



**Research article**

## Optimal channel-based sparse time-frequency blocks common spatial pattern feature extraction method for motor imagery classification

Xu Yin<sup>1</sup>, Ming Meng<sup>1,2,\*</sup>, Qingshan She<sup>1,2</sup>, Yunyuan Gao<sup>1,2</sup> and Zhizeng Luo<sup>1</sup>

<sup>1</sup> Institute of Intelligent Control and Robotics, Hangzhou Dianzi University, Hangzhou 310018, China

<sup>2</sup> Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province, Hangzhou 310018, China

\* Correspondence: Email: mnming@hdu.edu.cn; Tel: +8613957108993.

**Abstract:** Common spatial pattern (CSP) as a spatial filtering method has been most widely applied to electroencephalogram (EEG) feature extraction to classify motor imagery (MI) in brain-computer interface (BCI) applications. The effectiveness of CSP is determined by the quality of interception in a specific time window and frequency band. Although numerous algorithms have been designed to optimize CSP by splitting the EEG data with a sliding time window and dividing the frequency bands with a set of band-pass filters, simultaneously. However, they did not consider the drawbacks of the rapid increase in data volume and feature dimensions brought about by this method, which would reduce the classification accuracy and calculation efficiency of the model. Therefore, we propose an optimal channel-based sparse time-frequency blocks common spatial pattern (OCSB-CSP) feature extraction method to improve the model classification accuracy and computational efficiency. Comparative experiments on two public EEG datasets show that the proposed method can quickly select significant time-frequency blocks and improve classification performance. The average classification accuracies are higher than those of other winners' methods, providing a new idea for the improvement of BCI applications.

**Keywords:** brain-computer interface; electroencephalogram; motor imagery; common spatial pattern

---

### 1. Introduction

As one of the emerging technologies in the field of neurorehabilitation, the brain-computer

interface (BCI) aims to provide a new non-muscular channel for paralyzed people to communicate with outside world [1]. Electroencephalography (EEG) is widely used for BCI systems owing to its convenience, low cost, and high temporal resolution. At present, the commonly used paradigms for BCI system control include steady-state visual evoked potentials (SSVEPs) [2,3], event-related potentials (ERPs) [4,5], and motor imagery (MI) [6–8]. Compared with BCIs that require external active stimulation, MI-based BCI systems are easier to implement [9,10]. Sensory Motor Rhythm (SMR) is characterized as a band power change within a particular EEG frequency band appearing over a sensorimotor area of the brain during MI. Accordingly, BCI systems can be designed to use EEG band power changes associated with MI tasks as control signals [11]. Since EEG is a non-stationary, low amplitude and low signal-to-noise ratio bioelectric signal [12,13], and MI is an unstable, easily disturbed, and no obvious characteristic paradigm [14,15], which poses a huge challenge for correctly identifying MI intentions. Therefore, how to design reasonable algorithms to extract - discriminative features is particularly critical for MI-EEG recognition.

CSP is widely used for feature extraction of MI-related tasks [16–18]. However, the performance of CSP on specific subjects is closely related to the selection of time windows and frequency bands, so that some CSP methods for time windows and frequency bands improvements are proposed to increase the robustness and classification performance of the model [19–21]. Temporally constrained sparse group spatial pattern (TSGSP) [22] further enhance classification accuracy of MI EEG by the simultaneous optimization of filter bands and time window. Sparse linear discriminant analysis (GSDA) [23] extracted CSP features from the divided time-frequency blocks and used the generalized method to simultaneously select features and classify to improve the classification accuracy. And wrapped time-frequency combined selection in the source domain (WTFS-SD) [19] applied weighted minimum norm estimate and CSP based sub-band feature extraction to decode the MI-tasks.

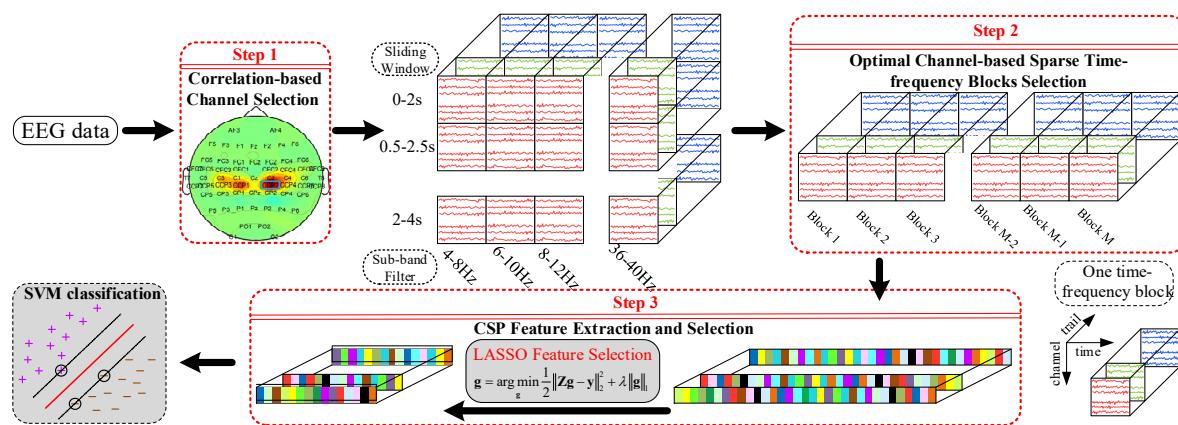
The main idea of above these algorithms is to extract, fuse and select the CSP features of EEG signals from multiple specific time windows and subbands in a single trial of specific subjects, so as to make up for the defects of high sensitivity to noise and low generalization capacity of CSP, and improve the representation ability of different motor imagery tasks. However, they did not consider the sharp increase of data volume and feature dimension, which will increase the computational complexity of the model, which is not conducive to the development and application of online BCI. In addition, the excessive division of time-frequency blocks will in turn lead to information redundancy, resulting in the reduction of classification accuracy.

To solve the above problem, we propose an optimal channel-based sparse time-frequency block common spatial pattern (OCSB-CSP) feature extraction method to reduce the computational burden and increase the classification accuracy of the model. First, a correlation-based method is implemented to select channels and mark the optimal one. Second, the discriminative ability of each time-frequency block is calculated based on the one-dimensional EEG data of the optimal channel, and the block selection is performed with this index. Then, feature extraction and selection are executed for sparse time-frequency blocks. Finally, support vector machine (SVM) is used for classification.

The rest of this paper is organized as follows. Section 2 explains the proposed method. Section 3 describes the experiment study. Section 4 provides the discussion. Finally, the conclusion is drawn in section 5.

