior to the traditional CSP method. In addition, the traditional CSP method lacks frequency domain information. Therefore, a lot of work has been done to select the optimal frequency band for CSP. For example, in literature [8], the original EEG signal was filtered into multiple subbands by filter bank method, and then CSP was used to extract features. Finally, mutual information was used to select the features of the optimal frequency band. Zhang et al. [9] proposed a sparse

filter band common spatial pattern (SFBCSP) method. SFBCSP carried out band-pass filtering on original EEG signals through multiple frequency subbands with a frequency range of 4–40 Hz, a bandwidth of 4 Hz, and an overlap rate of 2 Hz between subbands. CSP was used to extract features on each subband, and the least absolute shrinkage and selection operator (LASSO) was used for frequency band feature selection. Finally, the support vector machine (SVM) was used to classify selected features. Subsequently, Zhang et al. [10] proposed a subband optimization method that implements sparse Bayesian learning of frequency bands (SBLFB) for motor imagery classification. The subbands filtering method was the same as that in literature [9], but sparse Bayesian learning was used to select sparse frequency band features. The above spatial-frequency feature extraction methods filter the original EEG signals into multiple subbands, which requires a large amount of computation and a long time.

Existing feature selection methods mainly include filter, wrapper, and embedded [11]. The filter feature selection method uses evaluation criteria such as information measurement and distance measurement to select features. The wrapper feature selection method generates feature subsets in a specific way and then uses the results of classifiers as the evaluation criteria for feature selection. The embedded feature selection method can automatically remove some features during classifier training, so feature selection and classification can be carried out simultaneously. The above three types of methods have their advantages and disadvantages, and the organic combination of these methods can achieve complementary advantages. Therefore, the hybrid feature selection method has been studied widely in recent years. Moradi et al. [12] proposed a novel hybrid feature selection algorithm based on particle swarm optimization (PSO) and the local search strategy, and the local search strategy was embedded in the PSO to select the less correlated and salient feature subset. Jain et al. [13] proposed a hybrid model for gene selection and cancer classification, and the optimal gene subset was selected by correlation-based feature selection method combined with improved binary PSO. Lu et al. [14] proposed a hybrid feature selection algorithm for gene expression data classification. The algorithm combined the mutual information maximization and the adaptive genetic algorithm to reduce the dimension of gene expression data and remove the redundancies for classification. Ghareb et al. [15] combined six filtering feature selection methods and an improved genetic algorithm to form a new hybrid feature selection method. Literature [16] is a review article that comprehensively introduces the hybrid feature selection method for cancer classification. Although the hybrid feature selection method has been widely used, as far as we know, few hybrid feature selection methods have been applied to EEG decoding. In addition, the existing hybrid feature selection methods are mostly based on intelligent optimization algorithms such as PSO and genetic algorithms, so the feature selection time is relatively long.

In order to further improve the performance of motor imagery EEG decoding, a new spatial-frequency feature extraction method and hybrid feature selection method are

proposed in this paper. First, considering the effectiveness of the CSP-LBP method [7] and the frequency defects of CSP, a new spatial-frequency feature extraction method based on CSP transform, filter bank (FB), and logarithmic band power (LBP) is proposed; we call it CSP-FBLBP. The original EEG signals are preprocessed and then spatially filtered by CSP transform. After that, the spatially filtered signals are decomposed into multiple subbands using a filter bank, and the logarithmic band power of each subband is extracted as the feature. Second, a new hybrid feature selection method based on Fisher score (F-score) and SVM is proposed to select subject-specific spatial-frequency features; we call it F-score-h. The Fisher score of each feature is calculated, then a series of threshold parameters are set to generate different feature subsets, and finally, SVM and 10-fold cross-validation are used to select the optimal feature subset. After feature extraction and selection are completed, SVM is used for classification. Two public data sets and a self-collected data set are used to verify the proposed method.

The main contributions of this paper are in two aspects:

First, a new spatial-frequency feature extraction method is proposed. CSP is used for dimensionality reduction, and then the spatial projection signal is band-pass filtered using the filter bank method. Finally, logarithmic band power is used for feature extraction. CSP dimensionality reduction effectively reduces the number of signal channels, thereby reducing the calculation amount and time of band-pass filtering, which greatly improves the timeliness of feature extraction. In addition, the experimental results show that logarithmic band power is more effective than the logarithmic variance of the traditional CSP method.

Second, a new hybrid feature selection method is proposed. The Fisher score is used for feature sorting, and the threshold method, SVM, and cross-validation are combined for optimal feature subset selection. The proposed method takes full advantage of the simple calculation of the filtering method and the supervised selection of the wrapped method, which not only reduces the feature selection time but also improves the classification performance of EEG decoding.

2. Materials and Methods

2.1. EEG Data Description. Data set 1: data set IIa of BCI competition IV (2008) [17]: this data set contains 22 electrode channels, and the sampling rate is 250 Hz. Nine healthy subjects (A01, A02, A03, A04, A05, A06, A07, A08, and A09) performed left-hand, right-hand, foot, and tongue motor imagery tasks, respectively. Since we only consider binary classification tasks, $C_4^2 = 6$ groups of binary classification tasks are obtained by permutation and combination of four types of tasks. Since there are nine subjects, $9 \times 6 = 54$ data subsets could be obtained. The number of samples in the training set and test set of each subject is 144, respectively.

Data set 2: data set IIb of BCI competition IV (2008) [18]: this data set contains 3 electrode channels, and the sampling rate is 250 Hz. Nine healthy subjects (B01, B02, B03, B04, B05, B06, B07, B08, and B09) performed left-hand and right-hand motor imagery tasks, respectively. This data set has five

sessions, and we only analyzed the data of the third session [10]. The number of samples in the training set and test set of each subject is 80, respectively.

Data set 3: data set self-collected from our laboratory: NuAmps amplifier and electrode cap from Neuroscan company are used for scalp EEG signals collection, and the sampling rate is 250 Hz. A total of 36 electrode channel data in this data set, including 30 channels of EEG data, 4 channels of electrooculogram data, and two reference channels, and only 30 channels of EEG data, are analyzed in this paper. Six healthy subjects (S01, S02, S03, S04, S05, and S06) performed left-hand and right-hand motor imagery tasks, respectively.

2.2. The Proposed Method. The data processing flow of the proposed method is shown in Figure 1, which mainly includes preprocessing, feature extraction, feature selection, and feature classification. In the preprocessing stage, all data sets performed 8–30 Hz band-pass filtering using a 6-order Butterworth filter. The time window with 0.5–2.5 s is selected for single-trial data extraction. In the following content, we will introduce the core work of the proposed method in detail.

