Computer Methods and Programs in Biomedicine 193 (2020) 105464

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Motor imagery EEG classification based on ensemble support vector learning

Jing Luo[a,*], Xing Gao[a], Xiaobei Zhu[a], Bin Wang[a], Na Lu[b], Jie Wang[b]

[a] Shaanxi Key Laboratory for Network Computing and Security Technology, School of Computer Science and Engineering, Xi'an University of Technology, Xi'an Shaanxi, China

[b] State Key Laboratory for Manufacturing Systems Engineering, Systems Engineering Institute, Xi'an Jiaotong University, Xi'an, Shaanxi, China

ARTICLE INFO

Article history:
Received 9 October 2019
Revised 27 February 2020
Accepted 18 March 2020

Keywords:
Brain-computer interface
Motor imagery
Common spatial pattern
Support vector machine

ABSTRACT

Background and Objective: Brain-computer interfaces build a communication pathway from the human brain to a computer. Motor imagery-based electroencephalogram (EEG) classification is a widely applied paradigm in brain-computer interfaces. The common spatial pattern, based on the event-related desynchronization (ERD)/event-related synchronization (ERS) phenomenon, is one of the most popular algorithms for motor imagery-based EEG classification. Moreover, the spatiotemporal discrepancy feature based on the event-related potential phenomenon has been demonstrated to provide complementary information to ERD/ERS-based features. In this paper, aiming to improve the performance of motor imagery-based EEG classification in a few-channel situation, an ensemble support vector learning (ESVL)-based approach is proposed to combine the advantages of the ERD/ERS-based features and the event-related potential-based features in motor imagery-based EEG classification.

Methods: ESVL is an ensemble learning algorithm based on support vector machine classifier. Specifically, the decision boundary with the largest interclass margin is obtained using the support vector machine algorithm, and the distances between sample points and the decision boundary are mapped to posterior probabilities. The probabilities obtained from different support vector machine classifiers are combined to make prediction. Thus, ESVL leverages the advantages of multiple trained support vector machine classifiers and makes a better prediction based on the posterior probabilities. The class discrepancy-guided sub-band-based common spatial pattern and the spatiotemporal discrepancy feature are applied to extract discriminative features, and then, the extracted features are used to train the ESVL classifier and make predictions.

Results: The BCI Competition IV datasets 2a and 2b are employed to evaluate the performance of the proposed ESVL algorithm. Experimental comparisons with the state-of-the-art methods are performed, and the proposed ESVL-based approach achieves an average max kappa value of 0.60 and 0.71 on BCI Competition IV datasets 2a and 2b respectively. The results show that the proposed ESVL-based approach improves the performance of motor imagery-based brain-computer interfaces.

Conclusion: The proposed ESVL classifier could use the posterior probabilities to realize ensemble learning and the ESVL-based motor imagery classification approach takes advantage of the merits of ERD/ERS based feature and event-related potential based feature to improve the experimental performance.

1. Introduction

Brain-computer interface is a technology that establishes communication or control pathways between the human brain and external devices; in this way, brain activity can be used to control external devices [1–3]. BCIs have been successfully used in medicine, neurobiology and psychology and possess promising commercial prospects in the fields of virtual reality, educational technology, and smart homes [4,5]. The motor imagery (MI) electroencephalogram (EEG) is an endogenous, spontaneous EEG. It is simple, flexible, and noninvasive, and its advantages include low environmental requirements, good temporal resolution and low-cost equipment. Thus, MI EEG-based BCIs are one of the most widely ap-

* Corresponding author.
E-mail address: luojing@xaut.edu.cn (J. Luo).

https://doi.org/10.1016/j.cmpb.2020.105464
0169-2607/

plied paradigms [6,7]. The MI EEG paradigm collects the EEG signal when the subject imagines a specific motion, recognizes the imagined task according to the EEG signal, and then converts the recognition result into a control command to realize the control of a peripheral device. Recently, MI-based EEG classification-related research has mainly focused on three aspects, namely, the equipment and technology for collecting EEG signals, the method of feature extraction and the training of classifiers [8].

During the execution of MI, the frequency band power of the EEG signals changes according to the content of the imagined task, and these phenomena are called event-related desynchronization (ERD) and event-related synchronizationand (ERS) [9]. Features based on ERD/ERS are efficient in the classification of MI. The common spatial pattern (CSP) and its variants take advantage of the ERD/ERS phenomenon by maximizing or minimizing the variance ratio in the EEG signals from the two classes [10], and the features extracted based on the CSP algorithm are discriminative in two-class MI-based EEG classifications [11]. In recent years, CPS-based features have achieved great success in MI-based BCI systems [12]. Because the CSP algorithm is highly dependent on the covariance matrix of the EEG signals, the performance is greatly affected by the sample numbers. The regularized common spatial pattern was proposed to overcome performance degradation in a small-sample setting, and parameters are involved in the trade-off between the variance and the bias of the covariance matrix estimation [13]. Ang introduced the filter band CSP (FBCSP) algorithm that won the BCI Competition IV on datasets 2a and 2b. The CSP-based features are extracted from 9 subbands to efficiently capture the ERD/ERS phenomenon [14]. Kumar proposed a CSP-based algorithm that applied tangent space mapping (TSM) in the manifold of covariance matrixes for MI-based BCI [15]. The discriminative frequency bands vary from subject to subject. Therefore, many researchers try to recognize the most effective frequency bands and utilize the EEG signals in these bands to improve the performance of the CSP algorithm [14,16–18]. Thomas proposed the discriminative CSP method (DCSP). Fisher's ratio was imposed to select the most discriminative frequency bands, and the specific selected frequency bands of each subject are leveraged to extract features [16]. Two mutual information based alogrightms (The mutual information-based best individual feature algorithm and mutual information-based rough set reduction algorithms) were used to find the most discriminative features in the FBCSP [14]. Kumar proposed an improved discriminative FBCSP (IDCSP) method for MI-based EEG classification. The mutual information is calculated based on the extracted feature of different frequency bands, and the best bandpass filters are selected and applied for classification [19]. Zhang proposed a sparse Bayesian learning algorithm to obtain sparse EEG frequency band-based feature matrix. The features with large weight are employed to accomplish the classification with a linear discriminant criterion [18]. Yang proposed a Fisher wavelet packet decomposition (WPD)-CSP algorithm that found the most helpful subbands for a specific subject by the Fisher distance, and the WPD-based features and CSP-based features are combined and fed to the probabilistic neural network classifier [20].

The traditional approaches improve the classification results by selecting the discriminative features or frequency bands and abandoning the rest. In this way, the searching space of the classifier is shrunk, and the difficulty of the machine learning task is reduced. As a result, a better solution is easily obtained. However, feature selection procedures do not improve the separability of the extracted features. Second, the stability of the CSP algorithm decreases in few-channel EEG system. To overcome these weaknesses, the class discrepancy-guided subband filter-based common spatial pattern (CDFCSP) was proposed [21]. CDFCSP could make use of the a priori knowledge from the training set to augment the EEG signals in discriminative frequency bands, and extend the number of the EEG channel before feature extraction.