## 2. Methods

Our proposed framework consists of three main parts, Correlation-based channel selection, Optimal channel-based sparse time-frequency blocks selection, and CSP feature extraction and selection. Firstly, redundant information is removed, and the data dimensionality is initially reduced by step 1, and the optimal channel for subsequent processing is determined. Secondly, we select the time-frequency blocks rapidly and efficiently by step 2. Then the feature extraction and selection of sparse time-frequency blocks are performed by step 3. Finally, the classification is performed with SVM. The overall framework of the proposed method is illustrated in Figure 1.



**Figure 1.** The overall framework of the proposed OCSB-CSP method for MI classification. It includes three main parts: correlation-based channel selection, optimal channel-based sparse time-frequency blocks selection, and CSP feature extraction and selection.

### 2.1. Correlation-based channel selection

As the number of electrodes for recording EEG data increases, it is essential to adopt effective channel selection algorithms to reduce the computational complexity and channels redundancy. Different criteria functions are predefined, and different channel selection evaluation algorithms are generated, such as [24,25] proposed a filter method to select the significant channels by using Pearson correlation coefficient and bispectrum analysis; [26] adopted a wrapper method to select the important channels by genetic algorithms (GA) and adjust it according to the classification results of fisher discriminant analysis (FDA); [27] raised a hybrid method of two-stage channel selection. In the first stage, a single channel was formed by averaging the channels were combined according to the neurophysiological information about brain functions acquired from the literature, and in the second stage, selective channels were specified with the common spatial pattern-linear discriminant analysis (CSP-LDA)-based sequential channel removal.

The aim of this paper is to propose a method that can rapidly select time-frequency blocks to reduce the computational burden and improve the classification accuracy. Therefore, we use a filtering method in the first step to efficiently select the channels related to MI, which removes the redundant information between channels and facilitates the improvement of classification accuracy; on the other hand, we mark the optimal channel by this step and use the one-dimensional data based on the optimal

channel for the subsequent time-frequency block selection, which significantly reduces the computational complexity of the model compared to using multi-channel data. The combination of these two points shows the reasonableness and necessity of the method. As the specific subjects perform the same MI task, MI-related channels should contain common information, while other channels contain less common information. Based on this principle, we use the Person correlation coefficient, a classic method to quantify the statistical relationship between two or more variables, to measure the similarity between any two channels in reference [28]. In the following, we discuss the necessary steps:

Firstly, in order to reduce the error caused by individual variability or external interference, the raw broadband (1–42 Hz) EEG data is normalized with Z-score, so that the mean of each channel data is equal to 0 and the variance is equal to 1:

$$S_c = (X_c - \text{mean}(X_c)) / \text{std}(X_c), \quad c = 1, \dots, C, \quad (1)$$

where  $X_c$  and  $S_c$  are the  $c^{th}$  channel time series before and after normalizing, and  $C$  is the total number of channels.

Secondly, the Person correlation coefficient between the two channels is calculated by the following formula:

$$\rho(S_i, S_j) = \frac{1}{N-1} \sum_{n=1}^N \left( \frac{S_i(n) - \text{mean}(S_i)}{\text{std}(S_i)} \right) \left( \frac{S_j(n) - \text{mean}(S_j)}{\text{std}(S_j)} \right), \quad i, j = 1, 2, \dots, C, \quad (2)$$

where  $S_i$  and  $S_j$  are the time series of the  $i^{th}$  and  $j^{th}$  channels normalized by formula (1), and  $N$  is the number of sampling points in the sequence.

Thirdly, the correlation coefficient matrix  $R \in \mathbb{R}^{C \times C}$  is calculated:

$$R = \begin{bmatrix} \rho(S_1, S_1) & \cdots & \rho(S_1, S_C) \\ \vdots & \ddots & \vdots \\ \rho(S_C, S_1) & \cdots & \rho(S_C, S_C) \end{bmatrix}. \quad (3)$$

Then the average common information of each channel and all remaining channels is obtained by averaging each row of  $R$ , and the channel corresponding to the maximum value among them is regarded as the selection result of this trial. Repeat all trials as described above to obtain cumulative selection results for all channels.

Finally, count the selection results of all trials, the  $K$  channels with the most selections are regarded as the final channel selection results, and the channel corresponding to the maximum value is marked as the optimal channel  $k^*$ . The selection of channels by the way of this statistical voting does not generate additional hyperparameters that cause an increase in the computational burden.

## 2.2. Optimal channel-based sparse time-frequency blocks selection

For improving classification accuracy of MI-related EEG, we optimize the filter bands and time windows within CSP simultaneously. Specifically, with regards to bandpass filtering,  $K$  channels EEG data determined by section 2.1 is first decomposed into multiple specific frequencies signals at  $F$  overlapping filter bands. Each of the specific frequency signals is further segmented into multiple subseries using  $T$  overlapping sliding windows. Finally,  $F \times T$  time-frequency blocks are formed.

The time series of the optimal channel are taken out from each time-frequency block, and the

time-domain power feature and frequency-domain power spectral density (PSD) feature are extracted. Then, the two-dimensional features are projected to one-dimension through Fisher ratio, to characterize the MI classification ability of each time-frequency block, and to select the time-frequency blocks based on this. Fisher ratio is a statistic parameter that can be used to measure the discriminant ability of classes by projecting high-dimensional features into one dimension [29], It is defined as:

$$score_F = \frac{[mean(F^-) - mean(F^+)]^2}{var(F^-) + var(F^+)}, \quad (4)$$

where  $F^-$  and  $F^+$  are feature vectors from two different classes.

For the time series of the optimal channel in  $i^{th}$  trial  $X_{i,k^*}$  from the each time-frequency block, its power feature is defined as  $T_{i,k^*}$ , PSD feature is defined as  $P_{i,k^*}$  [30]:

$$T_{i,k^*} = \log[var(X_{i,k^*})], \quad (5)$$

$$P_{i,k^*} = \frac{1}{N} \left| \sum_{n=1}^N X_{i,k^*}(n) e^{-j\omega n} \right|^2. \quad (6)$$

