Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI*

Zhi-chuan TANG[†1,2], Chao LI[†2], Jian-feng WU[†‡1], Peng-cheng LIU[†3], Shi-wei CHENG[4]

[1]*Industrial Design Institute, Zhejiang University of Technology, Hangzhou 310014, China*

[2]*College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China*

[3]*Lincoln Centre for Autonomous Systems, University of Lincoln, Lincoln LN57DH, UK*

[4]*College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310014, China*

[†]E-mail: ttzzcc@zjut.edu.cn; superli@zju.edu.cn; jianfw@zjut.edu.cn; pliu@cardiffmet.ac.uk

Received Feb. 1, 2018; Revision accepted Apr. 24, 2018; Crosschecked Aug. 13, 2019

Abstract: Classifying single-trial electroencephalogram (EEG) based motor imagery (MI) tasks is extensively used to control brain-computer interface (BCI) applications, as a communication bridge between humans and computers. However, the low signal-to-noise ratio and individual differences of EEG can affect the classification results negatively. In this paper, we propose an improved common spatial pattern (B-CSP) method to extract features for alleviating these adverse effects. First, for different subjects, the method of Bhattacharyya distance is used to select the optimal frequency band of each electrode including strong event-related desynchronization (ERD) and event-related synchronization (ERS) patterns; then the signals of the optimal frequency band are decomposed into spatial patterns, and the features that can describe the maximum differences of two classes of MI are extracted from the EEG data. The proposed method is applied to the public data set and experimental data set to extract features which are input into a back propagation neural network (BPNN) classifier to classify single-trial MI EEG. Another two conventional feature extraction methods, original common spatial pattern (CSP) and autoregressive (AR), are used for comparison. An improved classification performance for both data sets (public data set: 91.25%±1.77% for left hand vs. foot and 84.50%±5.42% for left hand vs. right hand; experimental data set: 90.43%±4.26% for left hand vs. foot) verifies the advantages of the B-CSP method over conventional methods. The results demonstrate that our proposed B-CSP method can classify EEG-based MI tasks effectively, and this study provides practical and theoretical approaches to BCI applications.

Key words: Electroencephalogram (EEG); Motor imagery (MI); Improved common spatial pattern (B-CSP); Feature extraction; Classification

https://doi.org/10.1631/FITEE.1800083 **CLC number:** TP391.4

1 Introduction

As a communication bridge between humans and computers, brain-computer interface (BCI) based on electroencephalogram (EEG) aims to decode brain activity into different control commands directly (Lotte et al., 2007; Nicolas-Alonso and Gomez-Gil, 2012; Zhang et al., 2013). The BCI approach is a safe tool that enables seriously disabled people with normal brain to control an external device like an exoskeleton or neuro-prosthesis.

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61702454 and 61772468), the MOE Project of Humanities and Social Sciences, China (No. 17YJC870018), the Fundamental Research Funds for the Provincial Universities of Zhejiang Province, China (No. GB201901006), and the Philosophy and Social Science Planning Fund Project of Zhejiang Province, China (No. 20NDQN260YB)

ORCID: Zhi-chuan TANG, http://orcid.org/0000-0002-1730-1120

During human-computer communication, these people's intentions are translated from their EEG signals (Wolpaw et al., 2002; Moghimi et al., 2013; Salisbury et al., 2016).

Motor imagery (MI) EEG, as a new communication approach, has been widely used in BCI applications (Ang et al., 2010; Yang et al., 2015; Kevric and Subasi, 2017). Imagining the movement of different parts of limbs and body can lead to changes of brain rhythms (mu- and beta-rhythm) that can be measured at the corresponding scalp (Neuper et al., 2006). For example, when imagining left hand or right hand movement, the power decrease of mu- and beta-rhythm occurs in the contralateral motor region at the scalp. This is called an event-related desynchronization (ERD) pattern (Pfurtscheller and Neuper, 1997; Pfurtscheller, 2001; Graimann et al., 2002). The power increase of mu- and beta-rhythm occurs in the ipsilateral motor region at the scalp. This is called an event-related synchronization (ERS) pattern (Cassim et al., 2000; Pfurtscheller and Neuper, 2006; Nam et al., 2011). The MI of foot and tongue can produce similar patterns in the corresponding regions. We can use these physiological patterns caused by MI of different limbs (left hand movement versus right hand movement versus foot movement) to classify EEG and obtain control signals of external applications (Pfurtscheller et al., 1997; Gomarus et al., 2006).

EEG is a signal with a low signal-to-noise ratio (SNR); i.e., the signals are surrounded by high noise (Lemm et al., 2005). Therefore, how to define and extract appropriate features for effectively translating and encoding MI EEG is a challenge to classify EEG-based MI tasks. Feature extraction methods for EEG have been widely discussed by previous studies, including autoregressive (AR), wavelet coefficients (WC), and average power (AP). An AR model is used to describe certain time-varying processes like MI EEG, and the AR parameters can be used as the features for MI classification (Franaszczuk and Bergey, 1999). Pfurtscheller et al. (1998) used an adaptive AR model with a recursive-least-squares (RLS) algorithm to estimate AR parameters for left and right MI classification. Zhang et al. (2017) proposed an EEG feature extraction method combined with an AR model and wavelet packet decomposition to calculate AR parameters for five-mental-task classification. Some previous studies used another feature extraction method called common spatial pattern (CSP) for MI EEG classification (Müller-Gerking et al., 1999; Ramoser et al., 2000; Robinson et al., 2013). The original CSP method can find spatial filters that maximize differences in variance between two classes in a high-dimensional space (Lemm et al., 2005). Some of these studies using improved CSP methods have achieved a better classification performance (Qaraqe et al., 2015; Gaur et al., 2018). Kumar et al. (2016) applied Fisher discriminant analysis (FDA) for the CSP features and fed FDA scores to a classifier for MI EEG classification, resulting in an average reduction of 1.07% in the classification error rate. Afterwards, they proposed a discriminative filter band selection method combined with CSP and mutual information (Kumar et al., 2017a) and a single band CSP framework using the concept of tangent space mapping (TSM) (Kumar et al., 2017b) to further enhance the performance of CSP. Yuksel and Olmez (2015) developed an improved CSP framework called the spatial filter network (SFN), which calculates optimal spatial filters using a neural network approach and obtained an increased MI EEG classification accuracy in comparison with conventional CSP methods. Wu et al. (2015) proposed a probabilistic CSP method as a generic EEG spatio-temporal modeling framework that subsumes the CSP and regularized CSP algorithms, and this method was successfully applied to single-trial classifications of three MI EEG data sets. However, most of these methods are used to extract features from the specific and fixed frequency bands. Because of the individual differences of EEG signals, for different people, the ERD/ERS patterns will occur in different frequency bands; even for each electrode, the optimal frequency bands are different. Individual differences and low SNR are two factors that may affect the classification results negatively, leading to an unsatisfactory classification accuracy of below 85% in most previous studies.