The event-related potential (ERP) is an event-related brain activity that could cause time-locked fluctuation in the EEG signals, and the possible event includes sensory stimuli, imagery and so on [9,22]. The EEG signals amplitudes from the sensorimotor area of the brain will appear time-lock incresing or decreasing depending on the imagined task [22]. Recently, ERP-based MI classification studies have been proposed [23,24]. Some researchers have proved that the movement-related potential and the ERD provide complementary information when volitional finger movements are prepared and executed [25]. In our previous study of the two-class MI-based EEG classification problem, the spatiotemporal discrepancy feature (STDF) was proposed to take advantage of the ERP. Compared with the ERD/ERS-based features, the STDF performed well in different subjects. Therefore, combining the advantages of ERD/ERS-based features and ERP-based features becomes a key problem to further improve the performance of MI.

As the support vector machine (SVM) algorithms are the most popular method for MI classification [26,27], aiming to improve the performance of two-category MI-based EEG classification in a few-channel situation, an ensemble support vector learning (ESVL)-based approach is proposed to combine the merits of the ERD/ERS-based features and the ERP-based features in MI-based EEG classification. The decision boundary with the largest interclass margins is obtained using the SVM algorithm, and the distances between sample points and the decision boundary are mapped to posterior probabilities [28]. The probabilities obtained from different SVM classifiers are combined to predict the ESVL classifier. Specifically, the proposed method is composed of three steps. First, the a priori knowledge is extracted from the EEG in the training set to design CDF. The EEG signals augmented by the CDF is applied to extract features through the CSP algorithm. Second, the difference between the EEG signal from left and right sensorimotor area of the brain is evaluated to obtain the STDF. Finally, the extracted CDFCSP-based features and the STDF are used to train the ESVL classifier and make predictions. The experiments performance of the proposed ESVL based algorithm is evaluated based on the BCI Competition IV datasets 2a and 2b.

This paper is organized as follows. Section 2 describes the proposed ESVL-based approach. Section 3 presents BCI Competition IV datasets 2a and 2b as well as the experiments. Section 4 discusses the experiments results. Section 5 draws the conclusion.

2. Methods

Our proposed method extracts CDFCSP-based features and the STDF separately. The features are fed to the proposed ESVL classifier to make predictions. The Fig. 1 gives the block diagram of the proposed ESVL-based algorithm. The training stage is presented on the upper block and the testing stage is presented on the lower block.

First, the discrepancy signal of the EEG signals is calculated, and the STDF is extracted using an integral method. Second, the energy spectrum of the EEG trials in the training set is extracted, and the CDF is designed according to the energy spectrum. The segmented training EEG signals are filtered by the CDF to augment the discriminative frequency bands. And then, the FBCSP [14] algorithm (without feature selection procedure) is applied to extract the CDFCSP-based feature based on the filtered EEG signals. Finally, the ESVL classifier is trained based on the CDFCSP-based features and the STDF. The final prediction of the testing trial is made by the trained ESVL classifier. The algorithm details are described in the following.

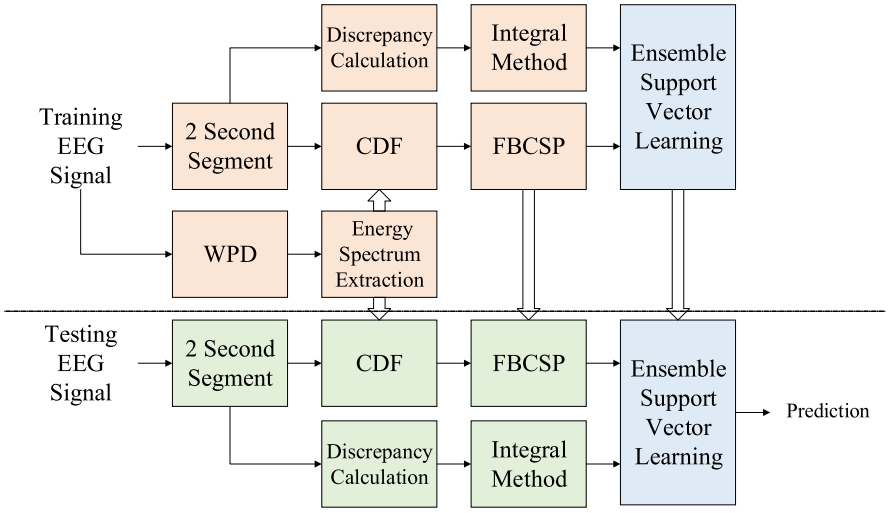


Fig. 1. Block diagram of the proposed ESVL method.

2.1. CDFCSP-Based feature extraction

2.1.1. CDF-Based preprocessing

Usually, a feature selection procedure is conducted following CSP-based feature extraction to find the most discriminative frequency bands and features [14]. Instead of enhancing the discriminant ability of the feature, feature selection reduces only the searching space of the machine learning algorithm. In contrast, the CDFCSP augments the EEG signals into discriminative frequency bands and enlarges the number of EEG signal channels reasonably, thus improving the effectiveness of the original CSP algorithm [21]. The designing guideline of the CDF is obtained by calculating the energy spectrum templates according to the EEG signals from the training set. WPD can accurately transform the signal into a time-frequency representation and realize denoising at the same time [29,30]. Therefore, WPD is applied to produce the energy spectrum templates. The WPD coefficients in level j are denoted by $WP^m(i, k)$, where m, i and k are the index of EEG signal channel, frequency component and coefficient in the frequency component respectively. Then, the energy spectrum of the EEG signals is calculated by the quadratic sum of the WPD coefficients:

$$s(m,i)=\sum_{k=1}^{K}(WP^m(i,k))^2 \tag{1}$$

where K is the maximum number of k in frequency component i. The energy spectrum templates are calculated as the average energy spectra in related class:

$$T_c(m,i)=\frac{1}{N}\sum_{\text{trials}\in C}s(m,i) \tag{2}$$

平均能量谱

where c and C are the class index and the training set of class c, respectively. And N represents the EEG trials number belonging to class c in the training set. Then, the energy spectrum template of each class is a matrix for which the number of rows equals the number of channels and the number of columns equals the frequency components number. The discriminative frequency bands relate to the frequency bands with significant template differences between classes. The energy spectrum templates are employed to guide the CDF designing to enhance the discriminant frequency components before feature extraction. Normalization mode CDF matrixes $H(m, i)$ are obtained based on the energy spectrum template of different classes as follows:

$$H_1(m,i)=\frac{T_{\text{class1}}(m,i)}{T_{\text{class1}}(m,i)+T_{\text{class2}}(m,i)} \tag{3}$$

$$H_2(m,i)=\frac{T_{\text{class2}}(m,i)}{T_{\text{class1}}(m,i)+T_{\text{class2}}(m,i)} \tag{4}$$

In this way, each element in H depends on the ratio of two templates. If the frequency band is indiscriminative, the related elements in the template of class 1 equal the elements in the template of class 2, and the related elements in H approximately equal 0.5. In contrast, if the frequency band is obviously discriminative, at least one related element in H would obviously be larger than the elements in the indiscriminative frequency bands.