According to the definition of Fisher ratio in formula (4) and the above two features, we define the value of the binary classification capability of each time-frequency block as  $score_F$  by calculating the ratio of the Euclidean distance between classes to the intra-class variances. The specific formula is:

$$score_F = \frac{\left[ \left( \frac{1}{|N_1|} \sum_{i \in N_1} T_{i,k^*} \right) - \left( \frac{1}{|N_2|} \sum_{i \in N_2} T_{i,k^*} \right) \right]^2}{\left( \frac{1}{|N_1|} \sum_{i \in N_1} (T_{i,k^*})^2 \right) + \left( \frac{1}{|N_2|} \sum_{i \in N_2} (T_{i,k^*})^2 \right)} + \frac{\left[ \left( \frac{1}{|N_1|} \sum_{i \in N_1} P_{i,k^*} \right) - \left( \frac{1}{|N_2|} \sum_{i \in N_2} P_{i,k^*} \right) \right]^2}{\left( \frac{1}{|N_1|} \sum_{i \in N_1} (P_{i,k^*})^2 \right) + \left( \frac{1}{|N_2|} \sum_{i \in N_2} (P_{i,k^*})^2 \right)}, \quad (7)$$

where  $N_1$  and  $N_2$  are the number of trails in class 1 and class 2, respectively. The  $score_F$  of all  $F \times T$  time-frequency blocks are obtained by the above formula, and the high-quality  $M$  time-frequency blocks are selected by setting reasonable threshold. A reasonable threshold setting will have a significant impact on the subsequent CSP performance. If the threshold is set too large, the number of selected time-frequency blocks will be less, resulting in the loss of a large number of effective information; adversely, if the threshold is set too small, the number of selected time-frequency blocks will be more, resulting in a large number of invalid redundant information. The power features in the time domain and the PSD features in the frequency domain of each time-frequency block are extracted and compressed into a one-dimensional indicator using the Fisher ratio to characterize the binary classification capability of each block. This method also ensures the rationality and efficiency of block selection to a certain extent.

### 2.3. CSP feature extraction and selection

As we all know, CSP is very sensitive to noise, and the CSP feature extraction and fusion of multiple time-frequency blocks are easy to make the model over-fitting. Therefore, many improved CSP feature extraction and selection methods have been proposed [31–33]. We perform CSP feature extraction and selection on the  $M$  time-frequency blocks selected by the method in section 2.2 above. For the  $i^{th}$  EEG data  $X_i \in \mathbb{R}^{K \times T}, i \in N_l, l = 1, 2$  ( $K$  is the number of channels after channel selection in section 2.1,  $T$  is the number of sampling points from any time-frequency block, and  $N_l$  is the number of trails corresponding to class  $l$ ). The average spatial covariance matrix of class  $l$  is:

$$\Sigma_l = \frac{1}{N_l} \sum_{i=1}^{N_l} X_i X_i^T, \quad (8)$$

where  $T$  denotes the transpose operator. The purpose of CSP is to find the optimal spatial filter to maximize the variance ratio between the two classes of data:

$$\max_w J(w) = \frac{w^T \Sigma_1 w}{w^T \Sigma_2 w} \quad s.t. \quad \|w\|_2 = 1, \quad (9)$$

Where  $w \in \mathbb{R}^K$  is a spatial filter,  $\|\cdot\|_2$  is the  $l_2$  norm. This maximization solution is equivalent to solving the generalized eigenvalue problem  $\Sigma_1 w = \lambda \Sigma_2 w$ . A set of spatial filters  $W = [w_1, \dots, w_{2P}]$  is obtained by combining the eigenvectors corresponding to the  $P$  largest and  $P$  smallest generalized eigenvalues. Finally, the feature vector  $z^{(m)} = [z_1^{(m)}, \dots, z_{2P}^{(m)}]$  of the  $m^{th}$  time-frequency block is extracted by the following formula:

$$z_p^{(m)} = \log \left( \text{var}(w_p^T \mathbf{X}_i) \right), \quad p = 1, \dots, 2P. \quad (10)$$

According to the above steps, CSP features are extracted from all selected  $M$  time-frequency blocks, and the fusion feature  $Z \in \mathbb{R}^{2PM}$  is obtained:

$$Z = [z^{(1)}, \dots, z^{(M)}]. \quad (11)$$

LASSO (Least Absolute Shrinkage and Selection Operator), as a filter method for feature selection, uses specific statistical criteria to select features without relying on any classifier, and has been widely used in MI based BCI [34,35]. Lasso aims to minimize the sum of squares of residuals and is constrained that the sum of absolute values of coefficient vectors is less than the given constant. The specific functions are as follows:

$$g = \underset{g}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^N \|Z_i g - y_i\|_2^2 + \lambda \|g\|_1, \quad (12)$$

where  $Z_i$  and  $y_i$  are the fusion features and labels corresponding to the  $i^{th}$  trial, and  $\lambda$  is a nonnegative hyperparameter that controls the sparsity of the coefficient vector  $g$ . The features in  $Z_i$  corresponding to those non-zero entries in  $g$  are selected to form an optimized feature vector.

After correlation-based channel selection, optimal channel-based sparse time-frequency blocks selection and CSP feature extraction and selection, the final feature vectors are obtained and input into SVM with the radial basis function (RBF) kernel for the classification.

### 3. Experiment study

#### 3.1. EEG data description

1) Dataset 1: This dataset is from BCI competition IV dataset 1 [36], which records 59 channels EEG data of 4 healthy subjects (a, b, f, g), and each subject was asked to complete 100 trials of left hand and foot motor imagery. For the first 6 seconds of each trial, a fixed cross will be displayed in the center of the computer screen. The arrow with direction (left: left hand motor imagery; down: foot motor imagery) will be superimposed on the cross in 2–6 seconds as a hint, and the subjects performed the motor imagery tasks according to the cue during this period. Then the screen appears black in 6–8 s.

2) Dataset 2: This dataset is from BCI Competition III dataset Iva [37], which records 118 channels EEG data of 5 healthy subjects (aa, al, av, aw, ay), and each subject was asked to complete 140 trials of right hand and foot motor imagery. For the first 3.5 seconds of each trial, an arrow with direction (right: right hand motor imagery; down: foot motor imagery) is displayed in the center of the computer screen as a prompt, and the subjects performed the motor imagery tasks according to the cue

during this period. The subjects were then allowed to relax for 1.75 to 2.25 s.

### 3.2. Experimental evaluation and result

To investigate the performance of the proposed method on the above two datasets, we performed a uniform preprocessing using the channel selection algorithm described in section 2.1, followed by feature extraction using the FBCSP, B-CSP, B-SCSP and OCSB-CSP methods, respectively (as implemented below), and finally classification using an SVM based on the RBF kernel, where the penalty parameter C was determined by a 5-fold cross-validation is determined. The above experiments were repeated 5 times to evaluate the classification accuracy and computation time.

(1) FBCSP (Filter band CSP): By extracting and fusing CSP features from different filter bands. Specifically, the 4–40 Hz broadband raw data was decomposed into 17 overlapping subbands (4–8 Hz, 6–10 Hz, 8–12 Hz, 10–14 Hz, 12–16 Hz, 14–18 Hz, 16–20 Hz, 18–22 Hz, 20–24 Hz, 22–26 Hz, 24–28 Hz, 26–30 Hz, 28–32 Hz, 30–34 Hz, 32–36 Hz, 34–38 Hz, 36–40 Hz) by the fourth order Butterworth bandpass filter.