2.2.1. The New Spatial-Frequency Feature Extraction Method. The new spatial-frequency feature extraction method includes CSP dimensionality reduction, filter bank band-pass filtering, and logarithmic band power feature extraction.

(1) CSP Dimensionality Reduction. The solution of the CSP objective function can be equivalent to a generalized eigenvalue problem [19]. After the eigenvector matrix is obtained, the eigenvectors corresponding to the first m largest eigenvalues and the last m smallest are selected to form the final spatial filter. Assuming that the spatial filter is \mathbf{W} and the single-trial data is \mathbf{D} , the spatial projection signal \mathbf{Z} can be calculated by the following calculation formula:

$$\mathbf{Z} = \mathbf{W}^T \mathbf{D}, \quad (1)$$

where $\mathbf{W} \in \mathbb{R}^{C \times 2m}$, $\mathbf{D} \in \mathbb{R}^{C \times K}$, C represents the total number of electrode channels, m represents the pair number of the spatial filters, and K represents the number of sampling points for each electrode channel.

After the single-trial data \mathbf{D} is transformed by CSP, the EEG signal has only $2m$ channels. The value of m is usually set to 3 or 1, so the number of EEG signal channels is significantly reduced after CSP transformation. For example,

data set 1 has 22 electrode channels; if m is set to 3, the EEG signal has only 6 channels after CSP dimensionality reduction. The specific form of the signal \mathbf{Z} is as follows:

$$\mathbf{Z} = \begin{bmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,K} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ z_{2m,1} & z_{2m,2} & \cdots & z_{2m,K} \end{bmatrix}. \quad (2)$$

(2) Filter Bank Band-Pass Filtering. The signal \mathbf{Z} is band-pass filtered using a filter bank with frequency subbands of 4–8 Hz, 6–10 Hz, . . . , 26–30 Hz. Specifically, band-pass filtering is performed on each channel of the signal \mathbf{Z} , which is shown in Figure 2.

(3) The Logarithmic Band Power Feature Extraction. The logarithmic variance is extracted as the feature in the traditional CSP method [6]. However, the experimental results in literature [7] prove that the logarithmic band power is more effective. Therefore, after band-pass filtering, the logarithmic band power is extracted as the feature in this paper, specifically as follows [7]:

$$LBP_p = \log \left(\frac{1}{K} \sum_{i=1}^K |Z_p(i)|^2 \right), \quad p = 1, 2, \dots, 2m, \quad (3)$$

where $Z_p(i)$ represents the i -th sample point of the p -th channel of the signals \mathbf{Z} .

In the newly proposed spatial-frequency feature extraction method, CSP spatial filtering is performed first, and then band-pass filtering and feature extraction are performed on the spatially filtered signal. This processing has two advantages. On the one hand, after the signal is spatially filtered by the CSP, the signal quality is improved, and the extracted features are more stable and more discriminative. On the other hand, after CSP dimensionality reduction, the signal channel is greatly reduced, thereby reducing the calculation amount of band-pass filtering. Therefore, the time of feature extraction is greatly reduced, and it is not affected by the actual number of electrode channels.

2.2.2. Hybrid Feature Selection Method. The Fisher score can measure the distinguishing ability of features between two categories [20]. The Fisher score is obtained by calculating the variance ratio between-classes and within-classes of each feature, details as follows:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{(1/(n_+ - 1)) \sum_{k=1}^{n_+} (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + (1/(n_- - 1)) \sum_{k=1}^{n_-} (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2}, \quad (4)$$

where $F(i)$ represents the Fisher score of i -th feature. n_+ and n_- , respectively, represent the number of positive samples

and negative samples, and $n = n_+ + n_-$ represents the number of total samples. \bar{x}_i , $\bar{x}_i^{(+)}$, and $\bar{x}_i^{(-)}$ are, respectively,

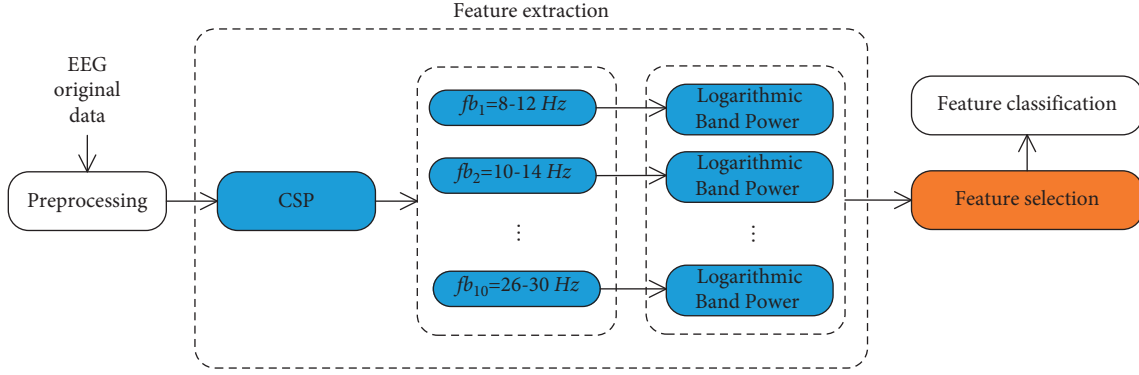
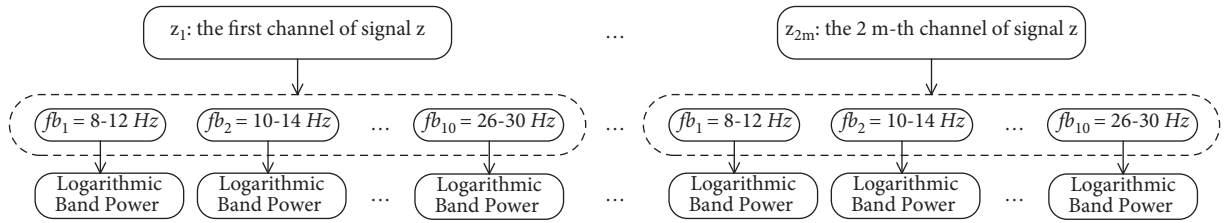


FIGURE 1: The data processing flow of the proposed method.

FIGURE 2: The band-pass filtering on each channel of the signal \mathbf{Z} .

the mean value of the i -th feature in the whole data sample set, the mean value of the i -th feature in the positive sample set, and the mean value of the i -th feature in the negative sample set. $x_k \in \mathbb{R}^p$ represents the k -th feature sample, $x_{k,i}^{(+)}$ is the feature value of the i -th feature of the k -th positive class sample, $x_{k,i}^{(-)}$ is the feature value of the i -th feature of the k -th negative class sample, and p is the feature dimension. The larger the F value, the stronger the discrimination of the corresponding features [21]. The traditional F-score method sorts the features according to the Fisher score and then selects the top K features for subsequent classification.