WPD is applied to get the WPD coefficients $WP^m(i, k)$, and a CDF is applied by changing the WPD coefficients of the raw EEG signals. Let the WPD coefficients in the discriminative components of the raw EEG signals be enlarged and the others be shrunk according to $H(m, i)$ as follows:

$$WP^m_{out}(i,k)=H(m,i)\cdot WP^m_{in}(i,k) \tag{5}$$

where $WP^m_{in}(i,k)$ represents the WPD coefficients of the raw EEG signals and $WP^m_{out}(i,k)$ denotes the filtered WPD coefficients. Two CDF matrixes from Eqs. (3) and (4) are applied to filter the raw EEG signals simultaneously. Then, the WPD coefficients will be enlarged or shrunk according to a priori knowledge of the classes in the training set, i.e., the differences in templates between classes. To reconstruct the filtered EEG signals, inverse WPD is employed based on the modified WPD coefficients $WP_{out}(i, k)$. Because two CDF matrixes obtained in Eqs. (3) and (4) are leveraged, the number of channels of the EEG signals are doubled after filtering.

2.1.2. CSP-Based feature extraction

The CSP algorithm finds filters that maximize the variance in the EEG signals for one class and minimize the variance for another class [12]. Recently, CSP-based feature extraction from different frequency bands has been shown to be efficient and is widely used in MI-based EEG classification [12,14,18]. Thus, the FBCSP algorithm is applied to extract features.

First, the EEG signals filtered by the CDF are leveraged to decompose into several subband bandpass filters. Chebyshev II bandpass filter bank is applied as in the FBCSP algorithm [14]. The filter bank has 9 bandpass filters including 4–8 Hz, 8–12 Hz, 12–16 Hz,

..., and 36–40 Hz. Then, the CSP algorithm is performed to find the projection matrix P of size $2e \times M$, where e is the hyperparameter controlling the channel of the projected EEG signals and M denotes the total channel number of the EEG signals. The projected EEG signals Z_t of size $2e \times L$ is calculated through P as follows:

$$Z_t = P \times E_t \tag{6}$$

投影EEG信号

where E_t is an EEG signals of size $M \times L$ and L is the sampling point of an EEG segment. Finally, the variance in the projected EEG signals is normalized, and the logarithm is calculated as follows:

$$f_{t,j} = \log\left(\frac{\mathrm{var}(Z_{t,j})}{\sum_{i=1}^{2e}\mathrm{var}(Z_{t,i})}\right), t = 1, 2; j = 1, 2, \cdots, 2e \tag{7}$$

f方差被归一化

Because a one-dimensional feature could be extracted from one channel of the projected EEG signals and 9 subband filters are employed, the dimension of the extracted feature is $18e$.

2.2. STDF Extraction

To capture the ERP from the EEG signals, the STDF is employed [31]. First, the spatiotemporal discrepancy signal is calculated based on the EEG signals collected from different area of the brain:

$$D = S_l - S_r \tag{8}$$

where S_l is the EEG signals collected from the left sensorimotor area (such as the C3 channel from the International 10–20 system) and S_r is the EEG signals collected from the right sensorimotor area (such as the C4 channel from the International 10–20 system). The spatiotemporal discrepancy signal describes the discrepancy in the EEG signals over both space and time. In this way, the contralateral difference would be enhanced, and the noise influencing both side of the sensorimotor area would be suppressed. The STDF is calculated by summation as follows:

$$\mathrm{STDF}(f) = \sum_{x=1}^{\frac{f\cdot X}{10}} D(x), f = 1, 2, 3, \ldots, 10 \tag{9}$$

where f and x are the indexes of the features and sampling points, respectively, and the maximum x in an EEG trial is denoted as X. The dimension of the feature obtained is a hyperparameter that is set based on a previous work [31].

2.3. Ensemble support vector learning classifier

As mentioned above, the ERD/ERS-based features show satisfactory discrimination on some subjects, and the ERP-based features are discriminative on other subjects. The CDFCSP-based features and the STDFs are applied to train linear SVM classifiers seperately to validate the performance on different subjects. The details of the experiment can be found in Section 4.1.

To incorporate the merits of both features, ESVL is proposed. SVMs are the most popular method for MI classification [26,32], so the proposed classification frame is built based on the SVM classifier. Because an increasing number of probabilistic-based algorithms have been proposed to improve the classification performance [18,33,34], the posterior probability is leveraged in ESVL to accomplish ensemble learning. Fig. 2 illustrates the proposed ESVL classifier. First, the linear SVM classifiers are separately trained on the CDFCSP-based feature, the STDF and the concatenated features, and the output posterior probability of the trained three SVM classifiers are combined to form the final prediction.

The SVM classifier attempts to find an optimal hyperplane that divides the data from two classes as:

$$f(x) = x'\beta + b = 0 \tag{10}$$

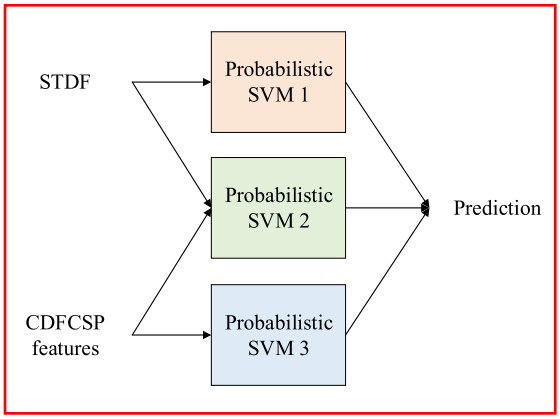


Fig. 2. Illustration of the ESVL classifier.

where x' is the transpose of input features and β and b are parameters of the hyperplane. This optimal hyperplane should maximize the margin between the data from the positive class and the negative class. For some inseparable data, a penalty is imposed on the objective function for the sample located on the inverse side of the class boundary as follows:

$$\begin{aligned} \beta, b = \arg\min_{\beta,b} \frac{1}{2}\|\beta\|^2 + \frac{1}{2}\|b\|^2 + g\sum \xi_j \\ \text{s.t.} \\ y_i f(x_j) \geq 1 - \xi_j \\ \xi_j \geq 0 \end{aligned} \tag{11}$$

where ξ_j represents the slack variables used to penalize the samples crossing the margin boundary and g is the penalty multiplier to control the maximum penalty imposed on margin-violating samples. The Lagrange multiplier method is applied to optimize the objective function and find the optimal hyperplane. After the optimal hyperplane is found based on the training set, the SVM is well trained.