The goal of this study is to propose an improved common spatial pattern (B-CSP) method to extract features for alleviating these adverse effects. First, for different subjects, the method of Bhattacharyya distance is used to select the optimal frequency band of each electrode including strong ERD and ERS patterns; then the signals of the optimal frequency band are decomposed into spatial patterns, and the features that can describe the maximum differences

of two classes of MI are extracted from the EEG data. The proposed method is applied in the public data set and experimental data set to extract features, which are input into a back propagation neural network (BPNN) classifier to classify single-trial MI EEG. Another two conventional feature extraction methods (original CSP and AR) are used for comparison.

2 Materials and methods

2.1 Public data set

Data set 1 of BCI Competition IV was the public data set in this study (Tangermann et al., 2012). The data were recorded from seven subjects. Five (ds1b, ds1c, ds1d, ds1e, and ds1g) performed left hand MI and right hand MI; two (ds1a and ds1f) performed left hand MI and foot MI. For each subject, we selected data in the first two runs, resulting in a total of 200 trials (100 trials for each class of MI). For details of the data set and experimental procedure, please refer to Tangermann et al. (2012).

2.2 Experimental data set

2.2.1 Subjects

Three healthy subjects (two males and one female, aged 27–30 years) were employed in this study. All were right handed by the Edinburgh Handedness Inventory (Oldfield, 1971). None had been informed of the experimental hypothesis.

2.2.2 Data acquisition

The EEG signals of 28 active electrodes (10-10 system, as shown in Fig. 1) attached on an EEG cap were collected using a Biosemi ActiveTwo EEG system (64-channel, DC amplifier, 24-bit resolution). In this system, the common mode sense (CMS) electrode and driven right leg (DRL) electrode replace the ground electrode. The reference electrode was located on the left mastoid. In our experiment, the sample rate of EEG was 1000 Hz, the notch filter was 50 Hz, the low-pass filter was 100 Hz, and the high-pass filter was 0.5 Hz. Before attaching the electrodes, medical alcohol and conductive gel were used for cleaning the skin and reducing electrode resistance, respectively.

2.2.3 Experimental procedure

The experimental procedure for the experimental data set was similar to that for the public data set. After the EEG electrodes were attached and all signals were normal, subjects sat before a computer screen, and put two arms on the table naturally. In the experiment, they should avoid head and body movement, eye blinks or swallowing, especially during the MI task (the visual cue onset). Each subject was asked to imagine two classes of movement (left hand and foot) in five runs in total. One MI was performed in one trial. In each run, 94 trials (47 left hand MI and 47 foot MI) were performed, and this made up a data set of a total of 470 trials for each subject. The time sequence of one trial is as shown in Fig. 2. A single trial lasted 8 s. The screen was blank during the first 2 s, and then a fixation cross

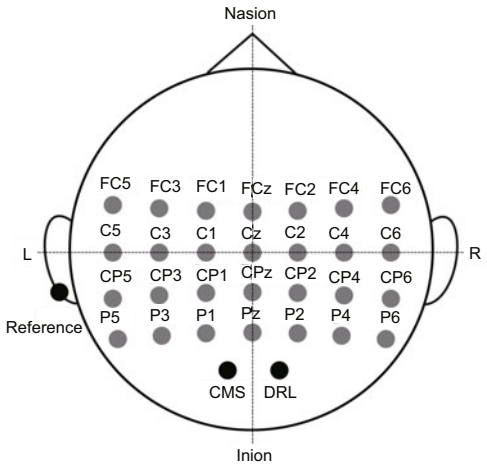

Fig. 1 Electroencephalogram (EEG) electrode location, including 28 active electrodes, common mode sense (CMS) and driven right leg (DRL) electrodes, and a reference electrode

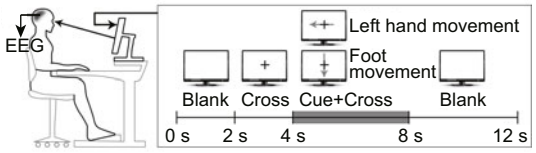

Fig. 2 Time sequence of one trial of the experimental data set. A single trial lasted 8 s. The screen was blank during the first 2 s, and then a fixation cross ("+") appeared in the center of the screen for 2 s, to remind subjects of the start of MI. During seconds 4–8, a random visual cue, i.e., left arrow "←" or down arrow "↓," floated on the cross to instruct the corresponding MI task. After the cue appeared, subjects were required to perform the cued MI task until the screen was blank. A 4-s break between two trials was given for relaxation

("+") appeared in the center of the screen for 2 s, to remind subjects of the start of MI. During seconds 4–8, a random visual cue, i.e., left arrow "←" or down arrow "↓," floated on the cross to instruct the corresponding MI task. After the cue appeared, subjects were required to perform the cued MI task until the screen was blank. A 4-s break between two trials was given for relaxation. There was a longer break of 10 min between two runs.

2.3 Data preprocessing

The important movement classification information exists mainly in the 8–30 Hz frequency band (alpha and beta) of EEG data (Pfurtscheller et al., 1997). For public and experimental data sets, the raw EEG signal was filtered in this frequency band. The time course graph of ERD/ERS and head topographies were created to analyze the different ERD/ERS patterns of each class of MI (left hand movement, right hand movement, and foot movement). The EEG signals of each trial during the time period of 4 s before and 4 s after cue appearance were extracted to draw the time course graph of ERD/ERS. In the time course graph, the extracted EEG signals of three channels (C3, Cz, and C4) from all trials of each subject were calculated by a superposed average method (Pfurtscheller and da Silva, 1999); i.e., the EEG signals of each channel were averaged across all trials. ERD and ERS can be described as the power decrease percentage and power increase percentage of a target period to a reference period (seconds 1–2 in our study) respectively, according to

$$\mathrm{ERD/ERS}=\frac{\mathrm{EEG_{target}}-\mathrm{EEG_{reference}}}{\mathrm{EEG_{reference}}}\times 100\%, \quad (1)$$

where $\mathrm{EEG_{target}}$ is the target period's EEG power and $\mathrm{EEG_{reference}}$ is the reference period's EEG power. To visually observe the ERD and ERS from the time course graph, the EEG data of the time period that presented the strongest ERD/ERS were selected for feature extraction and classification. Then the head topographies of this time period were plotted for further analysis. For the public and experimental data sets, each time period (t) was segmented by 50-ms time windows. Each input sample was a matrix ($\boldsymbol{M}$) of $N\times T$, where N ($N=28$) is the number of channels and T $\left(T=\dfrac{t\times 1000\ \mathrm{Hz}}{50\ \mathrm{ms}}\right)$ is the number of time points.