However, it is difficult to determine exactly how many features should be selected to achieve the best classification effect. Therefore, a hybrid feature selection method based on F-score and SVM classifier is proposed in this paper; we call it F-score-h. Unlike the filter feature selection method based on F-score, F-score-h uses feature weights (i.e., the Fisher score of the feature) to select the optimal feature subset, as shown in Figure 3. Specifically, after the features are sorted by the Fisher score, we set a series of thresholds to generate different feature subsets; the features greater than the set threshold will be selected. The set of candidate parameters for the threshold is $Th \in \{0, 0.05, 0.1, \dots, 0.8\}$. For each threshold parameter, the average verification accuracy of each feature subset is calculated by combining SVM and 10-fold cross-validation (CV). The threshold corresponding to the highest average accuracy is selected, and the optimal feature subset is further selected according to the optimal threshold.

The newly proposed hybrid feature selection method takes advantage of the small amount of calculation of the

filtering method and the supervised selection of the wrapped method, which can take into account the time efficiency and the classification performance for feature selection at the same time.

2.2.3. SVM Classification. SVM is used as the classifier. The SVM classification model used in this paper is as follows [22]:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i, \quad (5)$$

$$\text{subject to } y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, i = 1, \dots, l, n,$$

where $\mathbf{x}_i \in \mathbb{R}^p$ represents the i -th feature sample (feature vector). y_i represents the i -th label. ξ_i represents the i -th slack variable. x_i is mapped into a higher-dimensional space by $\phi(x_i)$, and $C > 0$ is the regularization parameter. Using the primal-dual algorithm to solve (5), the following decision function can be obtained:

$$f(x) = \text{sgn} \left(\sum_{i=1}^n y_i \alpha_i K(x, x_i) + b \right), \quad (6)$$

where α_i is the Lagrange multiplier, $K(x, x_i) = \phi(x)^T \phi(x_i)$ is the kernel function, and $\text{sgn}(\cdot)$ represents a symbolic function. SVM is implemented with the linear kernel using the LIBSVM toolbox [22]. The model parameter of SVM adopts the default setting of the toolbox [22].

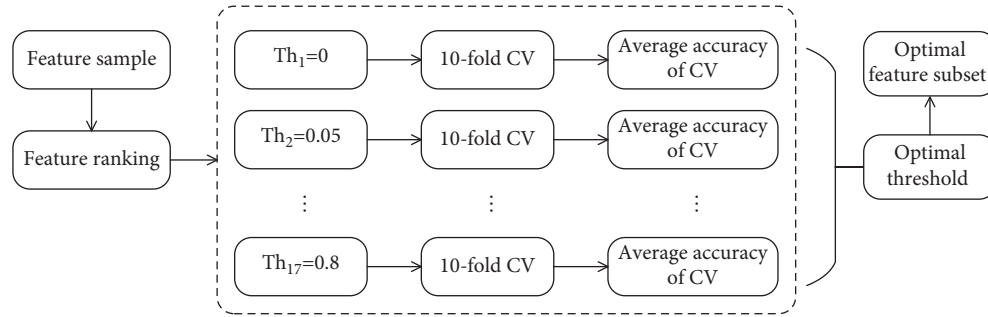


FIGURE 3: The hybrid feature selection method based on F-score and SVM.

3. Results

3.1. Results for Feature Extraction Methods

3.1.1. Comparison Methods and Parameter Settings. In order to verify the effectiveness of the proposed feature extraction method, the proposed method is compared with the other four CSP methods, which are the traditional CSP method [6, 19], CSP-FB [11], SFBCSP [9], SBLFB [10], and CSP-LBP [7]. If there is no special instruction, the pair number of spatial filters for CSP and its improvement methods are set as follows: $m = 3$ for data set 1 and data set 3; $m = 1$ for data set 2; SVM is used for classification. The comparison algorithms and their parameter settings are as follows:

CSP: CSP feature extraction refers to literature [6, 19].

CSP-FB: the parameter setting of the CSP-FB algorithm refers to literature [11]. F-score-h is used to select features.

SFBCSP: the parameter setting of the SFBCSP algorithm refers to literature [9]. Seventeen subbands (4–8 Hz, 6–10 Hz, ..., 36–40 Hz) with a bandwidth of 4 Hz and an overlap rate of 2 Hz are used for band-pass filtering. A 6-order Butterworth filter is used. LASSO is used to select sparse band features.

SBLFB: the parameter setting of the SBLFB algorithm refers to literature [10]. The subbands setting is the same as SFBCSP. Sparse Bayesian learning is used to select sparse band features.

CSP-LBP: CSP-LBP feature extraction refers to literature [7].

CSP-FBLBP: CSP-FBLBP is used for feature extraction, and F-score-h is used for feature selection.

3.1.2. Experimental Results. Tables 1–3, respectively, show the classification accuracy and the total average classification accuracy of all the subjects in the three data sets. The highest accuracy is marked in bold. In Table 1, the left-hand, right-hand, foot, and tongue motor imagery tasks in data set 1 were represented by letters L , R , F , and T , respectively. L versus R means left-hand and right-hand binary classification tasks, and the others can be deduced by analogy. Due to space constraints, only the average classification accuracy of each binary classification task is given. It can be seen from Table 1 that CSP-FBLBP achieves the highest average classification accuracy on data set 1, and the accuracy is 3.24% higher than that of CSP. CSP-LBP and CSP-FB are

better than CSP, but SFBCSP and SBLFB are lower than CSP. Similarly, CSP-FBLBP also achieved the highest average classification accuracy on data sets 2 and 3; see Table 2 and Table 3 for details.

In order to make a more intuitive comparison of the classification effect achieved by various methods, the average classification accuracy achieved by different feature extraction methods is shown in Figure 4. It can be seen from Figure 4 that the classification effect of CSP-FBLBP is significantly better than other methods. The total average classification accuracy of CSP, CSP-FB, SFBCSP, SBLFB, CSP-LBP, and CSP-FBLBP in all data is 79.40, 80.53, 75.88, 75.63, 80.01, and 82.39, respectively.

Furthermore, the distribution of classification accuracy achieved by various feature extraction methods is shown in Figure 5. The red line represents the median value of classification accuracy. It can be seen that the median value of CSP-FBLBP is higher than that of other methods. The maximum value of CSP-FBLBP is 100%, and the minimum value of CSP-FBLBP is also higher than other methods. In addition, the accuracy distribution of CSP-FBLBP is relatively compact and close to the top. Therefore, CSP-FBLBP is superior to other methods.