Then, the posterior probability is calculated. The signed distance from the sample point to the decision boundary is calculated as follows:

$$d(x) = x'\beta + b \tag{12}$$

If the distance is a positive value, the sample belongs to a positive class. In contrast, a negative distance means that the sample belongs to a negative class. The posterior probability, which is the probability that a sample belongs in a specific class, is calculated as a sigmoid function of the distance [28] as follows:

$$\rho_j = \frac{1}{1 + \exp(Ad_j + B)} \tag{13}$$

预测概率

where d_j is the distance calculated by (11) and A and B are the parameters representing the slope and the intercept. To fit this sigmoid function, the label $y_j \in (-1, 1)$ is transferred into the probability target, which is defined as the probability of the positive class as follows:

$$t_j = \begin{cases} 0 & y_i = -1 \\ 1 & y_i = 1 \end{cases} \tag{14}$$

Then, the cross-entropy error function is set as the objective function to find the optimal parameters A and B as follows:

$$A, B = \arg\min_{A,B} -\sum_j t_j \log(\rho_j) + (1 - t_j)\log(1 - \rho_j) \tag{15}$$

Then, the optimal parameters A and B are obtained using maximum likelihood estimation based on the training set. As shown in Fig. 2, two probabilistic SVM classifiers are trained based on

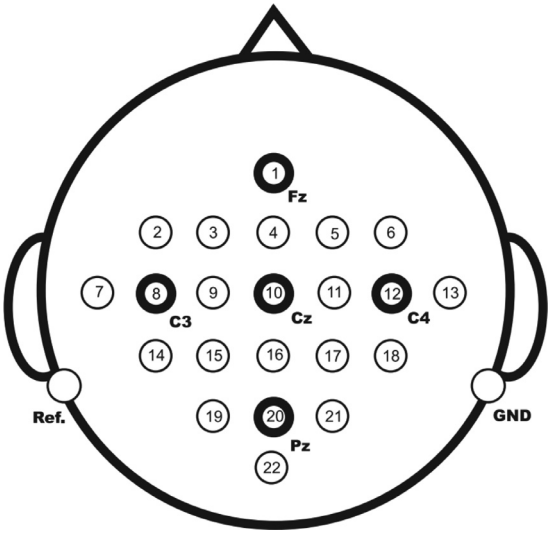


Fig. 3. The locations of the electrodes in the International 10–20 system.

the CDFCSP based feature and the STDF separately, and the CDFCSP based feature and STDF are concatenated to train the third probabilistic SVM classifier. Thus, the posterior probability of the testing sample belonging to the positive class can be computed by Eq. (12) and Eq. (13).

Based on the three trained probabilistic SVM classifiers, the posterior probabilities of the test trial belonging to the positive class are calculated. Finally, the three posterior probabilities are summed up to obtain the probability output and make the prediction of the ESVL classifier. If the final probability belonging to the positive class is larger than the probability of the negative class, the prediction is positive, and vice versa.

3. Data and experiment

We used BCI Competition IV datasets 2a and 2b in our experiments.

3.1. BCI Competition IV dataset 2a

This dataset collected MI-based EEG signals from 9 subjects [35,36]. For each subject, one training session including 288 trials (72 trials for each class) and one evaluation session including 288 trials (72 trials for each class) were recorded on different days. Based on the cue displayed on the screen, four different kinds of MI tasks (left hand, right hand, tongue and both feet) were performed, and the EEG signals were monitored on 22 channels. The locations of the electrodes employed are shown in Fig. 3 . The sampling rate of the EEG signals is 250 Hz, and the resolution of the amplifier is set to 100 μV. In every trial, a cue appears on the screen at 2 s, and the MI tasks are performed from 3 s to 6 s. A 0.5–100 Hz bandpass filter and a notching filter (50 Hz) are employed to preprocess the collected EEG signals. Abnormal trials are marked as artifacts by the experts. Continuous predictions for the sampling points were obtained, and the accuracy and the kappa value [37] are obtained on every sampling point. The maximum kappa value through all the sampling point is leveraged to evaluate the effectiveness. More details can be found in [36]. To evaluate the proposed method on few-channel MI EEG, electrodes C3, Cz and C4 from the International 10â20 system are applied to implement a two-category classification (left hand and right hand). The kappa value of the two-category classification is calculated based on the accuracy P_{acc}:

$$\text{kappa} = \frac{N_c \times P_{\text{acc}} - 1}{N_c - 1} \tag{16}$$

where N_c, which represents the categories number, equals 2 in the following experiments.

3.2. BCI Competition IV dataset 2b

BCI Competition IV dataset 2b collected EEG signals from 9 subjects for two categories of MI tasks (left hand and right hand) [35,38]. For each subject, 5 sessions were recorded in which the first 3 sessions were used as the training set and the last 2 sessions were used as the evaluation set. The first 2 sessions presented the cue on the screen without feedback, and the 3 last sessions were based on the smiley feedback. The trials numbers in a sessions range from 120 to 160. EEG signals were collected from three channels at 250 Hz (including channel C3, channel Cz and channel C4). A 0.5–100 Hz bandpass filter and a notching filter (50 Hz) are employed to preprocess the collected EEG signals. The trials with artifacts were labeled by specialists. The maximum kappa value through all the sampling point is leveraged to evaluate the effectiveness. Due to the different setup of the sessions, in this paper, only session 3 is employed to train the ESVL classifier, and the experimental performances are evaluated based on sessions 4 and 5.

3.3. Experiments

The implementation of the analysis in this paper is based on MATLAB R2018a and a computer with an Intel(R) Core(TM) i7-8700 3.2 GHz CPU and 16 GB RAM. The MATLAB-based binary SVM classifiers are employed to build the ESVL program. Because the cost of EEG collecting is high and the model is usually trained for specific subject, the MI-based EEG classification try to solve a small sample problem. Specifically, the trials number in the training sets range from 120 to 160 in the experiment of this paper. To increase the number of training samples and prevent overfitting, the trials are segmented into 2-s segmentations with 0.1 s of step. Every 2-s EEG segment is treated as a sample. As the requirement of the BCI Competition, the predictions for every sampling point are needed to generate a causal system. Therefore, the prediction of a sampling point is made based on 2-s EEG signals before this point. Given the equivalence among adjacent sampling points, the classification outputs are only made on every alternate 10th sampling point to reduce calculation load. The setup in this paragraph refers to that of the FBCSP algorithm [14]. The resolution of the energy spectrum is controlled by setting the decomposition level. Similar to the description of the CDFCSP [21], 5-level WPD is performed in the CDFCSP-based feature extraction process.