2.4 B-CSP method for feature extraction

The CSP method can find spatial filters that maximize differences in variance between two classes in a high-dimensional space, and this is effectively applied in the binary classification problem (Lemm et al., 2005). The original CSP method extracts features from the specific and fixed frequency bands of EEG signals. However, when different people perform the MI, the ERD/ERS patterns will happen in different frequency bands. Empirical selection cannot obtain the optimal frequency band of each person. Furthermore, for the same person, the optimal frequency bands of each electrode are different.

To alleviate the adverse effect of individual differences, we propose an improved CSP (B-CSP) method for feature extraction. The steps of the B-CSP method are shown in Fig. 3. First, for different subjects, the optimal frequency band of each electrode including strong ERD/ERS patterns was selected. The method of Bhattacharyya distance was used; then the signals of the optimal frequency band were decomposed into spatial patterns, and the features that could describe the maximum differences of two classes of MI were extracted from the EEG data. The details of this proposed method are as follows:

Step 1: According to the different MI, all input matrices are divided into two classes, i.e., CL1 and CL2. For all matrices of each class, EEG signals of each channel are filtered in k frequency bands and then averaged, where $k=6$ and the frequency bands are 8–12, 12–16, 16–20, 20–24, 24–28, and 28–30 Hz.

Step 2: For EEG signals of each frequency band of each channel, the Bhattacharyya distance is used to measure the separability between two classes of MI. A larger value of the Bhattacharyya distance in one frequency band means that the EEG data of this frequency band can result in a better classification performance (Bai et al., 2007). The Bhattacharyya distance (D_{B}) and Bhattacharyya coefficient (BC) can be calculated by

$$D_{\mathrm{B}}(\mathrm{CL1}_n^m,\mathrm{CL2}_n^m)=-\ln[\mathrm{BC}(\mathrm{CL1}_n^m,\mathrm{CL2}_n^m)], \quad (2)$$

$$\mathrm{BC}(\mathrm{CL1}_n^m,\mathrm{CL2}_n^m)=\sum_{x\le T}\sqrt{\mathrm{CL1}_n^m(x)\mathrm{CL2}_n^m(x)}, \quad (3)$$

where $\mathrm{CL1}_n^m(x)$ and $\mathrm{CL2}_n^m(x)$ are EEG signals of the m^{th} frequency band of the n^{th} channel of the m^{th} time point of two classes MI, $n\le N$, $m\le k$. The frequency band with the largest Bhattacharyya

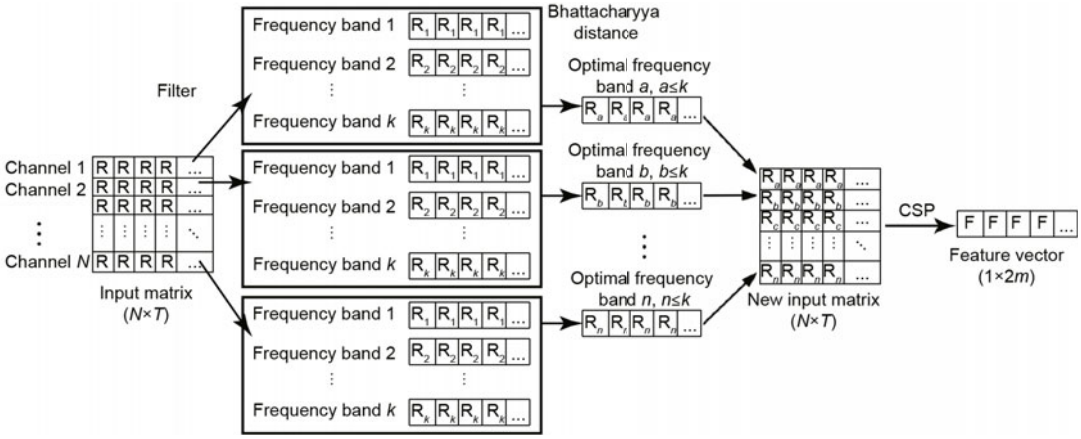

Fig. 3 Improved common spatial pattern (B-CSP) method for feature extraction

distance is the optimal frequency band of this channel.

Step 3: The EEG signals of each channel of the raw input matrix $\boldsymbol{M}$ are filtered in the corresponding optimal frequency, generating a new input matrix $\boldsymbol{M}_{D_\mathrm{B}}$ of $N \times T$. The two classes' covariance matrices are given as

$$\mathbf{Cov} = \frac{\boldsymbol{M}_{D_\mathrm{B}} \boldsymbol{M}_{D_\mathrm{B}}^\mathrm{T}}{\mathrm{tr}(\boldsymbol{M}_{D_\mathrm{B}} \boldsymbol{M}_{D_\mathrm{B}}^\mathrm{T})}, \tag{4}$$

where $\mathrm{tr}(\cdot)$ represents the sum of $\boldsymbol{M}$'s diagonal values. To remove the trial-to-trial variations, a normalization operation should be carried out. The averaged normalized covariance matrices of two classes are $\overline{\mathbf{Cov}_{\mathrm{CL1}}}$ and $\overline{\mathbf{Cov}_{\mathrm{CL2}}}$, which can be built by averaging all input matrices of each class. Then the composite covariance matrix is calculated by

$$\mathbf{Cov}_\mathrm{R} = \overline{\mathbf{Cov}_{\mathrm{CL1}}} + \overline{\mathbf{Cov}_{\mathrm{CL2}}}, \tag{5}$$

which can be factored into its eigenvectors by

$$\mathbf{Cov}_\mathrm{R} = \boldsymbol{E}_\mathrm{R} \boldsymbol{\lambda}_\mathrm{R} \boldsymbol{E}_\mathrm{R}^\mathrm{T}, \tag{6}$$

where $\boldsymbol{E}_\mathrm{R}$ is the eigenvector matrix and $\boldsymbol{\lambda}_\mathrm{R}$ is the eigenvalue matrix. The whitening matrix can be constructed by

$$\boldsymbol{W} = \sqrt{\boldsymbol{\lambda}_\mathrm{R}^{-1}} \boldsymbol{E}_\mathrm{R}^\mathrm{T}, \tag{7}$$

and according to the whitening transformation, $\overline{\mathbf{Cov}_{\mathrm{CL1}}}$ and $\overline{\mathbf{Cov}_{\mathrm{CL2}}}$ are transformed to