In order to fully reflect the advantages of CSP-FBLBP, we further studied the time efficiency of CSP-FBLBP; the running time of various feature extraction methods is shown in Table 4. The training sets of the three subjects (A01, B01, and S01) are selected to calculate the feature extraction time. The feature extraction time of CSP-FB, SFBCSP, SBLFB, and CSP-FBLBP includes two parts, namely, CSP spatial filtering time and band-pass filtering time. For CSP-FB, SFBCSP, SBLFB, and CSP-FBLBP methods, two types of time are listed in brackets. The first one represents the spatial filtering time, and the last one represents the band-pass filtering time. It can be seen from Table 4 that the feature extraction time of SFBCSP and SBLFB is the longest, mainly because their band-pass filtering is relatively time-consuming. It is worth pointing out that the feature extraction process of SFBCSP and SBLFB is the same, so the feature extraction time is the same. Although the feature extraction time of CSP-FBLBP is longer than that of CSP-LBP and CSP, such time does not affect the use of CSP-FBLBP in a real-time BCI system. In addition, we can see that CSP-FBLBP has a greater time advantage than CSP-FB, SFBCSP, and SBLFB.

TABLE 1: The classification accuracy achieved by various feature extraction methods (data set 1).

Subject	CSP	CSP-FB	SFBCSP	SBLFB	CSP-LBP	CSP-FBLBP
L versus R	77.62	79.01	76.39	77.01	77.47	77.78
L versus F	82.41	85.88	79.32	80.25	82.33	87.58
L versus T	85.26	85.42	82.87	81.25	85.80	85.65
R versus F	81.02	84.72	76.93	76.62	84.10	86.96
R versus T	82.95	81.79	79.78	79.32	83.34	85.80
F versus T	72.61	72.61	71.30	70.68	74.46	77.55
Mean \pm Std	80.31 \pm 14.25	81.57 \pm 12.56	77.76 \pm 12.06	77.52 \pm 12.03	81.25 \pm 12.73	83.55 \pm 12.32

TABLE 2: The classification accuracy achieved by various feature extraction methods (data set 2).

Subject	CSP	CSP-FB	SFBCSP	SBLFB	CSP-LBP	CSP-FBLBP
B01	80.00	83.75	78.75	77.50	80.00	76.25
B02	60.00	55.00	50.00	42.50	56.25	61.25
B03	43.75	41.25	47.50	46.25	43.75	50.00
B04	97.50	97.50	97.50	98.75	92.50	97.50
B05	87.50	93.75	81.25	80.00	91.25	92.50
B06	81.25	76.25	73.75	75.00	80.00	78.75
B07	81.25	87.50	85.00	82.50	82.50	87.50
B08	92.50	92.50	80.00	76.25	91.25	95.00
B09	80.00	80.00	70.00	70.00	83.75	78.75
Mean \pm Std	78.19 \pm 16.61	78.61 \pm 18.86	73.75 \pm 16.12	72.08 \pm 17.62	77.92 \pm 16.84	79.72 \pm 15.86

TABLE 3: The classification accuracy achieved by various feature extraction methods (data set 3).

Subject	CSP	CSP-FB	SFBCSP	SBLFB	CSP-LBP	CSP-FBLBP
S01	70.31	68.75	51.56	50.00	67.19	76.56
S02	68.75	72.92	65.63	68.75	64.06	70.83
S03	68.75	84.62	64.58	62.50	72.92	78.46
S04	78.13	75.00	70.31	73.44	79.69	78.13
S05	75.00	78.13	43.75	45.31	73.44	76.56
S06	77.08	64.62	77.08	83.33	75.00	75.38
Mean \pm Std	73.00 \pm 4.25	74.01 \pm 7.04	62.15 \pm 12.31	63.89 \pm 14.37	72.05 \pm 5.61	75.99 \pm 2.77

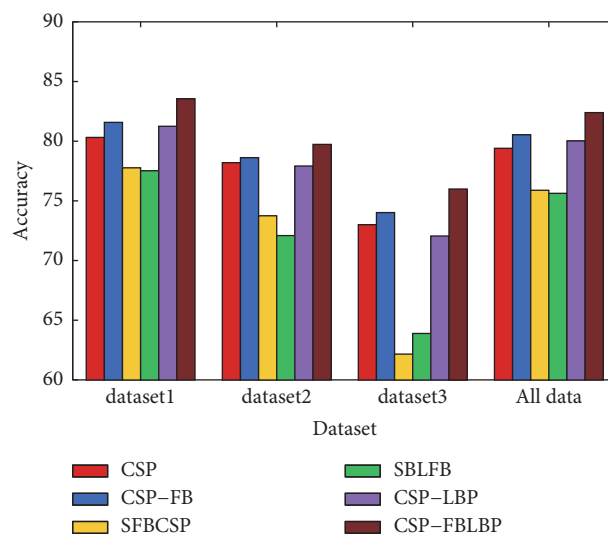


FIGURE 4: The average classification accuracy achieved by various feature extraction methods.

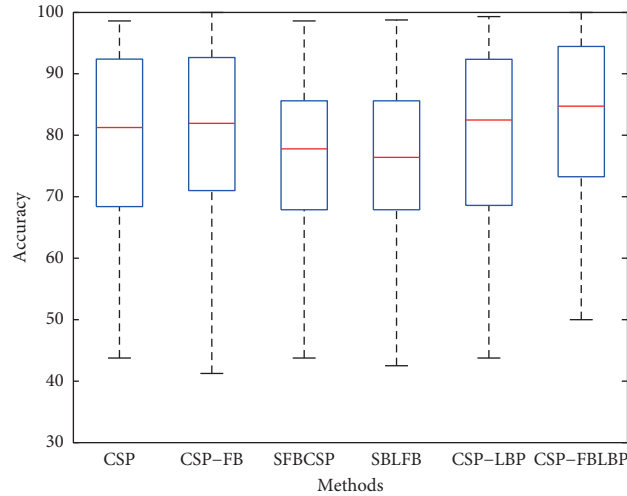


FIGURE 5: The distribution of classification accuracy achieved by various feature extraction methods.

TABLE 4: The feature extraction time of various methods (unit: second).