(2) B-CSP (Blocks-CSP): By extracting and fusing CSP features from each time-frequency block constructed by simultaneously sliding time windows and dividing frequency bands. Specifically, the data in Dataset 1 was divided into five overlapping time windows (i.e., 0–2 s, 0.5–2.5 s, 1–3 s, 1.5–3.5 s, 2–4 s), and the data in Dataset 2 was divided into four overlapping time windows (i.e., 0–2 s, 0.5–2.5 s, 1–3 s, 1.5–3.5 s). Then used the same method as in (1) to divide the frequency bands for each time window data.

(3) B-SCSP (Blocks-Sparse CSP): Based on (2), features were selected through Lasso. Specifically, 5-fold cross-validation was used to determine the hyperparameters  $\lambda$  in Lasso regression.

(4) OCSB-CSP: The construction method of time-frequency blocks was the same as (2). For specific subjects, optimal channel-based sparse time-frequency blocks CSP method was proposed, and the method described in (3) was used for feature extraction and selection.

After the above four algorithms were used to extract and select the respective CSP features, SVM with RBF kernel function was used for classification, and the penalty factor was determined by 5-fold cross-validation.

**Table 1.** Comparison of average classification accuracies  $\pm$  standard deviation (%) obtained by repeating the experiment 5 times for Dataset 1. For each subject, the highest average accuracy is marked in boldface.

Subjects	FBCSP	B-CSP	B-SCSP	OCSB-CSP
a	$80.23 \pm 3.82$	$82.67 \pm 3.91$	$84.76 \pm 3.80$	<b><math>86.91 \pm 3.58</math></b>
b	$70.62 \pm 5.34$	$70.86 \pm 5.49$	<b><math>74.66 \pm 4.93</math></b>	$74.17 \pm 4.81$
f	$80.92 \pm 4.11$	$82.60 \pm 4.01$	$82.63 \pm 3.58$	<b><math>84.55 \pm 3.49</math></b>
g	$89.84 \pm 2.13$	$91.34 \pm 2.09$	$91.86 \pm 2.00$	<b><math>93.46 \pm 1.95</math></b>
ave	$80.40 \pm 3.85$	$81.87 \pm 3.88$	$83.48 \pm 3.58$	<b><math>84.78 \pm 3.46</math></b>
p-value	$< 0.01$	$< 0.01$	0.12	–

Table 1 reports the comparison results of the classification accuracy rates for our proposed algorithm and three other algorithms on Dataset 1. We observe that our proposed method outperforms all these remaining methods. The average improvements achieved by our method were 4.37%, 2.91%

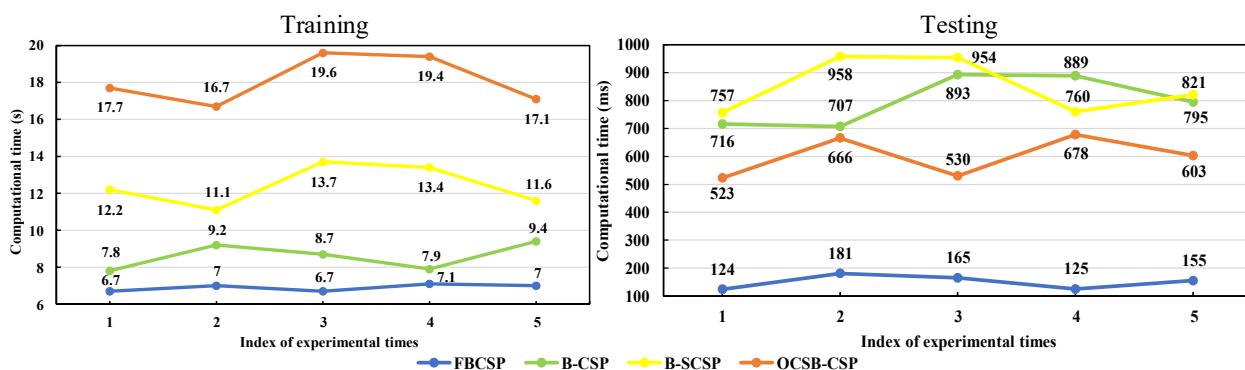
and 1.30% in comparison with FBCSP, B-CSP and B-SCSP, respectively. Our proposed method obtained very significant improvement on classification accuracy compared to FBCSP and B-CSP ( $p < 0.01$ ).

Table 2 shows the comparison results on Dataset 2. The average improvements achieved by our proposed method were 3.7%, 1.35% and 1.01% in comparison with FBCSP, B-CSP and B-SCSP, respectively. Our proposed method obtained very significant improvement on classification accuracy compared to FBCSP ( $p < 0.01$ ).

**Table 2.** Comparison of average classification accuracies  $\pm$  standard deviation (%) obtained by repeating the experiment five times for Dataset 2. For each subject, the highest average accuracy is marked in boldface.

Subjects	FBCSP	B-CSP	B-SCSP	OCSB-CSP
aa	$89.38 \pm 2.82$	$90.81 \pm 2.50$	$90.77 \pm 2.60$	<b><math>91.85 \pm 2.53</math></b>
al	$96.87 \pm 2.01$	$98.11 \pm 1.89$	<b><math>99.18 \pm 0.82</math></b>	$98.94 \pm 1.06$
av	$68.49 \pm 6.13$	$72.11 \pm 6.20$	<b><math>72.29 \pm 6.19</math></b>	$71.71 \pm 6.09$
aw	$91.18 \pm 2.65$	$91.98 \pm 2.77$	$94.83 \pm 2.49$	<b><math>97.14 \pm 2.50</math></b>
ay	$89.56 \pm 2.24$	$90.22 \pm 2.20$	$91.87 \pm 2.10$	<b><math>94.34 \pm 2.00</math></b>
ave	$87.10 \pm 3.17$	$88.65 \pm 3.11$	$89.79 \pm 2.84$	<b><math>90.80 \pm 2.84</math></b>
p-value	$< 0.01$	0.11	0.18	—

Additionally, we also compared the computational efficiency of each algorithm. Figure 2 shows the computational time evaluated for 5 replicated experiments of the model training and testing phases under the environment of python 3.7.3 on a desktop with 2.80 GHz CPU (i5-8400, 8 GB RAM). It can be seen from Figure 2(a) that our algorithm has relatively more hyperparameters, so it takes most of time for inner loop cross-validation to select the hyperparameters. Although our algorithm requires a longer computational time than other methods, the inner loop cross-validation is not necessary for testing but only for model training. The computational time for testing is shown in Figure 2(b), the proposed OCSB-CSP algorithm not only substantially improves the classification accuracy compared to the conventional FBCSP, but also reduces the computational burden compared to the B-CSP and B-SCSP that employ all time-frequency blocks, which is meaningful for the development and improvement of BCI applications.



**Figure 2.** The computational time spent for each comparison algorithm on training and testing by repeating the experiments 5 times respectively.