$$\begin{cases} \boldsymbol{S}_{\mathrm{CL1}} = \boldsymbol{W} \overline{\mathbf{Cov}_{\mathrm{CL1}}} \boldsymbol{W}^\mathrm{T}, \\ \boldsymbol{S}_{\mathrm{CL2}} = \boldsymbol{W} \overline{\mathbf{Cov}_{\mathrm{CL2}}} \boldsymbol{W}^\mathrm{T}, \end{cases} \tag{8}$$

where the sum of the corresponding eigenvalues of $\boldsymbol{S}_{\mathrm{CL1}}$ and $\boldsymbol{S}_{\mathrm{CL2}}$ equals 1. $\boldsymbol{S}_{\mathrm{CL1}}$ and $\boldsymbol{S}_{\mathrm{CL2}}$ are decomposed by the shared eigenvector $\boldsymbol{U}$:

$$\begin{cases} \boldsymbol{S}_{\mathrm{CL1}} = \boldsymbol{U} \boldsymbol{\lambda}_{\mathrm{CL1}} \boldsymbol{U}^\mathrm{T}, \\ \boldsymbol{S}_{\mathrm{CL2}} = \boldsymbol{U} \boldsymbol{\lambda}_{\mathrm{CL2}} \boldsymbol{U}^\mathrm{T}, \end{cases} \tag{9}$$

where $\boldsymbol{\lambda}_{\mathrm{CL1}} + \boldsymbol{\lambda}_{\mathrm{CL2}} = \boldsymbol{I}$ with $\boldsymbol{I}$ the identity matrix. Because the sum of two corresponding eigenvalues is constant, the eigenvalue for $\boldsymbol{S}_{\mathrm{CL1}}$ in eigenvector $\boldsymbol{U}$ is the largest when the eigenvalue for $\boldsymbol{S}_{\mathrm{CL2}}$ in eigenvector $\boldsymbol{U}$ is the smallest. The projection matrix $\boldsymbol{P}^\mathrm{T} = \boldsymbol{U}^\mathrm{T} \boldsymbol{W}$ multiplied by input matrix $\boldsymbol{M}_{D_\mathrm{B}}$ forms the decomposed signal matrix:

$$\boldsymbol{Z} = \boldsymbol{P}^\mathrm{T} \boldsymbol{M}_{D_\mathrm{B}}. \tag{10}$$

The columns of $(\boldsymbol{P}^\mathrm{T})^{-1}$ are the common spatial patterns and can be seen as time-invariant EEG source distribution vectors.

Step 4: The variances of the first and last m rows of $\boldsymbol{Z}$ are used as the features for classification, containing most of the discriminative information between the two classes (Müller-Gerking et al., 1999). The feature vector for input matrix $\boldsymbol{i}$ is

$$\boldsymbol{F}_i^k = \log\left(\frac{\mathrm{var}(\boldsymbol{Z}_i^k)}{\sum\limits_{k=1}^{2m} \mathrm{var}(\boldsymbol{Z}_i^k)}\right), \tag{11}$$

where $\mathrm{var}(\boldsymbol{Z}_i^k)$ is the variance of the k^{th} row of the i^{th} input matrix, and $k = 1, 2, \ldots, 2m$.

2.5 Classification

In this study, we used a back propagation neural network (BPNN) as the classifier, since it can learn

any nonlinear function and has been widely applied in EEG classification (Subasi, 2005). The inputs of a BPNN were the features extracted by the B-CSP method, and the output of a BPNN was one of three classes (left hand MI, right hand MI, and foot MI). In the model training step, the classification model of each subject was trained using his/her own EEG data. For the public data set and experimental data set, we randomly selected 80% data as the training set to train the models, and the remaining 20% data were used as the testing set to evaluate the models. Furthermore, the data of the training set were randomly divided into 10 folds (nine folds were used for training and one for validation) for 10-fold cross validation, to fine-tune the parameters and select the optimal model.

Another two conventional feature extraction methods, original CSP and AR, were employed to train the BPNN classification models for comparison with the B-CSP method. For the original CSP method, the raw EEG signals of each trial were filtered in the 8–30 Hz frequency band to form an input matrix ($\boldsymbol{M}$) of $N \times T$. These input matrices were used to extract features using the original CSP method, with the same steps as steps 3 and 4 of the B-CSP method. Six spatial filters with a bandwidth of 4 Hz covering the frequency band from 8 to 30 Hz were adopted for the original CSP method. Then these features were input into the BPNN for MI classification. For the AR method, an adaptive AR model was built following the procedure of Pfurtscheller et al. (1998). The EEG signal Y_d can be described by the following model:

$$Y_d = a_{1,d}Y_{d-1} + a_{2,d}Y_{d-2} + ... + a_{p,d}Y_{d-p} + X_d, \quad (12)$$

where d ($d \leq T$) is the time point, $a_{1,d}, a_{2,d}, ..., a_{p,d}$ are AR parameters which can vary with time, p is the order of the AR model, and X_d is a white noise process. The RLS algorithm was used to estimate the AR parameters. An AR model with an order of $p = 6$ was used. This resulted in a feature vector with a dimension of 28×6 for each time point of each trial. Then these feature vectors were input into the BPNN for MI classification.

Given the testing results, we applied the receiver operating characteristic (ROC) curve and accuracy to evaluate the classification models, and applied precision, recall, and F-score to evaluate each MI's classification performance.

3 Results and discussion

3.1 ERD/ERS analysis

For each MI of two data sets, the 8–12 Hz EEG signals of each trial during the time period of 4 s before and 4 s after cue appearance were extracted to draw the time course graph of ERD/ERS (Fig. 4). In the time course graphs, second 4 means the cue occurrence. For the public data set, during the left hand MI and right hand MI, a strong ERD can be seen in a time period of 3 s after cue appearance (seconds 4–7) over the contralateral side at the scalp, and a weak ERS was observed over the ipsilateral side at the scalp and around the Cz electrode. In addition, a strong post-movement ERS was found in the end of one MI trial (seconds 7–8) over the contralateral side at the scalp. During the foot MI, a strong ERD can be seen around the Cz electrode, and a strong ERS can be seen over the bilateral sides at the scalp (around the C3 and C4 electrodes). The experimental data set had similar results in the time course graphs to the public data set. There is an ERD phenomenon (i.e., a power decrease) in the hand area at the scalp and an ERS phenomenon (i.e., a power increase) in the foot area at the scalp when performing hand MI, and there is an opposite pattern when performing foot MI. Furthermore, the head topographies of the time period of seconds 4–7 were plotted beside the time course graphs. In the head topographies, the blue color represents an ERD phenomenon and the red color represents an ERS phenomenon