4. Results

4.1. Discrimination comparison between the CDFCSP- and STDF-based features

The discrimination of the CDFCSP- and STDF-based features is compared in this section. The features with the largest absolute values of the coefficients in the trained SVM classifier (β in Eq. (10)) are treated as the most discriminative features. To provide intuitive results, Fig. 4 presents an example (subjects 3 and 7 from BCI Competition IV dataset 2a) of the discrimination of features. The top 2 discriminative features from the CDFCSP and the STDF are included in Fig. 4 . We can conclude from the figure that (1) the CDFCSP performs well on subject 3 and the STDF performs

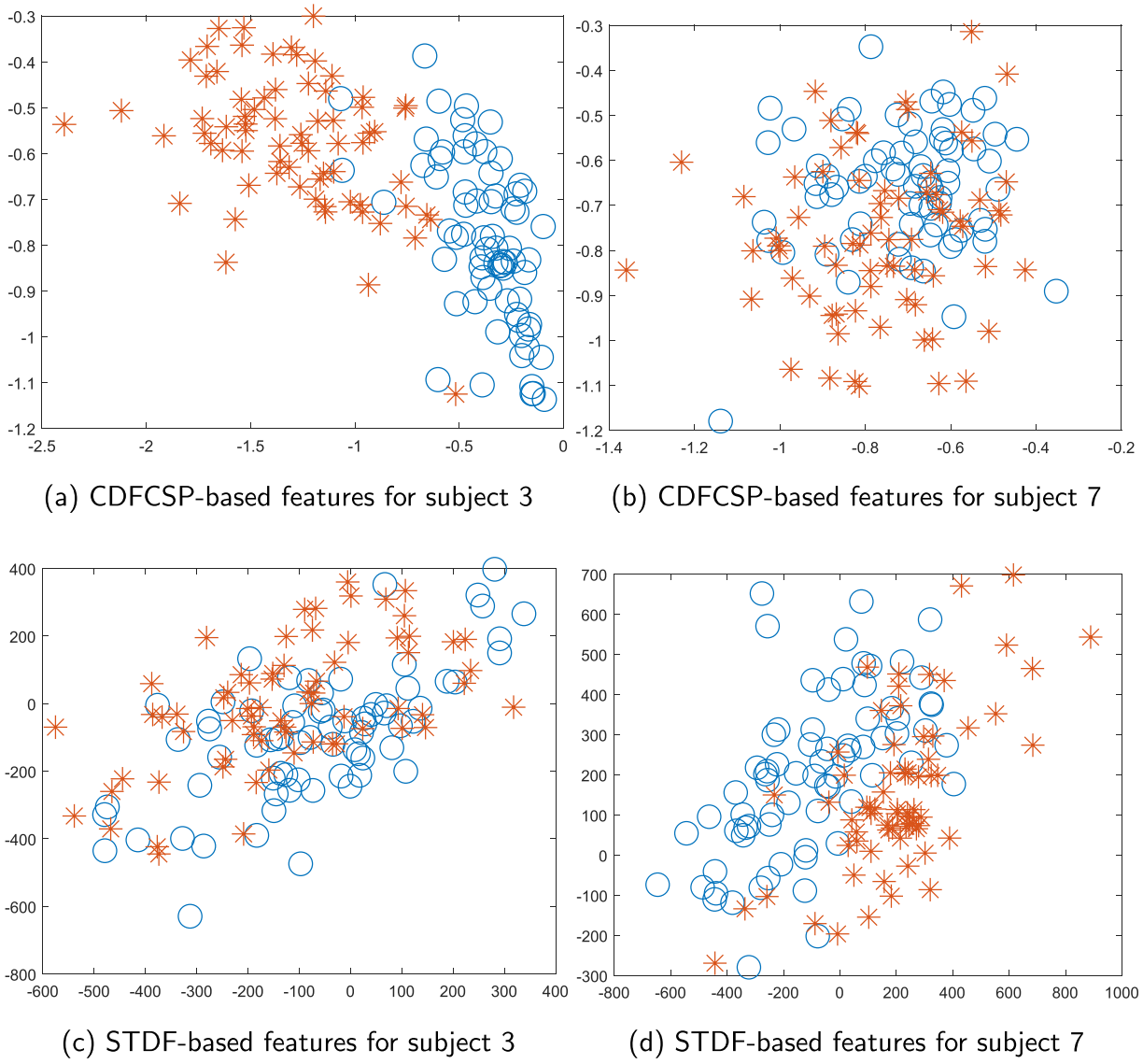


Fig. 4. Discrimination comparison between the CDFCSP-based features and STDF. In the figures, circle indicates the samples belonging to the left hand MI class and asterisk indicates the samples belonging to the right hand MI class.

well on subject 7; and (2) the discrimination of the CDFCSP for subject 7 and the STDF-based features for subject 3 are poor.

To provide an overview of the discriminability of features from different subjects, the squared point-biserial correlation coefficient r^2 of the features is computed [18] as follows:

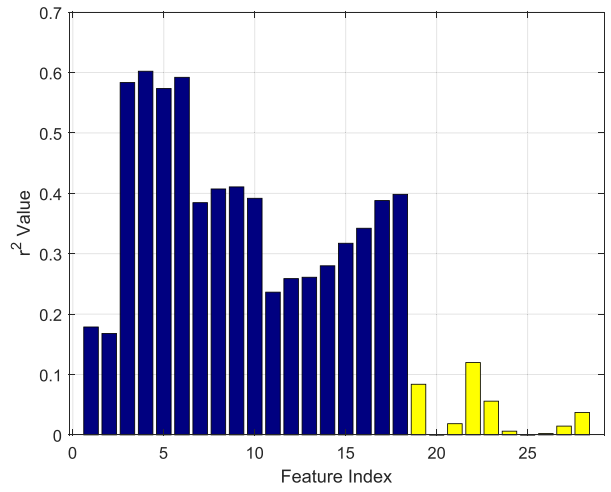
$$r^2(x) = \frac{N_1 N_2}{(N_1 - N_2)^2} \times \frac{(\text{mean}\{x_i|y_i=1\} - \text{mean}\{x_i|y_i=2\})^2}{\text{var}\{x_i|y_i=1,2\}} \tag{17}$$

where x_i is the feature of the ith sample and y_i is the label of the ith sample. N_1 and N_2 are the numbers of samples in the related class. A large r^2 indicates a feature with strong separability. In Fig. 5 , the r^2 values are calculated for features of subject 3 and subject 7 from BCI Competition IV dataset 2a to show the overall discriminability of the CDFCSP-and STDF-based features. The features indexed from 1 to 18 (marked in blue) are the CDFCSP-based features, and the features indexed from 19 to 28 (marked in yellow) are the STDF-based features. From Fig. 5 , the distribution of r^2 suggests that the CDFCSP-based features are more discriminant than the STDF-based features for subject 3, and the 2 most discriminant features of subject 7 are STDF-based features 20 and 21.