In order to demonstrate the superiority of the proposed method, it was also compared with existing winners' methods, including LRFCSP [38], SGRM [39], CCS-RCSP [28], BCS-CSP [40], and OCS-CSP [41]. Table 3 and Table 4 show the average classification accuracy of each method for multiple experiments on Dataset I and II, respectively. As shown in the tables, although the average classification accuracy of our proposed method does not display a significant advantage over existing methods, our method does not lose classification performance after a drastic dimensionality reduction process, but has some enhancement, especially in Dataset 2.

**Table 3.** Classification accuracies of the proposed OCSB-CSP method and other winners' methods on Dataset 1.

Subjects	methods				
	LRFCSP	CCS-RCSP	BCS-CSP	OCS-CSP	OCSB-CSP
a	87.0	85.5	78.5	86.5	<b>87.1</b>
b	69.0	67.0	<b>77.5</b>	69.5	74.2
f	87.5	79.5	<b>92.0</b>	87.5	84.6
g	92.5	94.5	87.0	94.0	<b>94.6</b>
ave	84.0	81.6	83.8	84.4	<b>85.1</b>
std	8.9	10.0	<b>6.0</b>	9.1	7.3
p-value	0.56	0.11	0.75	0.66	—

**Table 4.** Classification accuracies of the proposed OCSB-CSP method and other winners' methods on Dataset 2.

Subjects	methods				
	SGRM	CCS-RCSP	BCS-CSP	OCS-CSP	OCSB-CSP
aa	73.9	80.7	82.1	89.3	<b>91.9</b>
al	94.5	96.8	95.0	98.2	<b>98.9</b>
av	59.5	70.4	72.1	<b>73.5</b>	71.7
aw	80.7	92.9	90.7	92.9	<b>97.1</b>
ay	79.9	92.1	91.8	89.3	<b>94.3</b>
ave	77.7	86.6	86.3	88.6	<b>90.8</b>
std	11.3	9.7	8.3	<b>8.2</b>	9.8
p-value	<0.01	0.08	0.06	0.15	—

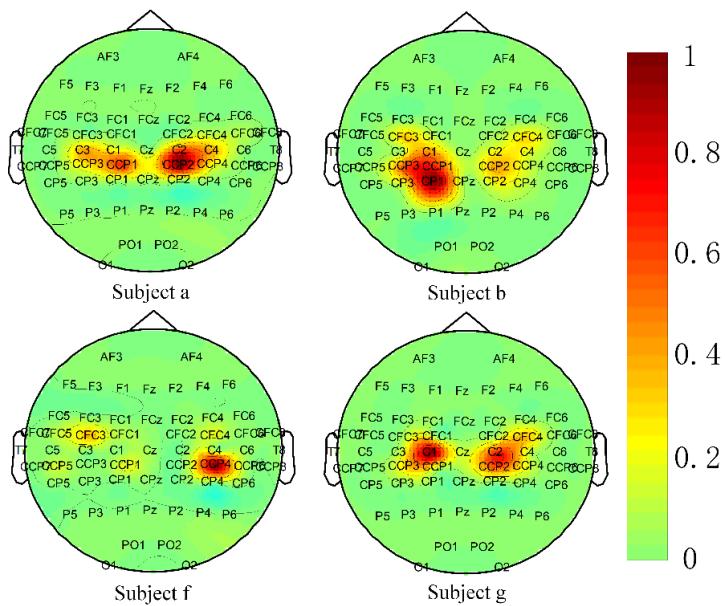
## 4. Discussion

### 4.1. Distribution of selected channels

We used the Pearson correlation coefficient-based channel selection method proposed in section 2.1 to count the channel selection results of all 160 training trials for each subject in the Dataset 1, as shown in Table 5. We can see that the results of channel selection for specific subjects are not the same, that is, different channels are selected, the number of channels selected is different, the number of times each channel is selected is different, and the optimal channels marked are different, so it is necessary to select channels for specific subjects.

**Table 5.** The specific channel selection results and the corresponding statistical times of four subjects in Dataset 1, and the selection results of the optimal channels are displayed in bold.

a	CFC4	CFC6	C3	C1	Cz	C2	C4	CCP3	CCP1	<b>CCP2</b>
num	6	4	18	13	1	27	17	12	30	<b>51</b>
	CCP4	CP3	CP1	CPz	CP2					
	10	1	4	3	3					
b	FCz	CFC3	CFC1	CFC2	CFC4	C3	C1	Cz	C2	C4
num	1	13	2	3	14	3	18	1	9	7
	CCP5	CCP3	CCP1	CCP2	CCP4	CP3	<b>CP1</b>	CPz	CP2	CP4
	2	26	23	14	5	6	<b>37</b>	1	12	3
f	FC3	FC2	FC4	CFC5	CFC3	CFC1	CFC4	CFC6	C3	Cz
num	2	1	2	11	32	8	21	1	2	2
	C4	C6	CCP5	CCP3	CCP1	<b>CCP4</b>	CCP6	CP3	CP1	CP2
	2	1	11	1	15	<b>82</b>	1	1	3	1
g	FC1	FC2	FC4	CFC3	CFC1	CFC2	CFC4	C3	<b>C1</b>	Cz
num	1	1	1	4	1	2	19	7	<b>49</b>	8
	C2	C4	CCP3	CCP1	CCP2	CCP4	CP3	CP1	CP2	
	41	10	13	5	26	8	1	2	1	



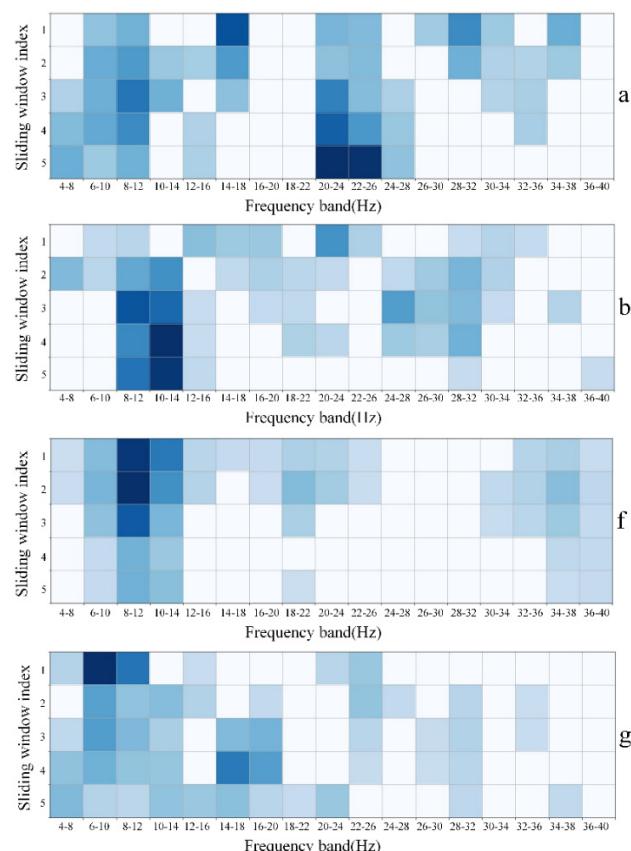
**Figure 3.** Brain topographic maps of channel selection distribution from four subjects in Dataset 1, the warmer the channel color, the more choices, that is, the higher the quality of the channel.