3.2 Classification results

According to Fig. 4, the time period of 3 s after the cue appearance (seconds 4–7) that presented the strongest ERD/ERS was selected for feature extraction and classification. During the feature extraction step, we first used the Bhattacharyya distance to select the optimal frequency band of each channel of each subject. Fig. 5 shows the Bhattacharyya distance values of different frequency bands of each channel from subject 2, experimental data set, as a representative example. The yellow color stands for a larger value, while the blue color stands for a smaller value. An optimal frequency band of one channel can be selected from a frequency band with the largest Bhattacharyya distance value, which means that the EEG data of this frequency band have the highest

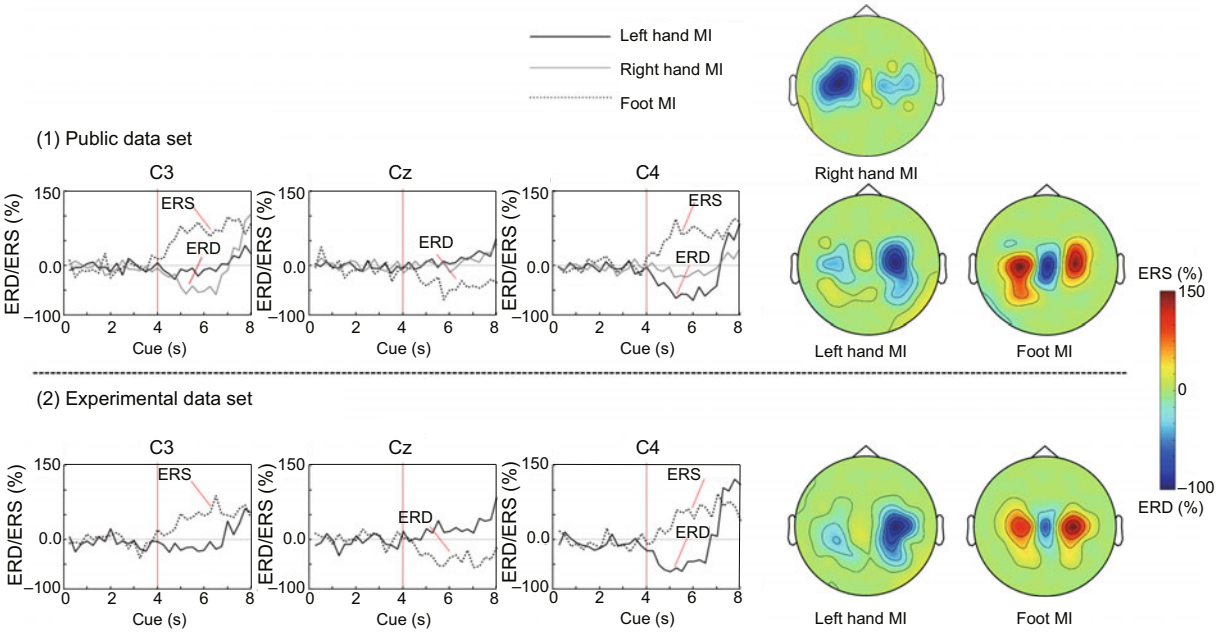

Fig. 4 ERD/ERS analysis including time course graphs and head topographies. For each MI of two data sets, the 8–12 Hz EEG signals of each trial during the time period of 4 s before and 4 s after cue appearance were extracted to draw the time course graph of ERD/ERS. In the time course graph, the extracted EEG signals of three channels (C3, Cz, and C4) from all trials of each subject were calculated using the superposed average method. ERD and ERS are displayed as the power decrease percentage and power increase percentage of a target period to a reference period. The head topographies of the time period of seconds 4–7 are plotted beside the time course graphs. In the head topographies, the blue color represents an ERD phenomenon (a power decrease) and the red color represents an ERS phenomenon (a power increase). References to color refer to the online version of this figure

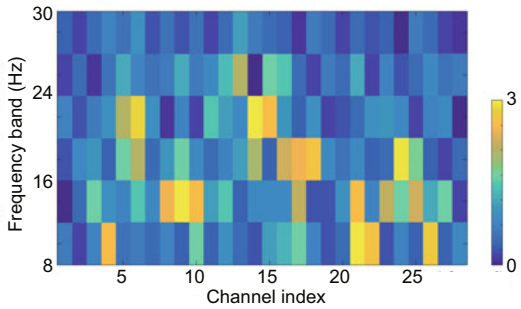

Fig. 5 Bhattacharyya distance values of different frequency bands of each channel from subject 2, experimental data set. The yellow color stands for a larger value, while the blue color stands for a smaller value. An optimal frequency band of one channel can be selected from a frequency band with the largest Bhattacharyya distance value, which means that the EEG data of this frequency band have the highest separability between two classes of MI in this channel. References to color refer to the online version of this figure

separability between two classes of MI in this channel. Then the EEG signals of each channel of the raw input matrix were filtered in the corresponding optimal frequency band, generating a new input matrix. The feature vector was extracted from the new input matrix using the CSP algorithm.

For both public and experimental data sets, the classification model of each subject was trained using his/her own EEG feature data. Another two conventional feature extraction methods, original CSP and AR, were employed to train the BPNN classification models for comparison. Fig. 6 shows the confusion matrix results of the three feature extraction methods using the same testing sets of the two data sets. The values of each element of the confusion matrix mean the average value (upper number) and standard deviation (lower number) over all subjects. The color element represents the correct classification percentage, and the white element represents the wrong classification percentage. For each subject, the accuracies of three methods for testing sets of the two data sets can be seen in Table 1. To draw the ROC curves, left hand movement is