Subject	CSP	CSP-FB	SFBCSP (SBLFB)	CSP-LBP	CSP-FBLBP
A01	0.076	0.912 (0.041, 0.871)	55.380 (0.774, 54.606)	0.052	0.542 (0.040, 0.502)
B01	0.034	0.264 (0.025, 0.239)	11.769 (0.499, 11.270)	0.027	0.173 (0.024, 0.149)
S01	0.050	0.537 (0.026, 0.511)	54.765 (0.756, 54.009)	0.031	0.326 (0.027, 0.299)

3.2. Results for Feature Selection Methods

3.2.1. Comparison Methods and Parameter Settings. In order to verify the effectiveness of the proposed feature selection method, the proposed method was compared with four other feature selection methods, namely, LASSO [23], genetic algorithm (GA) [24], binary particle swarm optimization (BPSO) algorithm [25], and binary differential evolution (BDE) algorithm [26]. CSP-FBLBP is used for feature extraction, and SVM is used for classification.

LASSO: the alternative parameter set for LASSO is $\lambda \in 0.1 \times \{1, 2, \dots, 30\}$. The optimal parameter is selected by 10-fold cross-validation. LASSO with regression model is implemented by the SLEP toolbox [27]. After the LASSO model is determined, the features with a weight coefficient greater than 0 are selected as the optimal feature subset.

GA: the parameter setting of GA refers to literature [24]. The binary encoding is selected as the feature encoding method. The fitness function is the classification accuracy of the k -nearest neighbor classifier, where $k = 5$. The population size is 10, the number of iterations is used as the termination condition of the algorithm, and the maximum number of iterations is 100. The crossover probability is 0.8, and the mutation probability is 0.01.

BPSO: the implementation of BPSO refers to literature [25], and the parameter setting is consistent with literature [25]. The fitness function is the classification accuracy of the k -nearest neighbor classifier, where $k = 5$. The population size is 10, the number of iterations is used as the termination condition of the algorithm, and the maximum number of iterations is 100. The acceleration coefficients of the BPSO

are set as $c_1 = 2, c_2 = 2$. The maximum and minimum velocities are 6 and -6 , respectively. The maximum and minimum inertial weights are 0.9 and 0.4, respectively.

BDE: the parameter setting of BDE refers to literature [26]. The population size is 10, and the number of iterations is used as the termination condition of the algorithm, and the maximum number of iterations is set to 100. The crossover probability is 0.9.

3.2.2. Experimental Results. The average classification accuracy achieved by different feature selection methods is shown in Figure 6. The total average classification accuracy of LASSO, BPSO, GA, BDE, and F-score-h in all data is 77.57, 80.62, 80.56, 80.43, and 82.39, respectively. F-score-h is significantly better than other feature selection methods. BPSO, GA, and BDE are equally effective, and the effect of LASSO is relatively poor.

The distribution of classification accuracy achieved by various feature selection methods is shown in Figure 7. It can be seen that the median value of F-score-h is higher than that of other methods. The maximum value of F-score-h is 100%, and the minimum value of F-score-h is also higher than other methods. In addition, the overall classification accuracy distribution of F-score-h is relatively compact and close to the top. These results fully prove the superiority of F-score-h.

The running time of various feature selection methods is shown in Table 5. The training sets of the three subjects (A01, B01, and S01) are selected to calculate the feature selection time. The feature selection time of F-score-h is the

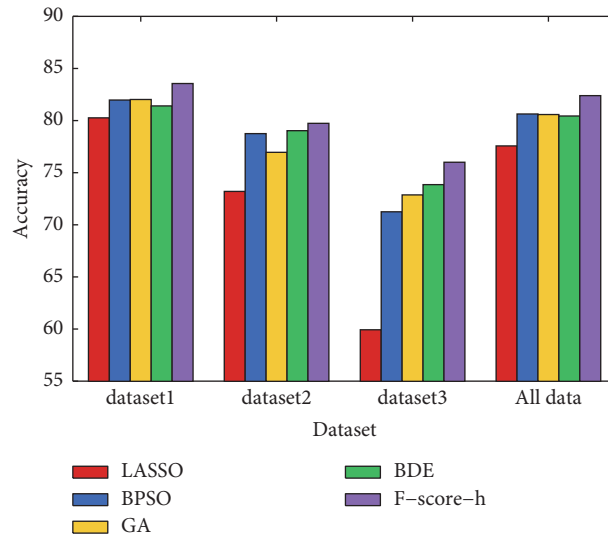


FIGURE 6: The average classification accuracy achieved by various feature selection methods.

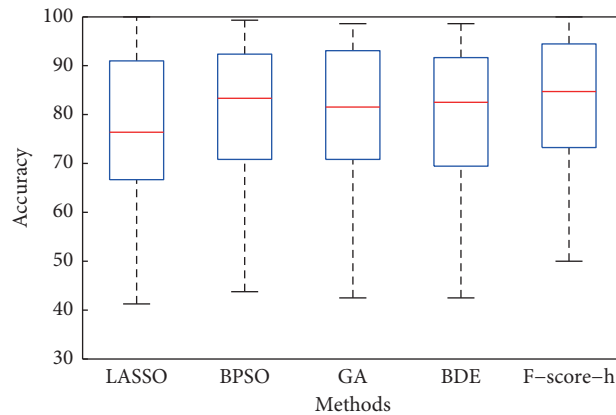


FIGURE 7: The distribution of classification accuracy achieved by various feature selection methods.

TABLE 5: The feature selection time of various feature selection methods (unit: second).

Subject	LASSO	BPSO	GA	BDE	F-score-h
A01	2.865	39.917	74.561	47.257	0.070
B01	1.879	40.337	77.432	45.693	0.050
S01	2.271	47.711	73.017	46.408	0.037

shortest, which is much lower than other methods. The feature selection time of BPSO, GA, and BDE methods is relatively long. In contrast, F-score-h has a huge time advantage.

In summary, F-score-h has great advantages in classification performance and feature selection time.

3.3. Results Compared with Other Existing Methods. In order to more fully reflect the advantages of the proposed method, the classification results of the proposed method are compared with that of the recently published papers. The classification results of data set 1 (L versus R binary classification task) are shown in Table 6, and the classification results of data set 2 are shown in Table 7. In data set 1, the proposed

method is superior to other existing methods. In data set 2, the proposed method is second only to the NCFS method [35] and is better than most existing methods. From the above experimental results, it can be seen that the classification effect of the proposed method has certain advantages.

4. Discussion

Comparing the classification results of CSP-FB and CSP as well as CSP-FBLBP and CSP-LBP can prove that selecting a subject-specific frequency band can improve the classification performance of CSP. The reason why CSP-FB and CSP-FBLBP obtain better classification results is mainly that these two methods use the filter bank method to make up for

TABLE 6: The classification accuracy achieved by various methods (data set 1: left-hand versus right-hand).