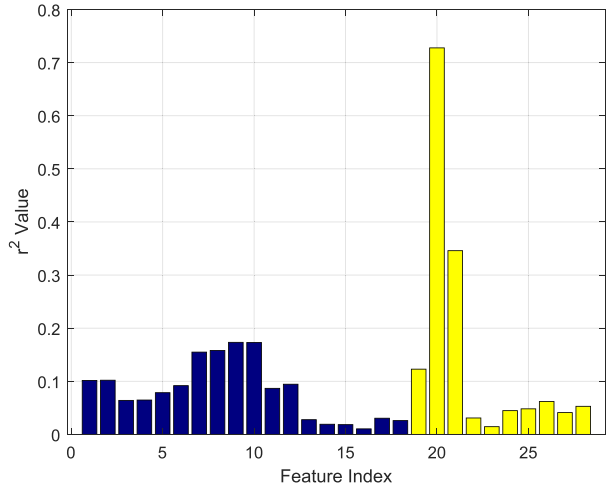
Furthermore, the SVM-based classifications are performed on BCI Competition IV dataset 2a to validate the assumption that the STDF and CDFCSP are good for different subjects. The classification results of the STDF and CDFCSP are evaluated separately based on the linear SVM classifiers. Table 1 presents a comparison of the experimental results based on different features. The maxi-

Table 1
Comparison of the classification results (maximum kappa value) based on the STDF and the CDFCSP. The largest kappa values are marked in bold.

Subject	CDFCSP	STDF
A01	**0.60**	0.18
A02	0.18	**0.25**
A03	**0.87**	0.20
A04	0.**34**	0.33
A05	0.47	**0.79**
A06	**0.37**	0.11
A07	0.50	**0.77**
A08	**0.78**	0.52
A09	**0.69**	0.38
Average	**0.53**	0.39



(a) r^2 values for subject 3



(b) r^2 values for subject 7

Fig. 5. r^2 values for features of subjects 3 and 7 in dataset 2a.

mum kappa value is calculated to evaluate the performance and the highest kappa values are marked in bold.

We can find the following from the experimental comparisons: (1) comparing the result between the STDF and CDFCSP, the STDF obtains a higher kappa on subjects 2, 5, and 7, and the CDFCSP obtains obviously higher kappas on subjects 1, 3, 6, and 9, and (2) the CDFCSP achieves a better average kappa value than the STDF. The experimental results concur with the conclusion from [31], and compared with the ERD/ERS-based features, the ERP-based STDF performs well on different subjects (i.e., the CDFCSP-based features in this paper).

4.2. Performance comparison of the proposed ESVL-based algorithms

To validate the performance improvement from the proposed ESVL-based approach, the classification is conducted according to the framework in Section 2. The histograms of the results are provided in Fig. 6 to elicit an intuitive comparison. Based on the histogram, we can conclude that (1) the ESVL-based approach achieves a kappa value approximately equal to the best feature for each subject; (2) the ESVL-based approach performs best on subjects 1, 3, 4, 6 and 9 and compares favorably with the best features of subjects 7 and 8; and (3) the average kappa values from ESVL based on 9 subjects (0.60) are significantly higher than the average kappa value from a single feature (0.53 and 0.39).

Table 2
Performance comparison (maximum kappa value) of the proposed ESVL-based approach with state-of-the-art algorithms.

Subject	RQNN	DCSP	IDCSP	FBCSP	SBL	CDFCSP	ESVL
A01	0.22	0.46	0.56	0.60	0.59	0.62	**0.63**
A02	0.22	**0.24**	0.21	0.20	0.01	0.20	0.17
A03	0.58	0.20	0.87	0.82	0.85	0.87	**0.88**
A04	0.21	0.09	0.34	0.38	0.36	**0.40**	0.38
A05	0.43	0.17	0.21	0.38	0.44	0.47	**0.69**
A06	0.22	0.31	0.35	0.35	**0.41**	0.37	**0.41**
A07	0.17	0.51	0.46	0.47	0.49	0.50	**0.76**
A08	0.35	0.60	**0.87**	**0.87**	0.81	0.78	0.76
A09	0.58	0.74	**0.82**	0.62	0.63	0.66	0.69
B01	**0.55**	0.31	0.30	0.39	0.41	0.48	0.47
B02	0.27	0.20	0.21	0.22	0.13	0.29	**0.31**
B03	**0.72**	0.30	0.25	0.32	0.28	0.25	0.63
B04	0.96	0.90	0.90	0.80	0.87	**0.97**	0.96
B05	0.83	0.83	0.82	0.90	0.89	**0.98**	0.97
B06	0.71	0.60	0.66	**0.75**	0.72	0.70	0.67
B07	0.46	0.67	0.67	0.66	0.60	0.67	**0.71**
B08	0.84	0.86	**0.88**	0.86	0.89	0.84	0.87
B09	0.75	0.67	0.73	0.62	0.68	**0.79**	**0.79**
Average (2a)	0.33	0.37	0.52	0.52	0.51	0.54	**0.60**
Average (2b)	0.68	0.59	0.60	0.61	0.61	0.66	**0.71**
Average	0.50	0.48	0.56	0.57	0.56	0.60	**0.65**

Table 3
Evaluation of the statistical significance of the performance difference between the proposed ESVL-based approach and the state-of-the-art algorithms compared in Table 2.

ESVL vs.	RQNN	DCSP	IDCSP	FBCSP	SBL	CDFCSP
p-value	0.002	0.001	0.013	0.004	0.001	0.039

The experimental results of the proposed methods are compared with the state-of-the-art algorithms on BCI Competition IV datasets 2a and 2b in Table 2. The compared methods include the recurrent quantum neural network (RQNN) model [39], the DCSP-based approach [16], the IDCSP-based approach [19], the FBCSP-based approach, the SBL-based approach [18] and the CDFCSP-based approach [21]. To evaluate the proposed ESVL based method on few-channel MI EEG, in BCI Competition IV dataset 2a, only EEG from channel C3, Cz and C4 are applied to implement a two-category classification (left hand and right hand).

Some conclusions can be drawn from the experiment results in Table 2: (1) the ESVL based apprach performs best among the state-of-the-art algorithms, and (2) the proposed method presents average performance increases of 82%, 62%, 15%, 15%, 18%, and 11% compared with others in dataset 2a and increases in the kappa value of 4%, 20%, 18%, 16%, 16%, and 8% in dataset 2b.

Student's t-test is also employed, and the p-values are presented in Table 3 to show the significance of the improvement when compared with the state-of-the-art algorithms. All the p-values shown in Table 3 are less than 0.05. In this way, we can conclude that the performance improvements in the experiments are significant.