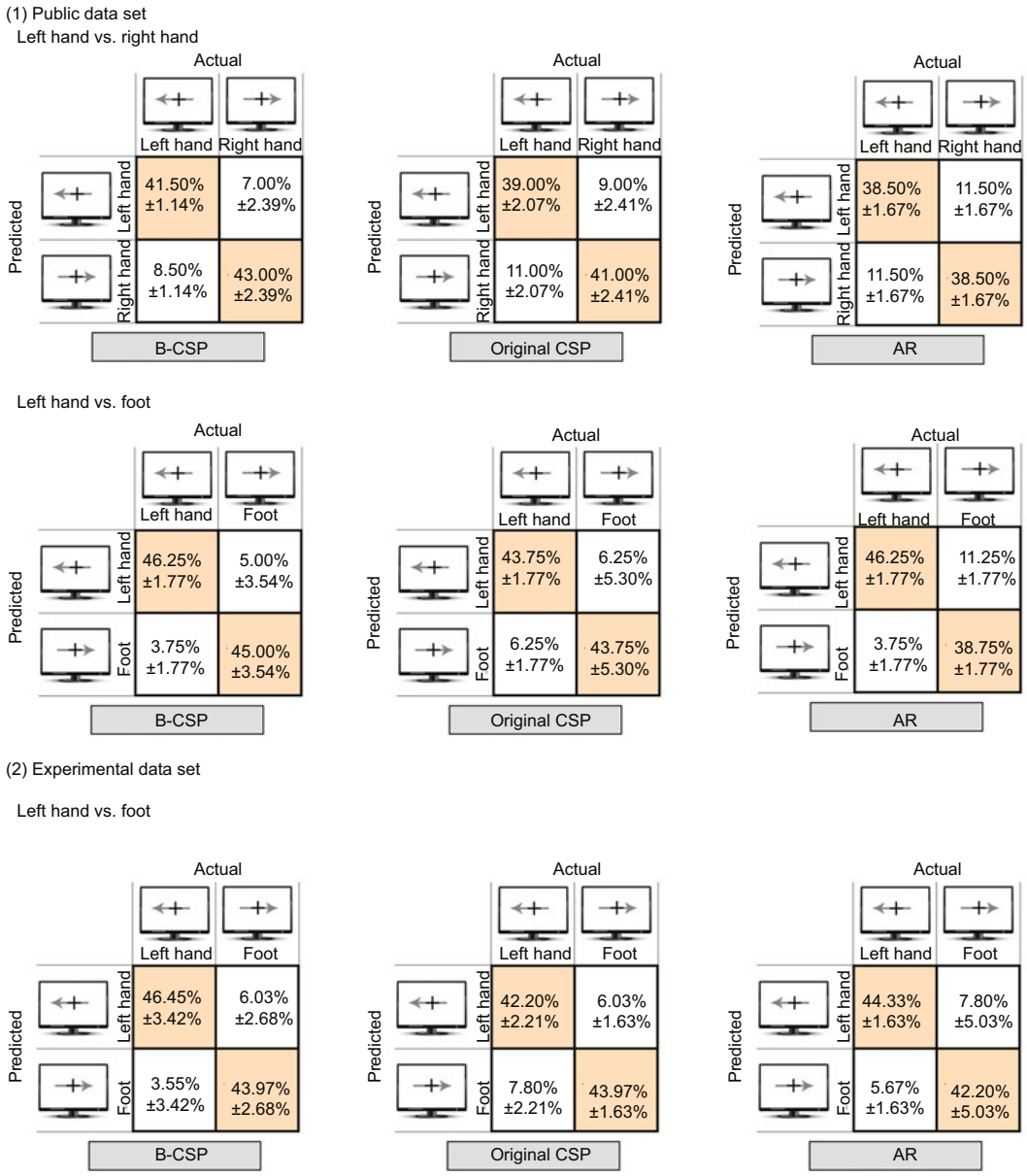

Fig. 6 Confusion matrix results of three feature extraction methods using the same two testing sets of data. The values of each element of the confusion matrix mean the average value (upper number) and standard deviation (lower number) over all subjects. The color element represents the correct classification percentage, and the white element represents the wrong classification percentage. References to color refer to the online version of this figure

defined as the positive class, and foot movement is defined as the negative class. The ROC curves of three feature extraction methods using testing sets of ds1a (public data set) and subject 2 (experimental data set) are shown in Fig. 7. According to Fig. 6 and Table 1, for two subjects performing left hand and foot MI of the public data set, the average accuracy using B-CSP (91.25%±1.77%) is 3.75% and 6.25% higher than those using original CSP (87.50%±3.54%) and AR (85.00%±3.54%) respectively; for five subjects performing left hand and right hand MI of the public data set, the average accuracy using B-CSP (84.50%±5.42%) is 4.50% and 7.50% higher than those using original CSP (80.00%±5.30%) and AR (77.00%±6.22%) respectively; for three subjects performing left hand and

Table 1 Accuracies of three feature extraction methods using the same testing sets of two data sets (all subjects)

Data set	Subject	Accuracy (%)		
		B-CSP	Original CSP	AR
Public data set (left hand vs. right hand)	ds1b	82.50	80.00	67.50
	ds1c	92.50	75.00	82.50
	ds1d	80.00	82.50	75.00
	ds1e	80.00	75.00	82.50
	ds1g	87.50	87.50	77.50
	Average	84.50±5.42	80.00±5.30	77.00±6.22
Public data set (left hand vs. foot)	ds1a	90.00	90.00	82.50
	ds1f	92.50	85.00	87.50
	Average	91.25±1.77	87.50±3.54	85.00±3.54
Experimental data set (left hand vs. foot)	Subject 1	94.68	84.04	86.17
	Subject 2	86.17	88.30	80.85
	Subject 3	90.43	86.17	92.55
	Average	90.43±4.26	86.17±2.13	86.52±5.86

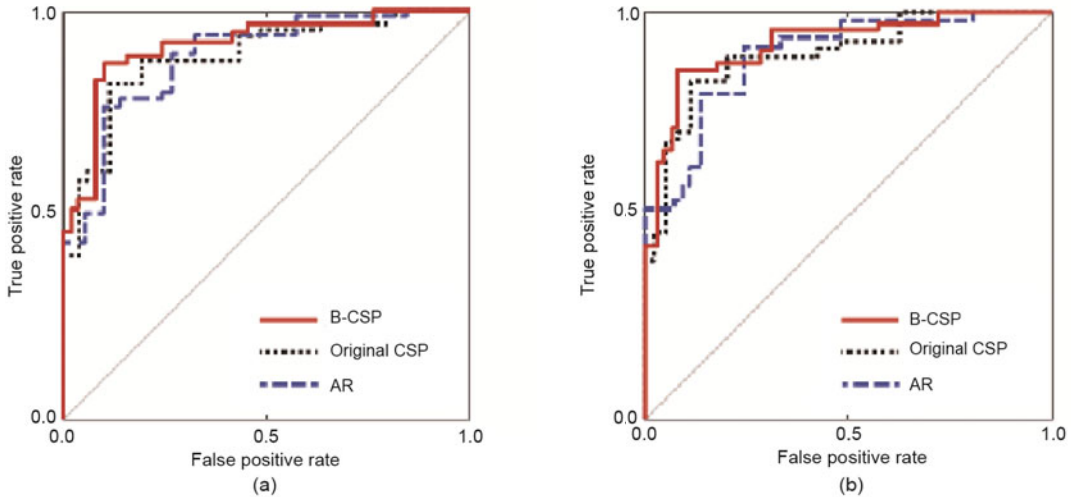

Fig. 7 Receiver operating characteristic (ROC) curves of three feature extraction methods using testing sets of ds1a (public data set) (a) and subject 2 (experimental data set) (b)

foot MI of the experimental data set, the average accuracy using B-CSP (90.43%±4.26%) is 4.26% and 3.91% higher than those using original CSP (86.17%±2.13%) and AR (86.52%±5.86%) respectively. In addition, an area under curve (AUC) using the B-CSP method larger than that using the other two methods is shown in Fig. 7. These similar results in two data sets demonstrate that the classification model trained by the B-CSP feature extraction method with a BPNN has better classification performance.