To intuitively reflect the distribution results of channel selection and the quality of each channel. We normalize the statistical results of each subject and draw brain topographic maps, as shown in Figure 3. It can be seen that the results of channel selection are indeed related to the specific subjects, but without exception, the channels selected by each subject are located in the motor perception region

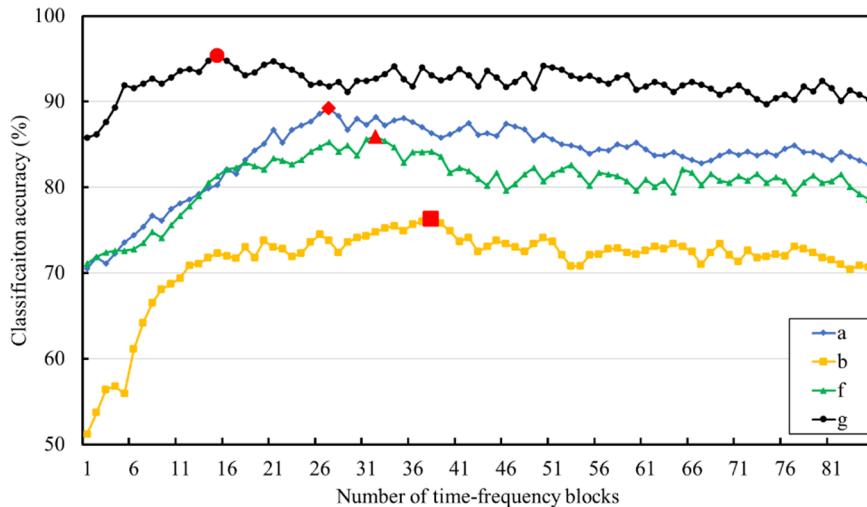
of the cerebral cortex, and the optimal channel (the warmest channel in the brain topographic map) appears near the CCP3 or CCP4.

#### 4.2. Sparse time-frequency blocks comparison

To improve the performance and efficiency of time-frequency blocks selection, as shown in 3.2, we propose a method that uses one-dimensional data of the optimal channel to calculate the Fisher ratio of each block as the basis for blocks selection. For the four subjects in Dataset 1, we set different thresholds to ensure that 45 blocks are selected for synchronous comparison, as shown in Figure 4. The blue blocks are the selected time-frequency blocks. The darker the color, the larger the Fisher ratio, which means that it has better MI classification ability. It can be found that the time-frequency blocks selected by our proposed algorithm from specific subjects are different, and the positions of significant time-frequency blocks are also different. Since MI tasks are usually unknown and varies between subjects, it is determined that a fixed time-frequency block cannot capture the most distinctive features, resulting in suboptimal accuracy, which further confirms the necessity of selecting time-frequency blocks. Furthermore, we consider that the EEG signal between adjoining sliding windows will not change too much, and the corresponding CSP features will not change too much, so the selection blocks of adjoining sliding windows should also ensure a certain consistency. The results presented in Figure 4 are consistent with the above.



**Figure 4.** Visualization of sparse time-frequency blocks by the OCSB-CSP algorithm for all subjects from Dataset 1. The blue blocks are the selected time-frequency blocks, and the darker the color, the larger the Fisher ratio and the better the MI classification ability.



**Figure 5.** Effects of varying time-frequency blocks selection number on the average classification accuracies of 5-fold cross-validation for all subjects in Dataset 1.

We used 5-fold cross-validation to determine a reasonable threshold for specific subjects. Figure 5 shows the impact of the change in the number of time-frequency blocks on the classification accuracies for all subjects in Dataset 1. It can be seen from this figure that when the number of blocks exceeds a certain number, the classification accuracy of each subject decreases to varying degrees. That is, we need to determine the number corresponding to the highest classification accuracy. Thus, the selection of sparse time-frequency blocks is very significant in improving the performance and operation speed of the model.

#### 4.3. Sparse feature selection

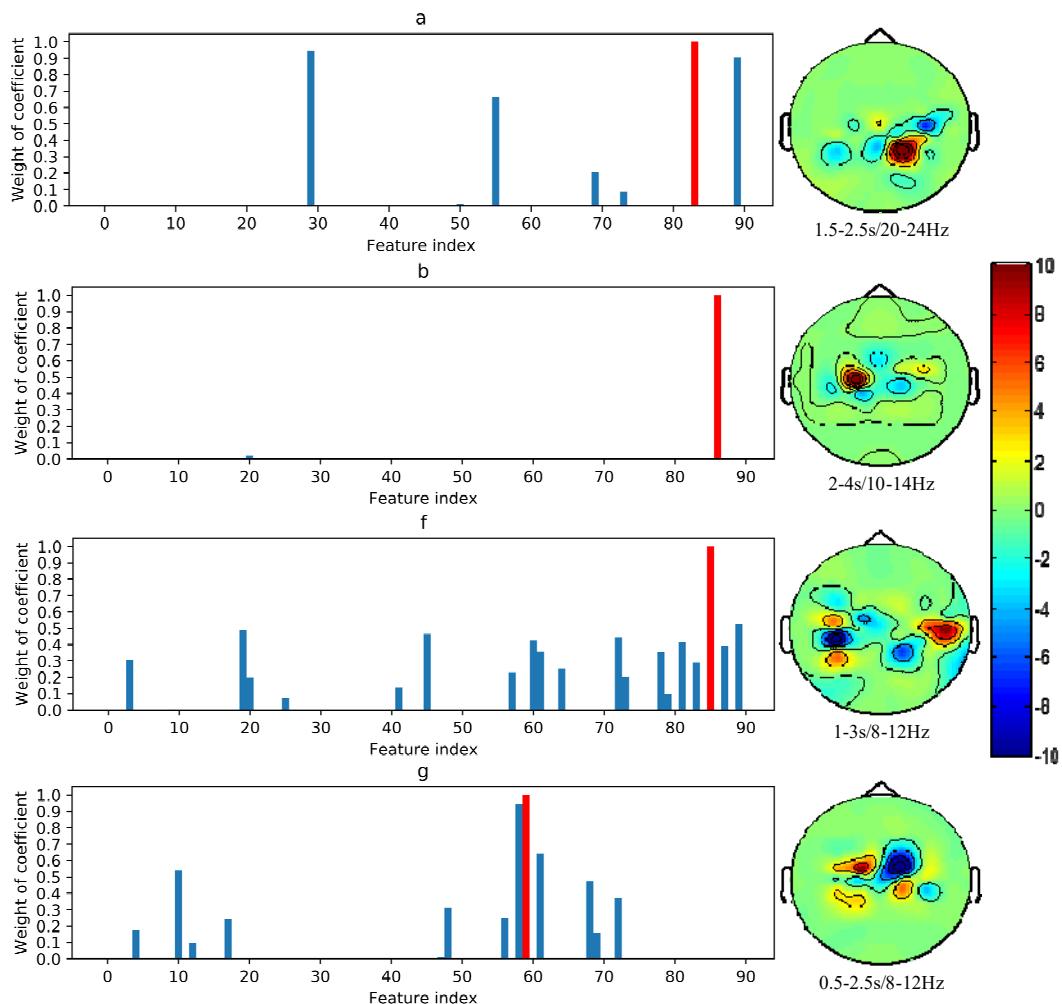
The regularization parameter  $\lambda$  plays an important role in the selection of CSP features based on lasso regression. A too large  $\lambda$  may exclude effective features while a too small one could not eliminate redundancy effectively. In this paper, the appropriate subject-specific  $\lambda$  was determined by 5-fold cross-validation, and the optimal sparse coefficient vector  $\mathbf{g}$  can be eventually learned from formula (12) to select the significant CSP features extracted from the sparse time-frequency blocks. Figure 6. reflects sparse coefficient vector and the most significant spatial filter learned by OCSB-CSP algorithm for each subject in Dataset 1. We can see that the significant features are sparse and subject-specific, and the brain topographic map of the most significant spatial filter further indicates the effectiveness of the proposed method for capturing the dynamical changing of SMRs.