Subject	GRU-RNN [28] (2018)	IST-TSVM [29] (2019)	ICA + PSR + CSP [30] (2020)	p-LTCSP [31] (2020)	CSP-FBLBP
A01	84.82	80.14	80	82.6	92.36
A02	65.32	51.55	65.36	70.23	54.17
A03	83.54	95.54	87.14	70.23	93.75
A04	67.67	53.6	67.5	55.15	67.36
A05	64	51.65	55.54	54.36	64.58
A06	70.87	56.83	50.18	60.14	64.58
A07	84.96	56.58	91.79	73.38	77.78
A08	71.95	93.42	84.11	85.29	93.75
A09	68.9	92.66	87.86	74.62	91.67
Mean \pm Std	73.56 \pm 8.53	70.22 \pm 19.74	74.39 \pm 15.18	69.56 \pm 11.10	77.78 \pm 15.53

TABLE 7: The classification accuracy achieved by various methods (data set 2).

Subject	DBN [32] (2018)	CapsNet [33] (2019)	SGRM [34] (2019)	NCFS [35] (2020)	CSP-FBLBP
B01	70.38	78.75	77.3	79.25	76.25
B02	70.34	55.71	59.1	63.48	61.25
B03	71.2	55	51.5	56.65	50
B04	71.24	95.93	97	99.28	97.5
B05	71.21	83.12	87.4	88.67	92.5
B06	70.52	83.43	72.5	79.96	78.75
B07	70.79	75.62	86.7	88.76	87.5
B08	70.49	91.25	84.7	92.66	95
B09	70.32	87.18	85.6	84.95	78.75
Mean \pm Std	70.72 \pm 0.4	78.44 \pm 14.44	78.0 \pm 14.65	81.52 \pm 13.72	79.72 \pm 15.86

the frequency information of CSP. To further illustrate the problem, two subjects (A01 and B01) have been selected to show the spatial-frequency feature selection results, where features are extracted by CSP-FBLBP and selected by F-score-h. A total of six channels (B01 only has two channels) of signals are retained after CSP dimensionality reduction. The feature index 1–10 in Figure 8 corresponds to the features of the first channel signal filtered by 8–12 Hz, . . . , 26–30 Hz band-pass filter. The other feature indexes can be deduced by analogy. Figures 8(a) and 8(c) show the feature weights calculated by Fisher score; Figures 8(b) and 8(d) show the features selected by F-score-h. From the feature index, it is possible to calculate which channel and which frequency band the selected feature belongs to. It can be seen from Figure 8 that only a few features with high scores are retained. The channel (spatial information) and frequency band (frequency information) selected for different subjects are different; that is, the optimal spatial-frequency features are subject-specific. CSP-FBLBP jointly considers subject-specific spatial-frequency features, so a better classification result is achieved.

In addition, CSP-FBLBP and CSP-FB use the same method to compensate for the frequency defects of CSP, but the classification effect of CSP-FBLBP is better than CSP-FB. This result shows that the feature type extracted after CSP spatial filtering is also very critical. It can be seen from the experimental results that the logarithmic band power is better than the traditional logarithmic variance features. Based on CSP transformation, it is worth studying to further improve the feature extraction method.

Compared with the existing spatial-frequency feature extraction methods (CSP-FB, SFBCSP, and SBLCSF), CSP-FBLBP has a greater time advantage, and the feature extraction time is significantly lower than the existing methods. The feature extraction time of SFBCSP and SBLFB is the longest, mainly due to the time of band-pass filtering. SFBCSP and SBLFB decompose the original EEG signals into 17 frequency subbands. The number of subbands and channels is relatively large, and the amount of calculation is relatively large, so the feature extraction time is long. After CSP dimensionality reduction, the number of signal channels of CSP-FBLBP is greatly reduced, so its feature extraction time is significantly reduced. In addition, comparing the feature extraction time of CSP-LBP and CSP as well as CSP-FBLBP and CSP-FB can show that the calculation time and complexity of logarithmic band power features are lower than a logarithmic variance.

From the comparative analysis of the above experimental results, it can be concluded that the F-score-h feature selection method has achieved better classification results, and its feature selection time also has significant advantages. F-score-h is a hybrid of filter and wrapper feature selection methods. On the one hand, the filter feature selection method has a small amount of calculation, so the calculation time is short; on the other hand, the wrapper feature selection method uses the classification performance of the classifier as an evaluation standard, its classification performance is generally better. F-score-h takes into account the advantages of both filter and wrapper methods, so it has achieved better classification performance.

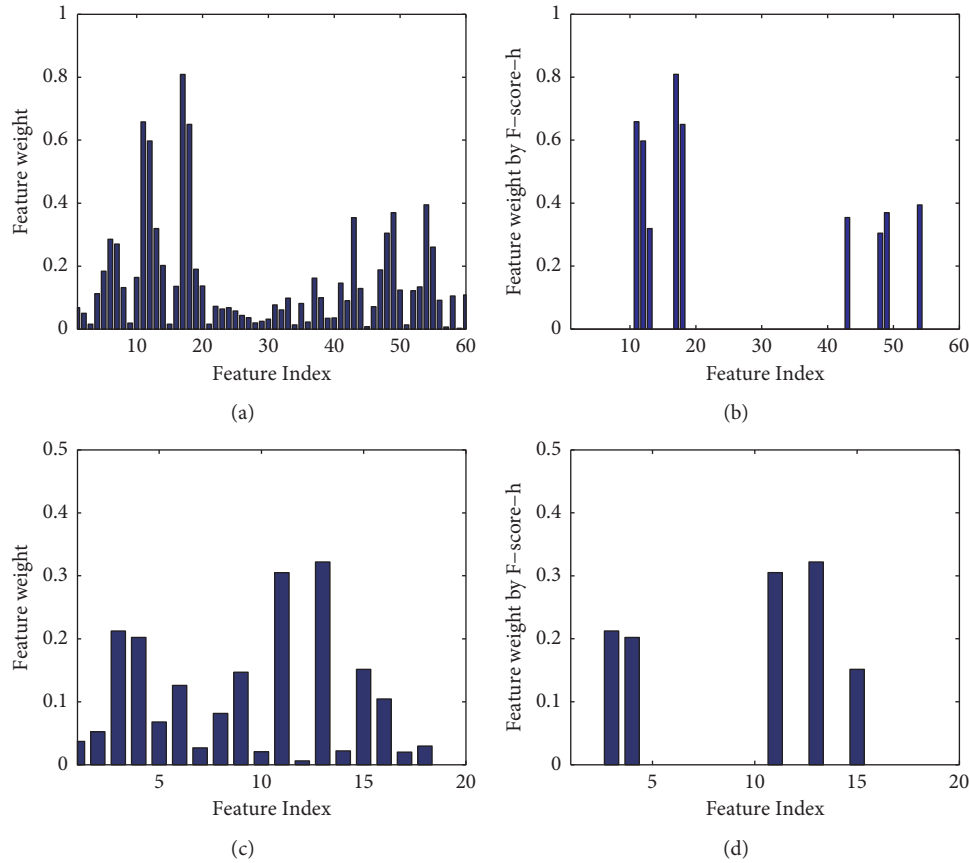


FIGURE 8: The feature selection by F-score-h.