For both public and experimental data sets, the precision, recall, and F-score values of each class of MI using the three methods can be seen in Table 2. The larger precision, recall, and F-score values mean better performance of the classification model B-CSP. A 3 (B-CSP, original CSP, and AR)×2 (left hand MI and foot/right hand MI) analysis of variance (ANOVA) was applied in this study to evaluate the interaction of "feature extraction method" × "MI," and to evaluate the main effects of the "feature extraction method" and "MI class." A 95% confidence interval is given for all statistics. The statistical results of ANOVA show that the interaction between the feature extraction method and MI is not significant ($p > 0.05$); the feature extraction method can significantly affect classification accuracy ($F = 9.450$, $p < 0.001$), but the MI

Table 2 Precision, recall, and *F*-score values of each class of MI using three methods (all subjects)

Data set	Subject	Evaluation	B-CSP		Original CSP		AR	
			Left hand	Foot/ right hand	Left hand	Foot/ right hand	Left hand	Foot/ right hand
Public data set (left hand vs. right hand)	ds1b	Precision	0.8095	0.8421	0.8750	0.7500	0.6842	0.6667
		Recall	0.8500	0.8000	0.7000	0.9000	0.6500	0.7000
		F-score	0.8293	0.8205	0.7778	0.8182	0.6667	0.6829
	ds1c	Precision	0.9474	0.9048	0.8125	0.7083	0.8095	0.8421
		Recall	0.9000	0.9500	0.6500	0.8500	0.8500	0.8000
		F-score	0.9231	0.9268	0.7222	0.7727	0.8293	0.8205
	ds1d	Precision	0.7727	0.8333	0.7826	0.8824	0.7500	0.7500
		Recall	0.8500	0.7500	0.9000	0.7500	0.7500	0.7500
		F-score	0.8095	0.7895	0.8372	0.8108	0.7500	0.7500
	ds1e	Precision	0.8000	0.8000	0.7083	0.8125	0.8824	0.7826
		Recall	0.8000	0.8000	0.8500	0.6500	0.7500	0.9000
		F-score	0.8000	0.8000	0.7727	0.7222	0.8108	0.8372
	ds1g	Precision	1.0000	0.8000	0.9412	0.8261	0.7391	0.8235
		Recall	0.7500	1.0000	0.8000	0.9500	0.8500	0.7000
		F-score	0.8571	0.8889	0.8649	0.8837	0.7907	0.7567
Public data set (left hand vs. foot)	ds1a	Precision	0.8636	0.9444	0.9444	0.8636	0.7826	0.8824
		Recall	0.9500	0.8500	0.8500	0.9500	0.9000	0.7500
		F-score	0.9048	0.8947	0.8947	0.9047	0.8372	0.8108
	ds1f	Precision	0.9474	0.9048	0.8182	0.8889	0.8261	0.9412
		Recall	0.9000	0.9500	0.9000	0.8000	0.9500	0.8000
		F-score	0.9231	0.9268	0.8571	0.8421	0.8837	0.8649
Experimental data set (left hand vs. foot)	Subject 1	Precision	0.9375	0.9565	0.8478	0.8333	0.8269	0.9048
		Recall	0.9574	0.9362	0.8298	0.8511	0.9149	0.8085
		F-score	0.9474	0.9462	0.8387	0.8421	0.8687	0.8539
	Subject 2	Precision	0.8696	0.8542	0.8750	0.8913	0.7843	0.8372
		Recall	0.8511	0.8723	0.8936	0.8723	0.8511	0.8298
		F-score	0.8602	0.8632	0.8842	0.8817	0.8163	0.8335
	Subject 3	Precision	0.8519	0.9750	0.9048	0.8269	0.9545	0.9000
		Recall	0.9787	0.8298	0.8085	0.9149	0.8936	0.9574
		F-score	0.9109	0.8966	0.8539	0.8687	0.9231	0.9278

has no significant effect on classification accuracy ($F = 0.082, p > 0.05$).

4 Conclusions

In this paper, we have proposed an improved common spatial pattern (B-CSP) method to extract features for alleviating the adverse effects of low SNR and individual differences. First, for different subjects, the optimal frequency band of each electrode including strong ERD/ERS patterns was selected, using the method of Bhattacharyya distance; then the signals of the optimal frequency band were decomposed into spatial patterns, and the features that can describe the maximum differences of two classes of MI were extracted from the EEG data. An improved classification performance for both data sets (public data set: 91.25%±1.77% for left hand vs. foot and 84.50%±5.42% for left hand vs. right hand; experimental data set: 90.43%±4.26% for left hand vs. foot) verified the advantages of the B-CSP method over conventional methods. The results demonstrated that our proposed B-CSP method can classify EEG-based MI tasks effectively, and this study provides practical and theoretical approaches to BCI applications.

Compliance with ethics guidelines

Zhi-chuan TANG, Chao LI, Jian-feng WU, Peng-cheng LIU, and Shi-wei CHENG declare that they have no conflict of interest.

The Ethics Committee of Zhejiang University of Technology had reviewed the experimental procedure and method, and approved this experiment. Before the experiment, all subjects signed the informed written consent and agreed to participate in this experiment.

References

Ang KK, Guan CT, Chua KSG, et al., 2010. Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. Annual Int Conf of the IEEE Engineering in Medicine and Biology, p.5549-5552. https://doi.org/10.1109/IEMBS.2010.5626782

Bai O, Lin P, Vorbach S, et al., 2007. A high performance sensorimotor beta rhythm-based brain-computer interface associated with human natural motor behavior. *J Neur Eng*, 5(1):24-35. https://doi.org/10.1088/1741-2560/5/1/003

Cassim F, Szurhaj W, Sediri H, et al., 2000. Brief and sustained movements: differences in event-related (de)synchronization (ERD/ERS) patterns. *Clin Neurophysiol*, 111(11):2032-2039. https://doi.org/10.1016/S1388-2457(00)00455-7

Franaszczuk PJ, Bergey GK, 1999. An autoregressive method for the measurement of synchronization of interictal and ictal EEG signals. *Biol Cybern*, 81(1):3-9. https://doi.org/10.1007/s004220050540

Gaur P, Pachori RB, Wang H, et al., 2018. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. *Exp Syst Appl*, 95:201-211. https://doi.org/10.1016/j.eswa.2017.11.007