In summary, we arrived at the following conclusions:

1) The channel selection section, for a single trial on a specific subject, using a filtered approach called Pearson correlation coefficient [17] to quickly eliminate redundant information between channels and diminish the dimensions of the data. In addition, all trials are counted in the form of voting, and the optimal channel with the most votes is marked. The experimental results show that the optimal channel is all from the area around CCP3 and CCP4 where ERD/ERS [42,43] is most obvious.

2) Aiming at the problem that the existing CSP feature extraction methods based on multiple time-frequency blocks have heavy computational burden, we propose an optimal channel-based sparse time-frequency blocks CSP feature extraction method to reduce the computational complexity and improve

the classification accuracy of the model. The specific innovations are as follows: the sparsity of time-frequency blocks based on one-dimensional data of the optimal channel, which ensures the accuracy of classification to a certain extent and improves the computational efficiency of the model. The innovative combination of extracting time-domain and frequency-domain features with Fisher ratio to define a reasonable indicator for time-frequency blocks selection, which has achieved better experimental results.



**Figure 6.** Weight of coefficient corresponding to each CSP feature and most significant spatial filter learned by the OCSB-CSP algorithm for each subject in Dataset 1.

3) The proposed framework involves more hyperparameters, such as the number of time-frequency blocks, penalty parameter of Lasso regression, penalty factor of SVM classifier, etc., and all these hyperparameters need to be determined according to specific subjects, and to some extent lead to an increased computational burden, which cannot be well applied across subjects, and this is not conducive to the development and application of BCI system. Future enhancements to the model are needed to improve the generalization capability of the model.

## 5. Conclusions

In this paper, an optimal channel-based sparse time-frequency blocks common Spatial pattern feature extraction method is introduced to enhance the classification accuracy and computational speed of MI tasks by efficiently selecting time-frequency blocks. In this proposed method framework, the channel selection method based on Pearson correlation coefficient was firstly cited to initially reduce the redundant information between channels and to mark the optimal channel for subsequent processing. The selection results show the reasonableness and efficiency of this method. Then the discriminative ability of each time-frequency block measured by defining Fisher ratio index based on the optimal channel of the one-dimensional EEG data was achieved to sparse the time-frequency blocks. The results indicate that the suggested method not only significantly reduces the data dimensionality, but also the selected time-frequency blocks are mostly distributed in the frequency bands relevant to the MI tasks. Finally, Lasso regression was performed to select the extracted multi-blocks CSP features and SVM was used for classification. Significant advantages were obtained when compared with the existing superior methods on both public datasets. The proposed OCSB-CSP algorithm achieves higher classification accuracy while reducing the computational burden of the model, which provides a new idea for the development and improvement of BCI applications.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 61671197, 61871427 and 61971168), Zhejiang Provincial Natural Science Foundation of China (No. LY18F030009), and the Foundation of Zhejiang Provincial Education Department of China (No. Y202044279). The authors would like to acknowledge the BCI competition IV dataset 1 and BCI Competition III dataset Iva which were used to test the algorithms proposed in this study.

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. R. Krepki, B. Blankertz, G. Curio, K. R. Müller, The Berlin Brain-Computer Interface (BBCI)-towards a new communication channel for online control in gaming applications, *Multimedia Tools Appl.*, **33** (2007), 73–90.
2. D. Saravanakumar, M. R. Reddy, A high performance hybrid SSVEP based BCI speller system, *Adv. Eng. Inf.*, **42** (2019), 100994–101003.
3. L. Shao, L. Y. Zhang, A. N. Belkacem, Y. M. Zhang, X. Q. Chen, J. Li, et al., EEG-Controlled Wall-Crawling Cleaning Robot Using SSVEP-Based Brain-Computer Interface, *J. Healthcare Eng.*, **2020** (2020), 1–11.
4. Z. Ma, Z. X. Xie, T. S. Qiu, J. Cheng, Driving event-related potential-based speller by localized posterior activities: An offline study, *Math. Biosci. Eng.*, **17** (2020), 789–801.
5. S. M. M. Martens, J. M. Leiva, A generative model approach for decoding in the visual event-related potential-based brain-computer interface speller, *J. Neural Eng.*, **7** (2010), 26003–26012.