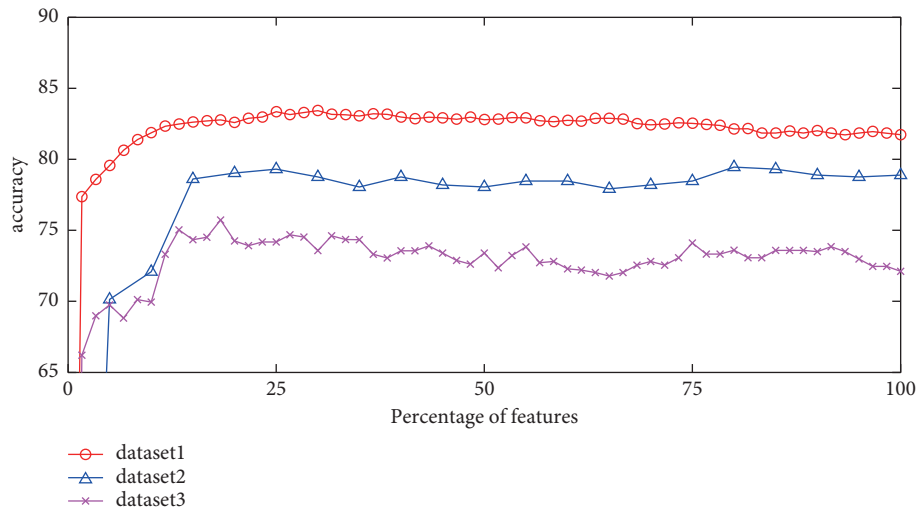


FIGURE 9: The variation of the average accuracy with the number of features (F-score).

In this paper, for the LASSO method, we select the features whose weight is greater than 0 as the optimal feature subset. Generally, the larger the feature weight, the more important the corresponding feature. However, the optimal feature subset is selected by the LASSO model, which is not necessarily the best on the SVM classifier [36]. Combined with LASSO and SVM for feature selection, the classification effect may be better [37]. The classification effect of BPSO,

GA, and BDE is relatively poor; there are many reasons. First of all, the genetic algorithm may fall into a local optimal situation, BPSO may appear a “premature” phenomenon [25], and BDE may not be able to effectively converge. In addition, the selection of initialization parameters for BPSO, GA, and BDE also has a great influence on feature selection. How to choose more suitable model parameters is a very critical issue.

TABLE 8: The classification accuracy achieved by F-score and F-score-h.

Subject	F-score				F-score-h
	25%	50%	75%	100%	
Dataset 1	83.35 ± 12.23	82.78 ± 12.65	82.52 ± 12.42	81.74 ± 13.17	83.55 ± 12.32
Dataset 2	79.31 ± 15.30	78.06 ± 17.35	78.47 ± 16.70	78.89 ± 16.30	79.72 ± 14.95
Dataset 3	74.17 ± 2.73	73.40 ± 5.58	74.11 ± 7.07	72.12 ± 5.66	75.99 ± 2.53
Dataset all	82.02 ± 12.49	81.35 ± 13.25	81.26 ± 12.97	80.53 ± 13.45	82.39 ± 12.41

To further compare the advantages of F-score-h, we compared F-score-h with F-score. F-score is a filter method that uses Fisher scores to rank the features and then select the top K features for classification. The classification accuracy when a different number of features is selected is shown in Figure 9. The feature dimension extracted in data set 2 is only 20, while the feature dimension of data sets 1 and 3 is 60; in order to compare the variation of the average classification accuracy of the three data sets with the number of features on the same axis, the number of features is mapped to a percentage. For example, 25 on the abscissa of Figure 9 indicates that the selected feature number is 25% of the total number of features. It can be seen from Figure 9 that the number of features corresponding to the optimal classification accuracy of each data set is different. The filtering method selects the same number of features for all data sets, and its classification effect is not good. F-score-h selects features through feature weights, so the selected feature subset is more discriminative and contains less redundant information. In addition, F-score-h can select subject-specific features. Therefore, F-score-h is better than F-score.

The average classification accuracy of each data set when the F-score takes a different number of features is shown in Table 8. It can be intuitively observed from Table 8 that the accuracy of F-score-h is optimal in each data set.

5. Conclusion

The new feature extraction and selection methods have been proposed in this paper. In the new feature extraction method, the logarithmic band power is used to replace the logarithmic variance, and the filter bank method is used to compensate for the frequency defects of CSP. In the new feature selection method, the Fisher score is used to sort the features, and then a series of threshold parameters are set; SVM combined with cross-validation is used to select the optimal threshold parameters so as to obtain the optimal feature subset. The experimental results show that the proposed feature extraction and feature selection method has better classification performance than existing methods, and both feature extraction time and feature selection time have greater advantages, which can be applied to real-time BCI systems.

Although the proposed method has achieved good classification results, the impact of the time window on CSP is not considered in the feature extraction process, and the proposed method still has a large room for improvement. In future work, we will jointly consider efficient time-spatial-frequency feature extraction and selection methods.

Data Availability

In this study, we used three data sets for experiments. Data sets 1 and 2 are public BCI competition data sets, which have been deposited on the BCI competition website. Data set 3 is self-collected by our laboratory and is not publicly available but can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest

All authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61967004, 11901137, and 81960324), Natural Science Foundation of Guangxi Province (no. 2018GXNSFB A281023), and Guangxi Key Laboratory of Automatic Testing Technology and Instruments (nos. YQ20113 and YQ19209).