Gomarus HK, Althaus M, Wijers AA, et al., 2006. The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task: an ERP and ERD/ERS study. *Clin Neurophysiol*, 117(4):871-884. https://doi.org/10.1016/j.clinph.2005.12.008

Graimann B, Huggins JE, Levine SP, et al., 2002. Visualization of significant ERD/ERS patterns in multichannel EEG and ECoG data. *Clin Neurophysiol*, 113(1):43-47. https://doi.org/10.1016/S1388-2457(01)00697-6

Kevric J, Subasi A, 2017. Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system. *Biomed Signal Process Contr*, 31:398-406. https://doi.org/10.1016/j.bspc.2016.09.007

Kumar S, Sharma R, Sharma A, et al., 2016. Decimation filter with common spatial pattern and Fishers discriminant analysis for motor imagery classification. Int Joint Conf on Neural Networks, p.2090-2095. https://doi.org/10.1109/IJCNN.2016.7727457

Kumar S, Mamun K, Sharma A, 2017a. CSP-TSM: optimizing the performance of Riemannian tangent space mapping using common spatial pattern for MI-BCI. *Comput Biol Med*, 91:231-242. https://doi.org/10.1016/j.compbiomed.2017.10.025

Kumar S, Sharma A, Tsunoda T, 2017b. An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information. *BMC Bioinform*, 18(S16), Article 125. https://doi.org/10.1186/s12859-017-1538-7

Lemm S, Blankertz B, Curio G, et al., 2005. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans Biomed Eng*, 52(9):1541-1548. https://doi.org/10.1109/TBME.2005.851521

Lotte F, Congedo M, Lécuyer A, et al., 2007. A review of classification algorithms for EEG-based brain-computer interfaces. *J Neural Eng*, 4(2):R1-R13. https://doi.org/10.1088/1741-2560/4/2/R01

Moghimi S, Kushki A, Marie Guerguerian A, et al., 2013. A review of EEG-based brain-computer interfaces as access pathways for individuals with severe disabilities. *Assist Technol*, 25(2):99-110. https://doi.org/10.1080/10400435.2012.723298

Müller-Gerking J, Pfurtscheller G, Flyvbjerg H, 1999. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin Neurophysiol*, 110(5):787-798. https://doi.org/10.1016/S1388-2457(98)00038-8

Nam CS, Jeon Y, Kim YJ, et al., 2011. Movement imagery-related lateralization of event-related (de)synchronization (ERD/ERS): motor-imagery duration effects. *Clin Neurophysiol*, 122(3):567-577. https://doi.org/10.1016/j.clinph.2010.08.002

Neuper C, Wörtz M, Pfurtscheller G, 2006. ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Prog Brain Res*, 159:211-222. https://doi.org/10.1016/S0079-6123(06)59014-4

Nicolas-Alonso LF, Gomez-Gil J, 2012. Brain computer interfaces, a review. *Sensors*, 12(2):1211-1279. https://doi.org/10.3390/s120201211

Oldfield RC, 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1):97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Pfurtscheller G, 2001. Functional brain imaging based on ERD/ERS. *Vis Res*, 41(10-11):1257-1260. https://doi.org/10.1016/S0042-6989(00)00235-2

Pfurtscheller G, da Silva FHL, 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol*, 110(11):1842-1857. https://doi.org/10.1016/S1388-2457(99)00141-8

Pfurtscheller G, Neuper C, 1997. Motor imagery activates primary sensorimotor area in humans. *Neurosci Lett*, 239(2-3):65-68. https://doi.org/10.1016/S0304-3940(97)00889-6

Pfurtscheller G, Neuper C, 2006. Future prospects of ERD/ERS in the context of brain-computer interface (BCI) developments. *Prog Brain Res*, 159:433-437. https://doi.org/10.1016/S0079-6123(06)59028-4

Pfurtscheller G, Neuper C, Flotzinger D, et al., 1997. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalogr Clin Neurophysiol*, 103(6):642-651. https://doi.org/10.1016/S0013-4694(97)00080-1

Pfurtscheller G, Neuper C, Schlogl A, et al., 1998. Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans Rehabil Eng*, 6(3):316-325. https://doi.org/10.1109/86.712230

Qaraqe M, Ismail M, Serpedin E, 2015. Band-sensitive seizure onset detection via CSP-enhanced EEG features. *Epilep Behav*, 50:77-87. https://doi.org/10.1016/j.yebeh.2015.06.002

Ramoser H, Muller-Gerking J, Pfurtscheller G, 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng*, 8(4):441-446. https://doi.org/10.1109/86.895946

Robinson N, Vinod AP, Ang KK, et al., 2013. EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm. *IEEE Trans Biomed Eng*,

60(8):2123-2132.
https://doi.org/10.1109/TBME.2013.2248153

Salisbury DB, Parsons TD, Monden KR, et al., 2016. Brain-computer interface for individuals after spinal cord injury. *Rehabil Psychol*, 61(4):435-441.
https://doi.org/10.1037/rep0000099

Subasi A, 2005. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Exp Syst Appl*, 28(4):701-711.
https://doi.org/10.1016/j.eswa.2004.12.027

Tangermann M, Müller KR, Aertsen A, et al., 2012. Review of the BCI competition IV. *Front Neurosci*, 6, Article 55. https://doi.org/10.3389/fnins.2012.00055

Wolpaw JR, Birbaumer N, McFarland DJ, et al., 2002. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113(6):767-791.
https://doi.org/10.1016/S1388-2457(02)00057-3

Wu W, Chen Z, Gao XR, et al., 2015. Probabilistic common spatial patterns for multichannel EEG analysis. *IEEE Trans Patt Anal Mach Intell*, 37(3):639-653.
https://doi.org/10.1109/TPAMI.2014.2330598

Yang HJ, Guan CT, Wang CC, et al., 2015. Detection of motor imagery of brisk walking from electroencephalogram. *J Neurosci Methods*, 244:33-44.
https://doi.org/10.1016/j.jneumeth.2014.05.007

Yuksel A, Olmez T, 2015. A neural network-based optimal spatial filter design method for motor imagery classification. *PLoS ONE*, 10(5):e0125039.
https://doi.org/10.1371/journal.pone.0125039

Zhang HH, Yang HJ, Guan CT, 2013. Bayesian learning for spatial filtering in an EEG-based brain-computer interface. *IEEE Trans Neur Netw Learn Syst*, 24(7):1049-1060. https://doi.org/10.1109/TNNLS.2013.2249087

Zhang Y, Liu B, Ji XM, et al., 2017. Classification of EEG signals based on autoregressive model and wavelet packet decomposition. *Neur Process Lett*, 45(2):365-378. https://doi.org/10.1007/s11063-016-9530-1