6. I. Majidov, T. Whangbo, Efficient Classification of Motor Imagery Electroencephalography Signals Using Deep Learning Methods, *Sensors*, **19** (2019), 1736–1748.
7. Y. Y. Rong, X. J. Wu, Y. M. Zhang, Classification of motor imagery electroencephalography signals using continuous small convolutional neural network, *Int. J. Imaging Syst. Technol.*, **30** (2020), 653–659.
8. A. Schlogl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, *J. Neural Eng.*, **2** (2005), 14–22.
9. F. Cincotti, D. Mattia, C. Babiloni, F. Carducci, S. Salinari, L. Bianchi, et al., The use of EEG modifications due to motor imagery for brain-computer interfaces, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **11** (2003), 131–133.
10. F. V. Alvarez, R. R. Angevin, L. S. Sauer, S. S. Ros, Audio-cued motor imagery-based brain-computer interface: Navigation through virtual and real environments, *Neurocomputing*, **121** (2013), 89–98.
11. H. Yuan, B. He, Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives, *IEEE Trans. Biomed. Eng.*, **61** (2014), 1425–1435.
12. A. Dietrich, R. Kanso, A Review of EEG, ERP, and Neuroimaging Studies of Creativity and Insight, *Psychol. Bull.*, **136** (2010), 822–848.
13. V. Sakkalis, Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG, *Comput. Biol. Med.*, **41** (2011), 1110–1117.
14. T. Aflalo, S. Kellis, C. Klaes, B. Lee, Y. Shi, K. Pejsa, et al., Decoding motor imagery from the posterior parietal cortex of a tetraplegic human, *Science*, **348** (2015), 906–910.
15. S. Hetu, M. Gregoire, A. Saimpont, M. P. Coll, F. Eugene, P. E. Michon, et al., The neural network of motor imagery: An ALE meta-analysis, *Neurosci. Biobehav. Rev.*, **37** (2013), 930–949.
16. H. Ramoser, J. M. Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Trans. Rehabil. Eng.*, **8** (2000), 441–446.
17. J. Feng, E. Yin, J. Jin, R. Saab, I. Daly, X. Wang, et al., Towards correlation-based time window selection method for motor imagery BCIs, *Neural Networks*, **102** (2018), 87–95.
18. K. Wang, M. P. Xu, Y. J. Wang, S. S. Zhang, L. Chen, D. Ming, Enhance decoding of pre-movement EEG patterns for brain-computer interfaces, *J. Neural Eng.*, **17** (2020), 016033–016051.
19. M. A. Li, Y. F. Wang, X. Q. Zhu, J. F. Yang, A wrapped time-frequency combined selection in the source domain, *Biomed. Signal Proc. Control*, **57** (2020), 101748–101757.
20. N. Yamawaki, C. Wilke, Z. M. Liu, B. He, An enhanced time-frequency-spatial approach for motor imagery classification, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **14** (2006), 250–254.
21. P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, H. Wang, A Sliding Window Common Spatial Pattern for Enhancing Motor Imagery Classification in EEG-BCI, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–9.
22. Y. Zhang, C. S. Nam, G. X. Zhou, J. Jin, X. Y. Wang, A. Cichocki, Temporally Constrained Sparse Group Spatial Patterns for Motor Imagery BCI, *IEEE Trans. Cyber.*, **49** (2019), 3322–3332.
23. V. Peterson, D. Wyser, O. Lambercy, R. Spies, R. Gassert, A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multichannel EEG, *J. Neural Eng.*, **16** (2019), 016019–016031.
24. P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, G. Prasad, An automatic subject specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation, *Biomed. Signal Proc. Control*, **68** (2021), 102574–102581

25. J. Jin, C. Liu, I. Daly, Y. Y. Miao, S. R. Li, X. Y. Wang, et al., Bispectrum-Based Channel Selection for Motor Imagery Based Brain-Computer Interfacing, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **28** (2020), 2153–2163.
26. Q. G. Wei, W. Tu, *Channel Selection by Genetic Algorithms for Classifying Single-Trial ECoG during Motor Imagery*, 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008. Available from: <https://ieeexplore.ieee.org/abstract/document/4649230/>.
27. F. K. Onay, C. Kose, Assessment of CSP-based two-stage channel selection approach and local transformation-based feature extraction for classification of motor imagery/movement EEG data, *Biomed. Eng. Biomed. Tech.*, **64** (2019), 643–653.
28. J. Jin, Y. Y. Miao, I. Daly, C. L. BZuo, D. W. Hu, A. Cichocki, Correlation-based channel selection and regularized feature optimization for MI-based BCI, *Neural Networks*, **118** (2019), 262–270.
29. K. R. Müller, M. Krauledat, G. Dornhege, G. Curio, B. Blankert, Machine learning techniques for brain-computer interfaces, *Biomed. Tech.*, **49** (2004), 11–22.
30. C. Vidaurre, N. Kramer, B. Blankertz, A. Schlogl, Time Domain Parameters as a feature for EEG-based Brain-Computer Interfaces, *Neural Networks*, **22** (2009), 1313–1319.
31. K. K. Ang, Z. Y. Chin, H. H. Zhang, C. T. Guan, Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface, *IEEE Int. Joint Conf. Neural Networks*, **2008** (2008), 2390–2397.
32. K. D. Ghanbar, T. Y. Rezaii, M. A. Tinati, A. Farzamnia, *Correlation-Based Regularized Common Spatial Patterns for Classification of Motor Imagery EEG Signals*, 27th Iranian Conference on Electrical Engineering (ICEE), 2019. Available from: <https://ieeexplore.ieee.org/abstract/document/8786490/>.
33. Y. Zhang, G. X. Zhou, J. Jin, X. Y. Wang, A. Cichocki, Optimizing spatial patterns with sparse filter bands for motor-imagery based brain-computer interface, *J. Neurosci. Methods*, **255** (2015), 85–91.
34. R. Tibshirani, Regression Shrinkage and Selection via the Lasso, *J. Royal Stat. Soc. Ser. B Stat. Methodol.*, **58** (1996), 267–288.
35. R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, *J. Royal Stat. Soc. Ser. B Stat. Methodol.*, **67** (2005), 91–108.
36. G. Dornhege, B. Blankertz, G. Curio, K. R. Muller, Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms, *IEEE Trans. Biomed. Eng.*, **51** (2004), 993–1002.
37. B. Blankertz, G. Dornhege, M. Krauledat, K. R. Muller, G. Curio, The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects, *Neuroimage*, **37** (2007), 539–550.
38. Y. Park, W. Chung, Frequency-Optimized Local Region Common Spatial Pattern Approach for Motor Imagery Classification, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **27** (2019), 1378–1388.
39. Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, et al., Sparse Group Representation Model for Motor Imagery EEG Classification, *IEEE J. Biomed. Health Inf.*, **23** (2019), 631–641.
40. J. Jin, C. Liu, I. Daly, Y. Y. Miao, S. R. Li, X. Y. Wang, et al., Bispectrum-Based Channel Selection for Motor Imagery Based Brain-Computer Interfacing, *IEEE Trans. Neural Syst. Rehabil. Eng.*, **28** (2020), 2153–2163.
41. Y. Park, W. Chung, Optimal Channel Selection Using Correlation Coefficient for CSP Based EEG Classification, *IEEE Access*, **8** (2020), 111514–111521.

42. B. Graimann, J. E. Huggins, S. P. Levine, G. Pfurtscheller, Visualization of significant ERD\_ERS patterns in multichannel EEG and ECoG data, *Clin. Neurophysiol.*, **113** (2002), 43–47.
43. G. Pfurtscheller, C. Neuper, C. Andrew, G. Edlinger, Foot and hand area mu rhythms, *Int. J. Psychophysiology*, **26** (1997), 121–135.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)