References

- [1] A. Khalaf and M. Akcakaya, "A probabilistic approach for calibration time reduction in hybrid EEG-fTCD brain-computer interfaces," *BioMedical Engineering Online*, vol. 19, no. 1, pp. 1-18, 2020.
- [2] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, 2019.
- [3] N. Gursel Ozmen, L. Gumusel, and Y. Yang, "A biologically inspired approach to frequency domain feature extraction for EEG classification," *Computational and Mathematical Methods in Medicine*, vol. 2018, 2018.
- [4] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. 45, no. 6, pp. 2137-2144, 2012.
- [5] B. Wang, C. M. Wong, Z. Kang et al., "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Transactions on Cybernetics*, vol. 51, no. 10, pp. 5008-5020, 2020.
- [6] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-r. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41-56, 2008.
- [7] M. Aljalal, R. Djemal, K. AlSharabi, and S. Ibrahim, "Feature extraction of EEG based motor imagery using CSP based on logarithmic band power, entropy and energy," in *Proceedings of the 2018 1st International Conference on Computer*

- Applications & Information Security (ICCAIS)*, pp. 1–6, IEEE, Riyadh, Saudi Arabia, April 2018.
- [8] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, “Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b,” *Frontiers in Neuroscience*, vol. 6, p. 39, 2012.
 - [9] Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, “Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface,” *Journal of Neuroscience Methods*, vol. 255, pp. 85–91, 2015.
 - [10] Y. Zhang, Y. Wang, J. Jin, and X. Wang, “Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification,” *International Journal of Neural Systems*, vol. 27, no. 02, Article ID 1650032, 2017.
 - [11] S. Zhang, Z. Zhu, B. Zhang, B. Feng, T. Yu, and Z. Li, “The CSP-based new features plus non-convex log sparse feature selection for motor imagery EEG classification,” *Sensors*, vol. 20, no. 17, p. 4749, 2020.
 - [12] P. Moradi and M. Gholampour, “A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy,” *Applied Soft Computing*, vol. 43, pp. 117–130, 2016.
 - [13] I. Jain, V. K. Jain, and R. Jain, “Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification,” *Applied Soft Computing*, vol. 62, pp. 203–215, 2018.
 - [14] H. Lu, J. Chen, K. Yan, Q. Jin, Y. Xue, and Z. Gao, “A hybrid feature selection algorithm for gene expression data classification,” *Neurocomputing*, vol. 256, pp. 56–62, 2017.
 - [15] A. S. Ghareb, A. A. Bakar, and A. R. Hamdan, “Hybrid feature selection based on enhanced genetic algorithm for text categorization,” *Expert Systems with Applications*, vol. 49, pp. 31–47, 2016.
 - [16] N. Almgren and H. Alshamlan, “A survey on hybrid feature selection methods in microarray gene expression data for cancer classification,” *IEEE Access*, vol. 7, pp. 78533–78548, 2019.
 - [17] C. Zuo, J. Jin, R. Xu et al., “Cluster decomposing and multi-objective optimization based-ensemble learning framework for motor imagery-based brain-computer interfaces,” *Journal of Neural Engineering*, vol. 18, no. 2, Article ID 26018, 2021.
 - [18] C. Zuo, Y. Miao, X. Wang, L. Wu, and J. Jin, “Temporal frequency joint sparse optimization and fuzzy fusion for motor imagery-based brain-computer interfaces,” *Journal of Neuroscience Methods*, vol. 340, p. 108725, 2020.
 - [19] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2010.
 - [20] A. Shoeibi, N. Ghassemi, R. Alizadehsani et al., “A comprehensive comparison of handcrafted features and convolutional autoencoders for epileptic seizures detection in EEG signals,” *Expert Systems with Applications*, vol. 163, Article ID 113788, 2021.
 - [21] M. Radman, M. Moradi, A. Chaibakhsh, M. Kordestani, and M. Saif, “Multi-feature fusion approach for epileptic seizure detection from EEG signals[],” *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3533–3543, 2020.
 - [22] C.-C. Chang and C.-J. Lin, “Libsvm,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
 - [23] A. Mezirani, K. Djouani, T. Medkour, and A. Chibani, “A Lasso quantile periodogram based feature extraction for EEG-based motor imagery,” *Journal of Neuroscience Methods*, vol. 328, Article ID 108434, 2019.
 - [24] J. Too, A. Abdullah, N. Mohd Saad, and W. Tee, “EMG feature selection and classification using a Pbest-guide binary particle swarm optimization,” *Computation*, vol. 7, no. 1, p. 12, 2019.
 - [25] J. Too, A. R. Abdullah, and N. Mohd Saad, “A new Co-evolution binary particle swarm optimization with multiple inertia weight strategy for feature selection,” *Informatics*, vol. 6, no. 2, p. 21, 2019.
 - [26] D. Datta and S. Dutta, “A binary-real-coded differential evolution for unit commitment problem,” *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 517–524, 2012.
 - [27] J. Liu, S. Ji, and J. Ye, “SLEP: sparse learning with efficient projections,” *Arizona State University*, vol. 6, no. 491, p. 7, 2009.
 - [28] E. C. Djamel and R. D. Putra, “Brain-computer interface of focus and motor imagery using wavelet and recurrent neural networks[],” *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 4, pp. 2748–2756, 2020.
 - [29] Y. Xu, J. Hua, H. Zhang et al., “Improved transductive support vector machine for a small labelled set in motor imagery-based brain-computer interface[],” *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 2087132, 2019.
 - [30] E. Dong, K. Zhou, J. Tong, and S. Du, “A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification,” *Biomedical Signal Processing and Control*, vol. 60, Article ID 101991, 2020.
 - [31] Z. Yu, T. Ma, N. Fang, H. Wang, Z. Li, and H. Fan, “Local temporal common spatial patterns modulated with phase locking value,” *Biomedical Signal Processing and Control*, vol. 59, Article ID 101882, 2020.
 - [32] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, and Y. Zhao, “A decoding scheme for incomplete motor imagery EEG with deep belief network,” *Frontiers in Neuroscience*, vol. 12, p. 680, 2018.
 - [33] K.-W. Ha and J.-W. Jeong, “Motor imagery EEG classification using capsule networks,” *Sensors*, vol. 19, no. 13, p. 2854, 2019.
 - [34] Y. Jiao, Y. Zhang, X. Chen et al., “Sparse group representation model for motor imagery EEG classification[],” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 631–641, 2018.
 - [35] M. K. I. Molla, A. A. Shiam, M. R. Islam, and T. Tanaka, “Discriminative feature selection-based motor imagery classification using EEG signal,” *IEEE Access*, vol. 8, pp. 98255–98265, 2020.
 - [36] J. Li, K. Cheng, S. Wang et al., “Feature selection: a data perspective,” *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2017.
 - [37] M. Miao, H. Zeng, A. Wang, C. Zhao, and F. Liu, “Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: an sparse regression and Weighted Naïve Bayesian Classifier-based approach,” *Journal of Neuroscience Methods*, vol. 278, pp. 13–24, 2017.