4.3. Parameter settings

The resolution of the CDF is determined by the level of the WPD as follows:

$$Re = 2^j \tag{18}$$

where j is the level WPD applied at. A higher-resolution filter will be obtained by a higher-level WPD with more computing load. The resolution will be doubled when one additional level is applied in

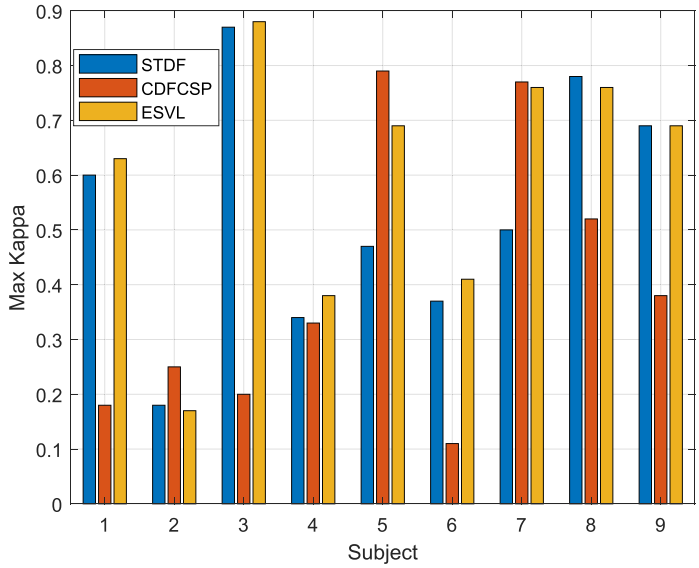


Fig. 6. Histograms of the classification results (maximum kappa value) based on the STDF-, CDFCSP- and ESVL-based approaches.

Table 4
The selected parameters for each single SVM classifier on BCI Competition IV dataset 2a.

SVM	CDFCSP		STDF		Concatenated	
Parameter	*C*	*S*	*C*	*S*	*C*	*S*
A01	0.004	0.001	0.101	977.321	163.873	236.700
A02	0.097	0.085	0.003	76.130	995.259	347.157
A03	18.752	0.515	0.001	114.239	0.040	1.694
A04	43.866	0.546	0.001	63.966	0.115	4.932
A05	274.439	50.679	0.011	412.314	995.219	613.467
A06	0.001	0.063	0.586	276.286	116.036	137.324
A07	0.016	0.199	0.007	225.855	20.096	46.718
A08	43.212	1.203	0.001	256.720	6.636	49.793
A09	36.779	1.148	0.313	948.874	1.111	15.522

WPD. Similar to the experiment performed with the CDFCSP [21], 5-level WPD is performed on the CDFCSP-based feature extraction.

As mentioned in Eq. (11), the penalty parameter *C* controls the penalty for the misclassified samples. The scale is another important parameter in the SVM classifier. All features are divided by the scale as the input features x of the SVM. To achieve the best setting of these two parameters, the fivefold cross-validation loss was minimized by Bayesian optimization. Because the penalty parameter and the scale are both positive values, the searching scopes are both set to [0.001 1000]. One specific ESVL classifier applied in this paper includes three single probabilistic SVM classifiers, and the selected parameters for each single SVM classifier are listed in Table 4 (*C* is the penalty parameter, and *S* is the scale).

5. Discussion

An ESVL-based approach is proposed for two-category MI-based EEG classification. The features applied in this frame include the CDFCSP [21] and STDF-based features [31] and the ESVL classifier is designed to combine the advantages of difficult features. Experiments on two famous public MI classification datasets including 18 subjects (BCI competition IV datasets 2a and 2b) are conducted. Our proposed method achieves an average accuracy of 82.5% which shown obviously improvement compared with the state-of-the-art algorithms. Now, the ESVL-based approach is applied in two-category classification, which limits the application in BCI with higher dimensional controlling requirement. In the next stage of work, we plan to generalize the application of the proposed ESVL-based approach to multicategory MI classification. Most of the existing multicategory MI classification approaches were built based on multiple two-category classifications, such as FBCSP algorithm [14], hierarchical SVM-based approach [40] and multidimensional MEG-based approach [41]. Compared with them, ESVL-based approach outperforms in two-category classifications (accuracy of 78.5% for FBCSP, 70.23% for multidimensional MEG-based approach). This demonstrates that the proposed approach should be promising in multicategory MI classification. The "divide-and-conquer" strategy, the "pairwise" strategy and the "one-versus-rest" strategy [14] will be applied in feature extraction and the training of the classifier.

Real-time BCI (also called asynchronous BCI or self-paced BCI) allows the subject to execute MI tasks at any time, which is more user friendly than a synchronous BCI system [6]. To apply the proposed method in asynchronous BCIs (real-time BCIs), a timing detection algorithm must be added to the system [42]. This MI detection approach is another part of future work.

6. Conclusion

In this paper, we proposed a novel ensemble learning algorithm called ESVL for MI-based EEG classification. ESVL is based on SVM classifiers, and the distances between the sampling point and the decision boundary and the sample labels are applied to fit a sigmoid function. Thus, the SVM outputs are mapped to probabilities. The probabilities of different SVM classifiers are combined to form the ESVL classifier. In this way, ESVL makes full use of the advantages of different classifiers and makes a better prediction based on the posterior probabilities. The CDFCSP and the STDF are applied to extracted discriminative features, and then the features are fed to the ESVL algorithm. The experiments on BCI Competition IV dataset 2a and 2b are conducted to verify the efficiency of the ESVL classifier. The CDFCSP-based SVM and the STDF-based SVM yield a better performance on different subjects, and ESVL shows efficiency on all the subjects in the dataset. The experimental comparison with the state-of-the-art methods suggests that the proposed ESVL-based approach improves the performance of MI-based EEG classifications.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgment

This work is supported by the National Natural Science Foundation of China grants 61906152 and 61876147.

Appendix A. List of acronyms

The list of acronyms is shown in Table 5.

A list of acronyms

Brain-computer interface (BCI)
Class discrepancy guided subband filter-based CSP (CDFCSP)
Common spatial pattern (CSP)
Discriminative CSP (DCSP)
Electroencephalogram (EEG)
Event-related desynchronization (ERD)
Event-related potential (ERP)
Event-related synchronization (ERS)
Ensemble support vector learning (ESVL)
Filter band CSP (FBCSP)
Improved discriminative FBCSP (IDCSP)
Motor Imagery (MI)
Recurrent quantum neural network (RQNN)
Spatiotemporal discrepancy feature (STDF)
Support vector machine (SVM)
Tangent space mapping (TSM)
Wavelet packet decomposition (WPD)

References

[1] P.J. Benson, Decoding brain-computer interfaces, Science 360 (615) (2018) 6.
[2] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, G. Pfurtscheller, Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment, IEEE Trans. Neural Syst. Rehabil. Eng. 15 (4) (2007) 473–482.
[3] S. Ge, R. Wang, Y. Leng, H. Wang, P. Lin, K. Iramina, A double-partial least--squares model for the detection of steady-state visual evoked potentials, IEEE J. Biomed. Health Inform. 21 (4) (2016) 897–903.
[4] L. Van Dokkum, T. Ward, I. Laffont, Brain computer interfaces for neurorehabilitation–its current status as a rehabilitation strategy post-stroke, Ann. Phys. Rehabil. Med. 58 (1) (2015) 3–8.
[5] Z. Qiu, B.Z. Allison, J. Jin, Y. Zhang, X. Wang, W. Li, A. Cichocki, Optimized motor imagery paradigm based on imagining chinese characters writing movement, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (7) (2017) 1009–1017.
[6] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, F. Yger, A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update, J. Neural Eng. 15 (3) (2018) 31005.
[7] S. Ge, Y.-H. Shi, R.-M. Wang, P. Lin, J.-F. Gao, G.-P. Sun, K. Iramina, Y.-K. Yang, Y. Leng, H.-X. Wang, et al., Sinusoidal signal assisted multivariate empirical mode decomposition for brain–computer interfaces, IEEE J. Biomed. Health Inform. 22 (5) (2017) 1373–1384.
[8] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophys. 113 (6) (2002) 767–791.
[9] G. Pfurtscheller, F.L. Da Silva, Event-related eeg/meg synchronization and desynchronization: basic principles, Clin. Neurophys. 110 (11) (1999) 1842–1857.
[10] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial eeg classification in a movement task, Clin. Neurophys. 110 (5) (1999) 787–798.
[11] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, IEEE Trans. Biomed. Eng. 58 (2) (2010) 355–362.
[12] S.-H. Park, D. Lee, S.-G. Lee, Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2) (2017) 498–505.
[13] H. Lu, H.-L. Eng, C. Guan, K.N. Plataniotis, A.N. Venetsanopoulos, Regularized common spatial pattern with aggregation for eeg classification in small-sample setting, IEEE Trans. Biomed. Eng. 57 (12) (2010) 2936–2946.
[14] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition iv datasets 2a and 2b, Front. Neurosci. 6 (2012) 39.
[15] S. Kumar, K. Mamun, A. Sharma, Csp-tsm: optimizing the performance of riemannian tangent space mapping using common spatial pattern for mi-bci, Comput. Biol. Med. 91 (2017) 231–242.
[16] K.P. Thomas, C. Guan, C.T. Lau, A.P. Vinod, K.K. Ang, A new discriminative common spatial pattern method for motor imagery brain–computer interfaces, IEEE Trans. Biomed. Eng. 56 (11) (2009) 2730–2733.
[17] J. Luo, Z. Feng, J. Zhang, N. Lu, Dynamic frequency feature selection based approach for classification of motor imageries, Comput. Biol. Med. 75 (2016) 45–53.
[18] Y. Zhang, Y. Wang, J. Jin, X. Wang, Sparse bayesian learning for obtaining sparsity of eeg frequency bands based feature vectors in motor imagery classification, Int. J. Neural Syst. 27 (2) (2017) 1650032.
[19] S. Kumar, A. Sharma, T. Tsunoda, An improved discriminative filter bank selection approach for motor imagery eeg signal classification using mutual information, BMC Bioinform. 18 (16) (2017) 545.
[20] B. Yang, H. Li, Q. Wang, Y. Zhang, Subject-based feature extraction by using fisher WPD-CSP in brain–computer interfaces, Comput. Methods Programs Biomed. 129 (2016) 21–28.
[21] J. Luo, J. Wang, R. Xu, K. Xu, Class discrepancy-guided sub-band filter-based common spatial pattern for motor imagery classification, J. Neurosci. Methods 323 (2019) 98–107.
[22] S.J. Luck, An Introduction to the Event-Related Potential Technique, MIT Press, 2014.
[23] N. Lu, T. Yin, Motor imagery classification via combinatory decomposition of ERP and ERSP using sparse nonnegative matrix factorization, J. Neurosci. Methods 249 (2015) 41–49.
[24] N. Lu, T. Li, J. Pan, X. Ren, Z. Feng, H. Miao, Structure constrained semi-nonnegative matrix factorization for eeg-based motor imagery classification, Comput. Biol. Med. 60 (2015) 32–39.
[25] C. Babiloni, F. Carducci, F. Cincotti, P.M. Rossini, C. Neuper, G. Pfurtscheller, F. Babiloni, Human movement-related potentials vs desynchronization of eeg alpha rhythm: a high-resolution eeg study, Neuroimage 10 (6) (1999) 658–665.
[26] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for eeg-based brain–computer interfaces, J. Neural Eng. 4 (2) (2007) R1.
[27] L. Quitadamo, F. Cavrini, L. Sbernini, F. Riillo, L. Bianchi, S. Seri, G. Saggio, Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review, J. Neural Eng. 14 (1) (2017) 11001.
[28] J. Platt, et al., Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Adv. Large Margin Classif. 10 (3) (1999) 61–74.
[29] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inf. Theory 36 (5) (1990) 961–1005.
[30] I. Daubechies, Ten Lectures on Wavelets, Siam, 1992.
[31] J. Luo, Z. Feng, N. Lu, Spatio-temporal discrepancy feature for classification of motor imageries, Biomed. Signal Process. Control 47 (2019) 137–144.
[32] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 2390–2397.
[33] W. Wu, C. Wu, S. Gao, B. Liu, Y. Li, X. Gao, Bayesian estimation of ERP components from multicondition and multichannel EEG, Neuroimage 88 (2014) 319–339.
[34] W. Wu, Z. Chen, X. Gao, Y. Li, E.N. Brown, S. Gao, Probabilistic common spatial patterns for multichannel eeg analysis, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2014) 639–653.
[35] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, et al., Review of the BCI competition IV, Front. Neurosci. 6 (2012) 55.
[36] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, G. Pfurtscheller, Bci competition 2008–graz data set a, Institute for Knowledge Discovery (Laboratory of Brain–Computer Interfaces), Graz University of Technology, 16, 2008.
[37] G. Dornhege, J.d.R. Millan, T. Hinterberger, D.J. McFarland, K.-R. Müller, Toward Brain-Computer Interfacing, MIT press, 2007.
[38] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, G. Pfurtscheller, Bci competition 2008–graz data set b, in: Graz University of Technology, Austria, 2008, pp. 1–6.
[39] V. Gandhi, G. Prasad, D. Coyle, L. Behera, T.M. McGinnity, Quantum neural network-based eeg filtering for a brain–computer interface, IEEE Trans Neural Netw Learn Syst 25 (2) (2013) 278–288.
[40] E. Dong, C. Li, L. Li, S. Du, A.N. Belkacem, C. Chen, Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces, Med. Biol. Eng. Comput. 55 (10) (2017) 1809–1818.
[41] A.N. Belkacem, S. Nishio, T. Suzuki, H. Ishiguro, M. Hirata, Neuromagnetic decoding of simultaneous bilateral hand movements for multidimensional brain–machine interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (6) (2018) 1301–1310.
[42] S. Shahid, G. Prasad, Bispectrum-based feature extraction technique for devising a practical brain–computer interface, J. Neural Eng. 8 (2) (2011) 25014.